HEMNMA-3D: Cryo Electron Tomography Method Based on Normal Mode Analysis to Study Continuous Conformational Variability of Macromolecular Complexes

Mohamad Harastani, Mikhail Eltsov, Amélie Leforestier, Slavica Jonic

frontiers

HEMNMA-3D: Cryo Electron Tomography Method Based on Normal Mode Analysis to Study Continuous Conformational Variability of Macromolecular Complexes

Mohamad Harastani [1] **Mikhail Eltsov** [2] **Amélie Leforestier** [3] **Slavica Jonic** [1,*]

[1] *IMPMC - UMR 7590 CNRS, Sorbonne Université, Muséum National d'Histoire Naturelle, Paris, France*

[2] *Department of Integrated Structural Biology, Institute of Genetics and Molecular and Cellular Biology, Illkirch, France*

[3] *Laboratoire de Physique des Solides (LPS), UMR 8502 CNRS, Université Paris-Saclay, Orsay, France*

Correspondence*:
Slavica Jonic, IMPMC - UMR CNRS 7590, Sorbonne Univerisité, 4 Place Jussieu, 75005 Paris, France
slavica.jonic@upmc.fr

ABSTRACT

Cryogenic electron tomography (cryo-ET) allows structural determination of biomolecules in their native environment (*in situ*). Its potential of providing information on the dynamics of macromolecular complexes in cells is still largely unexploited, due to the challenges of the data analysis. The crowded cell environment and continuous conformational changes of complexes make difficult disentangling the data heterogeneity. We present HEMNMA-3D, which is, to the best of our knowledge, the first method for analyzing cryo electron subtomograms in terms of continuous conformational changes of complexes. HEMNMA-3D uses a combination of elastic and rigid-body 3D-to-3D iterative alignments of a flexible 3D reference (atomic structure or electron microscopy density map) to match the conformation, orientation, and position of the complex in each subtomogram. The elastic matching combines molecular mechanics simulation (Normal Mode Analysis of the 3D reference) and experimental, subtomogram data analysis.

The rigid-body alignment includes compensation for the missing wedge, due to the limited tilt angle of cryo-ET. The conformational parameters (amplitudes of normal modes) of the complexes in subtomograms obtained through the alignment are processed to visualize the distribution of conformations in a space of lower dimension (typically, 2D or 3D) referred to as space of conformations. This allows a visually interpretable insight into the dynamics of the complexes, by calculating 3D averages of subtomograms with similar conformations from selected (densest) regions and by recording movies of the 3D reference's displacement along selected trajectories through the densest regions. We describe HEMNMA-3D and show its validation using synthetic datasets. We apply HEMNMA-3D to an experimental dataset describing *in situ* nucleosome conformational variability. HEMNMA-3D software is available freely (open-source) as part of ContinuousFlex plugin of Scipion V3.0 (http://scipion.i2pc.es).

Keywords: Cryo electron tomography, Continuous conformational changes, Flexible-reference alignment, Normal Mode Analysis, Nucleosome

1 INTRODUCTION

Cryogenic electron microscopy (cryo-EM) image collection and analysis technique referred to as single-particle analysis (SPA) allows near-atomic structural resolution of purified biomolecular complexes (*in vitro*). It is based on the principle of reconstructing a three dimensional (3D) structure from two dimensional (2D) parallel-beam projection images of vitrified specimens containing many copies of the same macromolecular complex at unknown orientations and positions. The 3D reconstruction requires extracting macromolecular complexes (particles) from the collected images into individual (single-particle) images and determining the particle orientation and position in every single-particle image. On the other hand, cryogenic electron tomography (cryo-ET) is gaining popularity for studying biomolecular complexes in their native environments (*in situ*). Cryo-ET requires the acquisition of multiple 2D projection images of the specimen in a range of orientations. In most practices, the specimen is physically rotated around a single axis (perpendicular to the electron beam) inside the cryo electron microscope. An image is collected at each tilting angle in a specific range (e.g., -60° to 60° with a step of 1°), yielding a tilt series representing 2D projections of the specimen. The tilt series is then used to computationally reconstruct a 3D volume called tomogram. A tomographic reconstruction typically contains hundreds of copies of a target biomolecular complex at unknown orientations and positions. These copies are then identified and extracted into individual (single-particle) volumes called subtomograms, either manually or semi-automatically (via template matching methods). Subtomograms suffer from a low signal-to-noise ratio (SNR), which is due to exposing the sample to a low electron dose during data acquisition in order to preserve the fragile biological structure. Additionally, subtomograms suffer from the so-called missing wedge artefacts, which are due to inability to include in the 3D reconstruction the images from all orientations (the maximum tilt angle in the microscope is usually limited to $\pm$ 60°). The missing wedge artefacts are often observed as elongation along the beam axis, blurring and distracting caustics in the subtomograms. Due to the low SNR and the missing wedge artefacts, cryo-ET data processing is mainly based on rigid-body aligning and averaging

many subtomograms to enhance the data quality and reveal the targeted biomolecular structure Leigh et al. (2019).

The primary technique for macromolecular structural determination is so-called subtomogram averaging (StA), in which subtomograms are classified, rigid-body aligned and averaged into 3D density maps iteratively (Albert et al. (2017); Böck et al. (2017); Bykov et al. (2017); Davies et al. (2018); Guo et al. (2018); Hutchings et al. (2018); Kaplan et al. (2019); Kovtun et al. (2018); Mahamid et al. (2016); Mosalaganti et al. (2018); Park et al. (2018); Pfeffer et al. (2017); Rapisarda et al. (2019); Riedel et al. (2017); Schur et al. (2016); Wan and Briggs (2016); Wan et al. (2017)). However, with recent instrumentation and software development, more research moves in the direction of studying single-particle subtomograms individually (with no or a minimum of averaging) by developing new methods for denoising, missing wedge correction, and 3D reconstruction Moebel and Kervrann (2020); Zhai et al. (2020); Zhang and Ren (2012).

Biomolecular complexes are not rigid but flexible entities with gradual (continuous) conformational transitions, and this flexibility is usually referred to as continuous conformational variability. If not properly taken into account, conformational heterogeneity limits the resolution of the resulting 3D structure. However, SPA research in the last decade has shown that disentangling the different conformations and identifying the conformational transitions from heterogeneous samples is valuable to study molecular mechanisms of action of complexes Abeyrathne et al. (2016); Banerjee et al. (2016); Dashti et al. (2014); Haselbach et al. (2018); Jin et al. (2014); Zhou et al. (2015). The majority of available SPA computational methods rely on optimized biochemical specimen preparation protocols and data classification, and simplify the problem of conformational heterogeneity by assuming that the data can be classified into a small number of different conformations Elad et al. (2008); Fu et al. (2007); Lyumkis et al. (2013); Penczek et al. (2006, 2011); Scheres (2012). However, some SPA methods explicitly take into account continuous conformational variability and aim at determining the full conformational distribution Andén and Singer (2018); Dashti et al. (2014); Frank and Ourmazd (2016); Harastani et al. (2020); Jin et al. (2014); Katsevich et al. (2015); Sorzano et al. (2014); Tagare et al. (2015). They represent images in a low-dimensional space, referred to as space of conformations or energy landscape, and allow a 3D visualization of conformational changes along trajectories in this space. For more information on SPA methods for continuous conformational variability analysis, the reader is referred to the recent reviews by Jonić (2017) and Sorzano et al. (2019).

Methods reported to deal with cryo-ET data heterogeneity are based on rigid-body alignment and can be classified into i) post-alignment classification approaches, and ii) simultaneous alignment and classification approaches (Chen et al. (2014); Förster et al. (2008); Himes and Zhang (2018); Stölken et al. (2011); Xu et al. (2012); Scheres et al. (2009); Bharat and Scheres (2016)). In the first family, the starting point is usually the covariance matrix representing the similarities of each pair of aligned subtomograms. The covariance matrix serves as a basis for a classification technique with some variants including dimensionality reduction. The second family of methods is based on competitive alignment. An example of the competitive alignment is a multireference alignment in which a subtomogram is compared with a set of different references provided by an expert user based on a prior knowledge and, then, attributed to the reference that yields the highest

similarity score. Another example is maximum-likelihood-based alignment, where each subtomogram contributes to all references with a probability. The main drawback of the post-alignment classification approaches is that the classification is heavily dependent on the alignment quality that degenerates with broadly heterogeneous specimens. The main drawback of the simultaneous alignment and classification methods is that the number of classes must be decided and set prior to the use of the methods. Besides, the methods that require prior knowledge of the specimen's anticipated conformations are prone to overfitting and data misinterpretation. Finally, as macromolecular complexes are not rigid but flexible entities with continuous conformational transitions, particles assigned to the same class will rarely, if ever, have perfectly identical conformations. For more information regarding the classification based techniques for cryo-ET conformational heterogeneity, the reader is referred to a recent review on the available techniques by Castaño-Díez and Zanetti (2019).

Existing multivariate statistical analysis techniques adapted to cryo-ET analysis of continuous flexibility of particular systems have been used previously (e.g., Mattei et al. (2016)). However, to the best of our knowledge, no method is currently available that has been specifically designed for cryo-ET analysis of continuous conformational variability of a general-case macromolecular complex. In this article, we present one such method, named HEMNMA-3D, which allows analyzing continuous conformational variability of macromolecular complexes by cryo-ET. It is inspired by HEMNMA, a method for continuous conformational variability analysis in SPA Harastani et al. (2020); Jin et al. (2014); Sorzano et al. (2014). HEMNMA interprets the conformation in each cryo-EM single-particle image by comparing this image with 2D projections of a 3D reference (an atomic structure or a density map) deformed elastically using normal modes. Normal Mode Analysis (NMA) is a method for molecular mechanics simulation. One of its main applications is elastic deformation of an existing atomic structure of one conformation to fit an electron microscopy density map (EM map) of a different conformation of the same macromolecule, which is usually known as normal mode flexible fitting and allows obtaining atomic resolution models for the EM map Tama et al. (2004b,a). However, as HEMNMA, HEMNMA-3D can use normal modes of an atomic structure or a density map. As in the case of the reference density-map structure in HEMNMA, HEMNMA-3D converts the density map into a collection of 3D Gaussian functions, referred to as pseudoatoms, and computes normal modes of this pseudoatomic structure, following the procedures described in Jin et al. (2014); Jonić and Sorzano (2016); Nogales-Cadenas et al. (2013). HEMNMA-3D uses the atomic or pseudoatomic normal modes to elastically deform the 3D reference to match the conformation of the complex in the given series of subtomograms (3D data) (see Figure 1). More precisely, HEMNMA-3D uses a combination of elastic (based on normal modes) and rigid-body 3D-to-3D iterative alignments of the 3D reference to match the conformation, orientation, and position of the complex in each subtomogram, and includes compensation for the missing wedge. The conformational parameters (amplitudes of normal modes) of the complexes in subtomograms obtained through the alignment are then processed to visualize the distribution of conformations in a space of lower dimension (typically, 2D or 3D) referred to as space of conformations. This space allows a visually interpretable insight into the dynamics of complexes, by

calculating 3D averages of subtomograms with similar conformations from selected (densest) regions and by recording movies of the 3D reference's displacement along selected trajectories in the densest regions.

In this article, we describe HEMNMA-3D and show its validation using synthetic datasets. Additionally, we show an application of HEMNMA-3D with an experimentally obtained dataset for *in situ* nucleosome conformational variability. HEMNMA-3D software is available freely (open-source) as part of ContinuousFlex plugin of Scipion V3.0 (http://scipion.i2pc.es). The article is organized as follows: Section 2 describes building blocks of HEMNMA-3D workflow, notably, in 2.5 we describe the software developed for this method. In Section 3, we present i) the process of synthesis of test datasets and HEMNMA-3D validation using these synthetic test data, and ii) use of HEMNMA-3D with experimental, *in situ* nucleosome data, and we discuss these results. The conclusions are provided in Section 4.

2 MATERIALS AND METHODS

The flowchart in Figure 2 describes the workflow of the proposed method, which was inspired by the workflow of HEMNMA (Harastani et al. (2020); Jin et al. (2014)). A graphical summary of the method is presented in Figure 3. The workflow comprises the following steps: (1) Input: the input to the method are a reference structure and a set of subtomograms. In the case where the reference structure is a density map (a 3D volume such as an EM map or a subtomogram average), a conversion to 3D Gaussian functions (pseudoatoms) takes place. (2) Normal mode analysis of the reference atomic structure or the reference pseudoatomic structure (obtained by converting the reference density map into 3D Gaussian functions in the previous step). (3) Combined iterative elastic and rigid-body 3D-to-3D alignment of the reference structure with each input subtomogram independently from other subtomograms, with missing wedge compensation. (4) Visualization of the computed conformations, after projecting the conformational parameters obtained for all subtomograms onto a low-dimensional space. In the remaining part of this section, we describe these steps in more detail. Please note that the first two steps of the workflow are exactly as those of HEMNMA and were thoroughly presented, tested and discussed in our previously published works on HEMNMA, its tools and applications (Harastani et al. (2020); Jin et al. (2014); Jonić and Sorzano (2016); Nogales-Cadenas et al. (2013); Sorzano et al. (2014)). However, for completeness of the present article, we here recall their basic principles. We close this section by a brief description of the software implemented for HEMNMA-3D.

2.1 Input reference and conversion of reference density maps into pseudoatoms

A reference structure of the molecule targeted in the subtomograms can be used in the form of an atomic model (PDB formatted files) or a density map, such as an EM map (SPA reconstruction) or a subtomogram average (obtained using classical StA without taking into account conformational heterogeneity). Although our method can be used with both atomic and density-map reference structures, one should prefer the use of a reference density map from the data at hand, if it can be obtained. If a reference density map is used, it must be converted into a collection of Gaussian functions (pseudoatoms) with a carefully selected standard deviation (pseudoatom size, whose default value is 1 voxel Jonić and Sorzano (2016);

Nogales-Cadenas et al. (2013)). The pseudoatom size should lead to a structure (called pseudoatomic structure) that, converted back to a density map, approximates the input density map with a small error (given a target approximation error, whose default value is 5% Jonić and Sorzano (2016); Nogales-Cadenas et al. (2013)). Optionally, a mask on the density map can be used prior to the conversion into pseudoatoms (e.g., a spherical binary mask of a given radius) to reduce background noise. Such masks may also be useful if applied on input cryo-ET subtomograms to maximize the chance of having a single molecular complex in each subtomogram (Preprocessing block in the workflow in Figure 2(A)).

2.2 Normal mode analysis

This step involves computing normal modes of a reference atomic or pseudoatomic structure, for the 3D-to-3D elastic alignment in the next step. The computation of normal modes is based on the elastic network model (Tama et al. (2002); Tirion (1996)) by representing the interaction between the (pseudo-)atoms as if they are locally connected by elastic springs (within a cutoff distance). Normal Mode Analysis requires the diagonalization of a $3N \times 3N$ matrix of second derivatives of the potential energy (Hessian matrix), where N is the number of nodes in the elastic network model determined by the total number of atoms (or pseudoatoms) in the input reference. In the case of atomic structures, we use the rotation-translation block (RTB) method, which divides the structure into blocks (one or a few consecutive residues per block) whose rotations and translations are considered rather than all degrees of freedom for all atoms Durand et al. (1994); Tama et al. (2000). Since the RTB method reduces the basis for Hessian diagonalization, it allows fast computing of normal modes. Since pseudoatomic structures usually contain fewer nodes (pseudoatoms) than atomic structures, normal modes can be obtained by a direct diagonalization of the $3N \times 3N$ Hessian, which is referred to as the Cartesian method. As in the case of HEMNMA, we here use the RTB and Cartesian method implementations of Suhre and Sanejouand (2004) and Tama et al. (2002), respectively. Larger values of the interaction cutoff distance (the distance below which atoms or pseudoatoms do not interact) lead to more rigid motions. The atomic interaction cutoff distance may be set manually (by default 8 Å) and the pseudoatomic cutoff distance is recommended to be computed automatically based on the distribution of the pseudoatomic pairwise distances (e.g., as the value below which is a given percentage of all distances as in Harastani et al. (2020); Jin et al. (2014); Jonić and Sorzano (2016); Nogales-Cadenas et al. (2013)). The modes are computed along with their respective collectivity degrees, which count the number of atoms or pseudoatoms affected by the mode as in Brüschweiler (1995). To allow faster data analysis and avoid noise overfitting in the 3D-to-3D elastic alignment in the next step, we select a subset of normal modes (usually, less than 10) with lowest frequencies and highest collectivities, as previously described (Harastani et al. (2020); Jin et al. (2014); Sorzano et al. (2014)). Low-frequency high-collectivity normal modes have been shown to be relevant to functional conformational changes Delarue and Dumas (2004); Ma (2005); Suhre et al. (2006); Tama and Sanejouand (2001); Tama and Brooks III (2006); Wang et al. (2004). The first six (lowest-frequency) normal modes are related to rigid-body transformations and are thus not used for the 3D-to-3D elastic alignment in the next step. The rigid-body 3D-to-3D alignment is done without using these rigid-body normal modes, as explained in the next paragraph. Additionally, a prior knowledge about the conformational transitions of the complex under study can be used to select the

normal modes for the use in the next step. The HEMNMA-3D graphical interface helps the user decide which normal modes to select. The reader is referred to Ma (2005) and Tama and Brooks III (2006) for reviews on the usefulness and limitations of NMA.

2.3 Combined iterative elastic and rigid-body 3D-to-3D alignment

This step, represented in Figure 2(B), is the backbone of the proposed method. It has been inspired by the combined iterative elastic and rigid-body 3D-to-3D alignment step of StructMap method (Sorzano et al. (2016)), which was proposed for pairwise similarity analysis of SPA high-resolution EM maps (no missing wedge). In HEMNMA-3D proposed here, this step comprises simultaneous NMA-based elastic alignment (search for amplitudes of a linear combination of normal modes) and rigid-body alignment (search for orientation and position, meaning 3 Euler angles and x, y, and z shifts) of the reference structure with each given subtomogram. It refines the amplitudes of displacement along each used normal mode (elastic parameters) as well as the angles and shifts (rigid-body parameters) of the reference structure until the best match is obtained between this reference structure and the given subtomogram. The latter is achieved by maximizing the similarity between the subtomogram and the density volume from the elastically deformed, oriented and shifted reference, and includes missing wedge compensation. The missing wedge compensation is done by calculating the cross-correlation between the reference and subtomogram density maps only in the region of the Fourier space where the data can be trusted, i.e., by constraining the cross-correlation evaluation to the Fourier space region that excludes the missing wedge region (the region outside of the one specified by the tilt angle range, e.g., -60° to +60°). To maximize this constrained cross-correlation (CCC), we use a variant of Powell's UOBYQA method, which subjects the objective function to a trust-region radius Berghen and Bersini (2005). To control the elastic deformation with highly noisy data, the radius of the trust region is adjusted iteratively. The scaling factor of the initial trust-region radius is a parameter that controls the normal-mode amplitude search range and can be modified by the user. It should have a positive value and its default value of 1 produces good results in general. It may be increased (typically to a value between 1 and 2) or decreased (e.g., between 0.5 and 0.9), if expecting larger or smaller conformational changes, respectively. For each subtomogram, the normal mode amplitudes are initiated with zeros, meaning that the non-deformed reference is used in the first iteration. As the iterations evolve, the reference model is displaced with the new guesses of the normal mode displacement amplitudes, converted into a volume and rigid-body aligned with the subtomogram using the method of fast rotational matching. Fast rotational matching has been largely used for rigid-body fitting of atomic models to high-SNR consensus EM maps Kovacs and Wriggers (2002); Kovacs et al. (2003). It has been extended to alignment of noisy subtomograms in Chen et al. (2013) and this implementation is used in our work. At the end of each iteration, the CCC is found and fed to the numerical optimizer Berghen and Bersini (2005). The iterations repeat until the final value of the trust-region radius or the maximum number of iterations is reached.

2.4 Visualizing and utilizing the space of conformations

The number of elastic alignment parameters (normal mode amplitudes) is determined by the number of selected normal modes for the 3D-to-3D elastic alignment. The ensemble of normal mode amplitudes (for

all subtomograms) can be projected onto a lower-dimensional space, so-called conformational space, using a dimensionality reduction technique. Here, we use linear Principal Component Analysis (PCA) as it is the most widely known and intuitively clear dimensionality reduction method, but other dimension reduction methods could also be used (linear or nonlinear). The dimensionality reduction is usually performed to two or three dimensions, which allows a global data display and easier modeling of conformational changes. Each point in the conformational space represents a subtomogram and close points correspond to similar conformations in the subtomograms. The points that differ significantly from the remaining observations (too isolated, outlier points) may be excluded from the further analysis, by excluding the points below a certain p-value based on the Mahalanobis distance (the distance between each point and the whole distribution) Mahalanobis (1936). The excluded points can be explained by the fact that some orientations of the molecule combined with the missing wedge artifacts and the high noise make some volumes more difficult to align with the elastically deformed reference. After excluding such outlier points, the space of conformations can be analyzed to reveal molecular dynamics. This can be done by averaging subtomograms of similar conformations in the densest regions of the conformational space or by exploring the densest regions by fitting curves (approximation by line segments) through the data and displacing the reference structure along these curves (referred to as trajectories) to animate the motion along them. The 1-CCC color bar of the conformational space shows coloring the points according to the value of the CCC between the subtomograms and the density maps from the elastically deformed reference model. Subsequently, the colors provide the level of confidence in the obtained conformations. Those subtomograms in which we have less confidence (subtomograms with lower CCC values) than in the "consensus" observations (subtomograms with higher CCC values) can be eliminated from the group averages.

2.4.1 Averaging subtomograms of similar conformations

Close points in the conformational space can be grouped, which results in grouping subtomograms of similar conformations and averaging them. Before computing group averages, the rigid-body alignment parameters found during the combined iterative elastic and rigid-body alignment are applied on the subtomograms. Optionally, before computing group averages, the missing-wedge Fourier space region of individual subtomograms may be filled in with the corresponding region of the global average computed from all subtomograms. A similar procedure of missing wedge filling of individual subtomograms is used in EMAN2 software package Galaz-Montoya et al. (2015). The subtomogram averages obtained from the selected groups of subtomograms can be overlapped and compared to understand the conformational changes of the complex in the given set of subtomograms.

2.4.2 Animating motions (trajectories)

Distinct trajectories can be determined through the data in the conformational space, and animated to see the motion of the biomolecule while it is displaced along the trajectory. To animate a trajectory, several points (e.g. 10) along the trajectory should be mapped back to the original displacement space (e.g. using inverse PCA), resulting in elastic alignment parameters that can be used to deform the reference atomic or pseudoatomic structure. Concatenating and displaying the resulting structures can show a movie-like

animation of the reference biomolecule traveling across the specified trajectory.

2.5 Software implementation and technical details

The software of HEMNMA-3D method proposed here is freely available (open-source). It is a part of ContinuousFlex plugin for the open-source software Scipion3 De la Rosa-Trevín et al. (2016). ContinuousFlex was introduced in Harastani et al. (2020) and also contains HEMNMA software. The software provides a graphical user interface (GUI) and is empowered with a C++ backend with a message passing interface (MPI) parallelization scheme to efficiently analyze large datasets (simultaneous analysis of N subtomograms using N computing threads). We tested the software on our local workstations and on supercomputer centres. On our local workstations (2.2GHz Intel Xeon Silver 4214 CPU processors), the current implementation takes around 10 and 30 minutes to analyze a subtomogram of size 64^3 voxels with three and six normal modes, respectively. These times are reported for the two types of complexes used in this article, together with the number of normal modes used in these two cases. They are the average times required for all iterations of analyzing one subtomogram using a single computing thread while the different subtomograms are analyzed in parallel using different computing threads (if the time is measured for the entire dataset, it will vary with the size of the dataset). It should be noted that the software allows the use of any number of MPI threads and any number of normal modes. However, the more modes are used, the slower the processing. Finally, it should be noted that there is no constraint regarding the size of the dataset (the number of subtomograms) or the size of the individual subtomograms (the number of voxels) that can be analyzed with our software.

3 RESULTS AND DISCUSSION

In this section, we present and discuss the results of HEMNMA-3D with synthetic and experimental subtomograms.

3.1 Synthesizing datasets for testing the method performance

For testing HEMNMA-3D in general, and the combined elastic and rigid-body 3D-to-3D alignment module in particular (which is the core module of the proposed method), we synthesized two datasets of conformationally heterogeneous subtomograms that mimic discrete and continuous conformational variability, called ”Discrete” and ”Continuous” datasets respectively. The flowchart for the data generation procedure is shown in Figure 4 and is detailed in the following.

The ”Discrete” dataset comprises 900 synthetic subtomograms representing three different (synthetic) conformations of the atomic PDB:4AKE structure (Müller et al. (1996)) of adenylate kinase chain A (1656 atoms), i.e. 300 subtomograms per conformation. We generated this dataset using the atomic PDB:4AKE structure and its first two non-rigid-body normal modes, i.e. modes 7 and 8. Precisely, the three conformations are represented by the following amplitudes of modes 7 and 8: (mode 7, mode 8) $\in$ $\{$(-150, 0), (+150, 0), (0, +150)$\}$.

The ”Continuous” dataset comprises 1000 synthetic subtomograms representing a continuum of conformations of the same PDB:4AKE structure. We generated this dataset using this atomic structure and

its modes 7 and 8 using a linear relationship between the amplitudes of the two modes. More precisely, the synthesized amplitudes of modes 7 and 8 were identical and randomly distributed in the range [-200, +200] (uniform distribution).

Normal-mode amplitudes do not have a physical unit. Nonetheless, the Root Mean Square Deviation (RMSD) Kufareva and Abagyan (2011) between the reference atomic coordinates and these coordinates displaced using normal-mode amplitudes transforms the normal mode amplitudes in physical units. To provide a basis for further evaluation of the method performance, we found a RMSD of 6.95 Å corresponding to the displacement using the amplitude of 200 for each of the two combined modes 7 and 8 (this represents one half of the full range of the synthesized motion).

To generate a subtomogram, first, we deform the atomic structure using appropriate amplitudes for the selected normal modes depending on the dataset in hand, i.e. we use (mode 7, mode 8) = (+150, 0) or (-150, 0) or (0, +150) to create a subtomogram in the "Discrete" dataset, while we assign a random value in the range [-200, 200] for both mode 7 and mode 8 to generate a subtomogram in the "Continuous" dataset. Then, we convert the deformed structure to a volume of size 64^3 voxels and the voxel size of 2.2 $Å^3$ (Peng et al. (1996)). Afterwards, we rotate and shift this volume in 3D space using random Euler angles (each of the three Euler angles was randomized in the range [0°, 360°]) and random shifts (the shift along each of the x, y, and z axes was randomized in the range $[-5, +5]$ voxels), and we project the rotated and shifted volume using tilt values -60° to +60° to obtain a tilt series. We simulate microscope conditions by adding heavy noise (signal to noise ratio SNR = 0.01) and modulating the images with a contrast transfer function (CTF) of defocus -1 μm, so that one part of the noise is affected by the CTF and the other is not (Chen et al. (2013); Sorzano et al. (2007)). Finally, we reconstruct a volume (our synthetic subtomogram) from the tilt series using a Fourier reconstruction method Sorzano et al. (2013). A few examples of the synthesized subtomograms (SNR = 0.01) and their less noisy version (SNR = 0.5, for illustration) is presented in Figure 5.

3.2 Synthetic discrete-type conformational variability

In this experiment, our goal is to retrieve the ground-truth amplitudes of normal modes 7 and 8 by the combined elastic and rigid-body alignment (the core module of HEMNMA-3D) of a reference model with the subtomograms in the "Discrete" dataset. In other words, the goal is to find a solution for the challenging inverse problem of finding the conformation of the structure in each subtomogram. Since the proposed method can use two choices for the reference model, namely, an atomic structure and a density map (e.g., an EM map or a subtomogram average), we performed two types of tests. In the first test type, the atomic structure used to generate the synthetic subtomograms (chain A of the PDB:4AKE) was used as a reference for retrieving normal mode amplitudes of the synthetic subtomograms. In the second test type, we converted (Peng et al. (1996)) the atomic structure into a density map (volume) of size 128^3 voxels and voxel size of 1 $Å^3$, and we used this density map as a reference for retrieving normal mode amplitudes of the synthetic subtomograms. In the case of the reference density map, normal modes were computed from the corresponding structure obtained by converting the density map into pseudoatoms (1675 pseudoatoms for the given pseudoatom radius of 1.25 voxels and the target approximation error of 5%). In both cases

(reference atomic structure and reference pseudoatomic structure, with their corresponding normal modes), we used three modes (modes 7, 8 and 9) instead of only two modes (modes 7 and 8 that were used to generate synthetic subtomograms), to make the 3D-to-3D elastic and rigid-body alignment task even more challenging. Figure 6 presents the estimated amplitudes of normal modes 7 and 8 (the estimated amplitude of normal mode 9 is close to 0 and is therefore not shown graphically). Table 1 presents the mean absolute error and the standard deviation between the estimated and ground-truth normal-mode amplitudes along with the angular and shift distances. In both test cases, the three distinct synthetic groups of subtomograms are correctly separated, taking into account the extreme noise level. The results show a less accurate alignment in the second case, which is expected since, in that case, the atomic structure was used to generate the dataset and the pseudoatomic structure was used as the reference model for the method to estimate the normal-mode amplitudes from this generated dataset. This is in contrast to the first test case where the same atomic structure was used to create the dataset and as the reference for the method to estimate the normal-mode amplitudes from this dataset. We found a RMSD of 0.55 Å and 0.60 Å corresponding to a combined displacement along modes 7, 8 and 9 with the mean absolute errors in Table 1 for the tests with atomic and pseudoatomic structures respectively. Similarly, we found a RMSD of 0.94 Å and 1.06 Å corresponding a combined displacement along modes 7, 8 and 9 with the sum of the mean and standard deviation of the absolute errors in Table 1 for the tests with atomic and pseudoatomic structures respectively. Hence, the error range is significantly inferior to the half range of the synthesized motion (6.95 Å) and the pixel size used to create the data (2.2 Å). Figure 7 shows grouping and averaging the subtomograms in the first test type (atomic reference). We compared the obtained subtomogram averages with the corresponding ground-truth volumes (density maps from ground-truth deformed models, without noise and missing wedge, used for synthesizing noisy and CTF-affected tilt-series from which subtomograms were obtained by 3D reconstruction). The visual comparison shows no significant difference between them. The resolutions calculated using Fourier shell correlation - FSC (threshold value of 0.143), after applying onto the subtomogram average a large spherical mask (radius of 28 voxels) with smooth edges (Gaussian smoothing with Gaussian standard deviation of 5 voxels), are 5.30 Å, 5.35 Å, and 5.74 Å for the three subtomogram averages shown from left to right in Figure 7, respectively. Without masking subtomogram averages, these resolutions are 6.31 Å, 6.10 Å, and 6.43 Å, respectively. As a basis for comparison, we provide the resolutions of three individual subtomograms arbitrarily chosen from the three corresponding subtomogram averaging groups. The resolutions of individual subtomograms with masking (the mask already described) are 10.58 Å, 10.57 Å, and 12.75 Å. The resolutions of the same subtomograms without masking are 12.67 Å, 13.55 Å, and 14.72 Å. These results show that a twice better resolution is obtained after averaging only about 300 individual subtomograms per group (Fig. 7).

3.3 Synthetic continuous-type conformational variability

Similarly to the previous experiment, our goal in this experiment is to find a solution for the inverse problem of finding the conformation of the structure in each subtomogram using the combined elastic and rigid-body alignment of a reference model with the subtomograms in the ”Continuous” dataset. We used the same two reference models as in the previous experiment to estimate the normal-mode amplitudes:

an atomic structure (chain A of PDB:4AKE) and a density map from this atomic structure. Also, as in the previous experiment, we used three modes for both tests (atomic or pseudoatomic modes 7, 8 and 9). Figure 8 presents the estimated amplitudes of modes 7 and 8 (the estimated amplitude of mode 9 is close to 0 and is not shown in the plots). Table 2 shows the mean absolute error and the standard deviation between the estimated and ground-truth normal-mode amplitudes along with the angular and shift distances. In both test cases, a linear relationship between the estimated amplitudes of normal modes 7 and 8 is clearly distinguishable, which is close to the identity relationship between the ground-truth amplitudes taking into account strong noise present in the data. As in the previous experiment, the results show a slightly less accurate alignment in the second test type (pseudoatomic reference) for the same aforementioned reason. We found a RMSD of 0.66 Å and 0.75 Å corresponding to a combined displacement along modes 7, 8 and 9 with the mean absolute errors in Table 2 for the tests with atomic and pseudoatomic structures respectively. Similarly, we found a RMSD of 1.09 Å and 1.21 Å corresponding a combined displacement along modes 7, 8 and 9 with the sum of the mean and standard deviation of the absolute errors in Table 2 for the tests with atomic and pseudoatomic structures respectively. Hence, the error range is significantly inferior to the half range of the synthesized motion (6.95 Å) and the pixel size used to create the data (2.2 Å). Figure 9 shows grouping and averaging of subtomograms in this experiment, with 8 subtomogram averages calculated along the distribution of the points for the first test type (atomic reference). The subtomogram averages show different conformations of adenylate kinase chain A. Note that the noise contained in the individual subtomograms (SNR = 0.01, Figure 5(B)) was reduced through subtomogram averaging (Figure 9). Additional experiments, for other noise levels in input subtomograms, can be found in the supplementary material Figure S1 and Table S1.

3.4 Experimental cryo-ET data: nucleosomes *in situ*

We applied our method on a dataset comprising 650 *in situ* subtomograms of nucleosomes collected from a cell of a Drosophila embryonic brain, whose conformational variability was detected but not fully explored in a previous work Eltsov et al. (2018). The subtomograms had the size of 64^3 voxels and the voxel size of 4.4 $Å^3$. A density map obtained with classical subtomogram averaging (without taking into account conformational heterogeneity) was used as the reference density map for HEMNMA-3D (Figure 10(C)). The resolution of this reference density map is around 2 nm (as determined by Fourier Shell Correlation between the reference density map and the density map from the atomic nucleosome structure PDB:3w98 Iwasaki et al. (2013) shown in Figure 10(B)). For more information on how this reference density map (global initial subtomogram average) was obtained, please see Section 1 of the supplementary material (Nucleosome data preparation and acquisition). This reference density map was converted into pseudoatoms (1368 pseudoatoms for the pseudoatom radius of 0.5 voxels and the target approximation error of 5%) and normal mode analysis of the obtained reference pseudoatomic structure was performed. The combined elastic and rigid-body alignment was performed using the pseudoatomic structure and a set of its six low-frequency high-collectivity normal modes. We selected modes 7-11 and mode 16, as described above and in our previous works (Jin et al. (2014); Sorzano et al. (2014); Harastani et al. (2020)). Modes 7-11 were selected as being the five lowest-frequency non-rigid-body modes with

collectivities above 0.5. They include the mode related to gaping motion (mode 7) and the mode related to breathing motion (mode 9), which have been described in previous nucleosome studies (Eltsov et al. (2018); Zlatanova et al. (2009)). Mode 16 was selected as being related to a motion that could be potentially interesting but it is more complex (a slightly higher frequency motion), potentially including gaping- and breathing-like motions. The normal-mode amplitudes estimated through the alignment (six normal mode amplitudes per subtomogram) were then projected onto a 2D space of conformations using PCA. The space of conformations is presented in Figure 10. Recall that each of the points represents a subtomogram, and close points represent similar conformations. By inspecting this conformational space, we identified four densest regions with 70, 183, 74, and 64 points from left to right in Figure 10(D). Following this analysis, we grouped the subtomograms in each of these four regions and averaged them. Before averaging, we filled in the missing-wedge Fourier space region of the individual subtomograms with the corresponding region of the global average computed from all subtomograms (please note that this global average was computed after aligning subtomograms using the rigid-body alignment parameters found along with the 3D-to-3D elastic alignment by HEMNMA-3D, which is a similar density map to the initial global average map shown in Figure 10(C) as both density maps result from averaging conformational heterogeneous subtomograms). The displacement of the reference pseudoatomic structure (converted into a density map) along two directions D1 and D2 in the space of conformations is shown in Figure 11 and in supplementary material Movie S1 and Movie S2. The significant difference between the four group averages (Figure 10(D)) and the reference density map (Figure 10(C)) as well as the motion observed along the two directions D1 and D2 (Figure 11, supplementary Movie S1 and Movie S2) can be described, mainly in terms of opening the nucleosome by increasing the distance between the two gyres of the DNA superhelix. This result consents the previous findings, observed but not fully explored in a previous study (manual analysis) of the nucleosome conformational variability Eltsov et al. (2018). The group averages are also compared with the atomic nucleosome structure PDB:3w98 in supplementary material Figure S2.

4 DISCUSSION AND CONCLUSIONS

This article presents HEMNMA-3D, the first cryo-ET subtomogram data analysis approach to study continuous conformational variability of biomolecular complexes, which maps a set of subtomograms into a space of conformations using a reference model and its normal modes. The conformational space permits i) grouping (and averaging) subtomograms with similar conformations and revealing hidden conformations and ii) recording animated displacements of the reference model along the densest regions of the space, along trajectories identified by curve fitting of the data in these regions. These HEMNMA-3D outputs could be valuable to cryo-ET studies of molecular mechanisms involved in conformational changes of complexes *in vitro* and *in situ*. HEMNMA-3D is thoroughly tested using synthetic subtomograms and applied to a cryo-ET experimental dataset (nucleosome subtomograms recorded *in situ* in Drosophila interphase nucleus). It provides promising results coherent with previous findings. An open-source software with a graphical user interface is provided for this method with a C++ backend and a Message Passing Interface parallelization scheme.

Both HEMNMA-3D and NMA-based flexible fitting are NMA applications concerned with estimating the molecular conformation in density maps based on an atomic or pseudoatomic reference. However, the purpose of HEMNMA-3D is different from that of the classical NMA-based fitting methods. Classical NMA-based fitting methods aim at determining an atomic representation of an EM density map, which is done by flexible fitting of a given atomic structure into that EM map. The purpose of HEMNMA-3D is to get a low-dimensional representation of the heterogeneity of a given set of EM maps such as subtomograms. Such low-dimensional representations do not require pushing the limits of the fitting accuracy as in the case of classical flexible fitting of atomic structures into EM maps, which also prevents overfitting. Besides, HEMNMA-3D performs a rigid-body alignment simultaneously with the flexible alignment, which accounts for the missing wedge of the low-SNR subtomograms, whereas classical NMA-based fitting methods typically use high-SNR average consensus EM maps reconstructed from single particle cryo-EM images without missing wedge.

The uniform random distribution of conformations was used in our experiments with synthetic data to show that HEMNMA-3D finds the correct values (within an acceptable error) for any conformation, rotation and translation, and that it does not yield wrong biased solutions (e.g., systematic alignment errors such as a wrong biased alignment to one or the other conformation and systematic rotational or translational errors). Taking into account the independent analysis of each individual subtomogram, HEMNMA-3D should be able to recover any other conformational distribution, with similar errors to those obtained with the conformational distributions used in this article.

The tests with synthetic data using a pseudoatomic structure from a simulated density map as a reference were used to demonstrate the ability of the method to retrieve the ground truth conformations with a comparable accuracy to the accuracy achieved using an atomic reference despite that i) the pseudoatomic reference was not used to synthesize the data (the datasets were synthesized using the atomic reference), ii) pseudoatomic coordinates unlikely coincide with atomic coordinates (pseudoatomic coordinates are obtained through volume-to-pseudoatoms conversion, which does not use any prior information about atoms), and iii) the method for calculating normal modes is different for a pseudoatomic reference (Cartesian method) and an atomic reference (RTB method). In experimental cases, one could obtain pseudoatoms from a density map of higher resolution than the data itself, if such density map is already available (e.g., an EMDB map obtained by high-resolution single particle cryo-EM reconstruction) or if it can be simulated from an available atomic structure (as in our synthetic data experiments). However, a preferred choice for the reference density map should be a density map from the data itself, which can be obtained by classical subtomogram averaging (without taking into account conformational heterogeneity), as was the case in the nucleosome experiments shown in this article.

The synthetic subtomogram datasets used in this work do not account for crowded molecular environments, radiation dose accumulation during tilting, and differences in CTF defocus over the tilted planes. Nevertheless, the synthesized subtomograms used here were challenging, as containing a small number of voxels and as being obtained by 3D reconstruction from synthetic tilt images affected by strong noise and CTF, which altogether lowered the resolution of the reconstructed subtomograms. The

subtomograms were additionally affected by the missing wedge artifacts. Despite such difficult conditions, HEMNMA-3D finds the correct conformational, orientational, and translational parameters, which suggests that it can be useful in practice, and this usefulness was here demonstrated with experimental nucleosome subtomograms.

In experimental cases such as the nucleosome study shown in this article, the number of used normal modes will always be smaller than the actual number of normal modes (the entire set of modes is too large to be included in our calculations as this would require too long computing times). Therefore, the conformational landscape will always be an approximation of the actual conformational landscape. Small sets of selected potentially relevant modes have been shown to produce good approximation of the actual conformational landscape of the nucleosome studied here by HEMNMA-3D as well as of other complexes studied by HEMNMA. In some cases, a single normal mode could be enough such as in the case of 70S ribosomes, where the normal mode describing the rotation between the two subunits 30S and 50S was used to analyze conformational and compositional variability of EF-G bound and unbound 70S ribosome cryo-EM dataset Jin et al. (2014). HEMNMA-3D uses the same software for NMA and the same numerical optimizer (to estimate the amplitudes of normal modes iteratively) as HEMNMA, and it should thus have similar performance as HEMNMA regarding the determination of the conformational landscape using a smaller number of normal modes. HEMNMA-3D software (ConinuousFlex plugin of Scipion V3.0) helps the user decide which normal modes to select, based on the lowest-frequency highest-collectivity criterion and including or not a prior knowledge about the conformational transitions, as is the case with HEMNMA software Harastani et al. (2020).

HEMNMA-3D can deal with larger sets of subtomograms than those shown in this article. Each subtomogram is analyzed independantly of other subtomograms, meaning on a separate computing thread. The same computing time per subtomogram is required for larger and smaller datasets of the same molecular complex and the time required to process the entire dataset varies with the size of the dataset.

5 SUPPLEMENTARY MATERIAL

The Supplementary Material is provided in the form of one PDF file and two movies.

6 FUNDING

We acknowledge the support of the French National Research Agency—ANR (ANR-20-CE11-0020-01 to A.L., ANR-20-CE11-0020-02 to M.E., ANR-20-CE11-0020-03 and ANR-19-CE11-0008-01 to S.J.); German Research Foundation (DFG EL 861/1 to M.E.); the French Microscopy Society (2019 Master internship grant to M.H.); the Sorbonne University (2019 ”Interface pour le Vivant” PhD scholarship grant to M.H.); and the access to the HPC resources of CINES and IDRIS granted by GENCI (2019-A0070710998, AP010712190, AD011012188 to S.J.).

7 ACKNOWLEDGMENTS

MH and SJ thank the Scipion team (CNB-CSIC, Madrid, Spain) for the help with porting the HEMNMA-3D code from Scipion V2.0 to Scipion V3.0, and Ilyes Hamitouche and Rémi Vuillemot (IMPMC - UMR 7590, Paris, France) for constructive discussions on HEMNMA-3D development. ME thanks Diana Grewe for cryo-sectioning, and Achilleas Frangakis (BMLS, Frankfurt, Germany) for providing scripts for subtomogram averaging, and the access to computational resources.

8 AUTHOR CONTRIBUTIONS

MH and SJ designed the method and the experiments with synthetic data. All authors designed the experiments with nucleosome data and contributed to the manuscript preparation. ME obtained nucleosome subtomograms and their initial average used as the reference for the method in nucleosome experiments. MH implemented the method and performed all experiments, under the supervision of SJ. ME and AL participated in the results interpretation. MH wrote the first draft of the article manuscript, which was finalized by SJ with input from all authors. All authors have read and approved the final version of the manuscript.

9 DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary files. HEMNMA-3D software code is publicly available on Github (https://github.com/scipion-em/scipion-em-continuousflex) and is also part of the open-source ContinuousFlex plugin of Scipion V3.0. The nucleosome data used in this article have been deposited in EMPIAR and EMDB databases under the accession codes EMPIAR-10679 and EMD-12699, respectively. Further inquiries can be directed to the corresponding author.

10 CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

Abeyrathne, P. D., San Koh, C., Grant, T., Grigorieff, N., and Korostelev, A. A. (2016). Ensemble cryo-em uncovers inchworm-like translocation of a viral ires through the ribosome. *Elife* 5, e14874

Albert, S., Schaffer, M., Beck, F., Mosalaganti, S., Asano, S., Thomas, H. F., et al. (2017). Proteasomes tether to two distinct sites at the nuclear pore complex. *Proceedings of the National Academy of Sciences* 114, 13726–13731

Andén, J. and Singer, A. (2018). Structural variability from noisy tomographic projections. *SIAM Journal on Imaging Sciences* 11, 1441–1492

Banerjee, S., Bartesaghi, A., Merk, A., Rao, P., Bulfer, S. L., Yan, Y., et al. (2016). 2.3 å resolution cryo-em structure of human p97 and mechanism of allosteric inhibition. *Science* 351, 871–875

559 Berghen, F. V. and Bersini, H. (2005). Condor, a new parallel, constrained extension of powell's uobyqa
560 algorithm: Experimental results and comparison with the dfo algorithm. *Journal of computational and*
561 *applied mathematics* 181, 157–175
562 Bharat, T. A. and Scheres, S. H. (2016). Resolving macromolecular structures from electron cryo-
563 tomography data using subtomogram averaging in relion. *Nature protocols* 11, 2054–2065
564 Böck, D., Medeiros, J. M., Tsao, H.-F., Penz, T., Weiss, G. L., Aistleitner, K., et al. (2017). In situ
565 architecture, function, and evolution of a contractile injection system. *Science* 357, 713–717
566 Brüschweiler, R. (1995). Collective protein dynamics and nuclear spin relaxation. *The Journal of chemical*
567 *physics* 102, 3396–3403
568 Bykov, Y. S., Schaffer, M., Dodonova, S. O., Albert, S., Plitzko, J. M., Baumeister, W., et al. (2017). The
569 structure of the copi coat determined within the cell. *Elife* 6, e32493
570 Castaño-Díez, D. and Zanetti, G. (2019). In situ structure determination by subtomogram averaging.
571 *Current opinion in structural biology* 58, 68–75
572 Chen, Y., Pfeffer, S., Fernández, J. J., Sorzano, C. O. S., and Förster, F. (2014). Autofocused 3d
573 classification of cryoelectron subtomograms. *Structure* 22, 1528–1537
574 Chen, Y., Pfeffer, S., Hrabe, T., Schuller, J. M., and Förster, F. (2013). Fast and accurate reference-free
575 alignment of subtomograms. *Journal of Structural Biology* 182, 235–245
576 Dashti, A., Schwander, P., Langlois, R., Fung, R., Li, W., Hosseinizadeh, A., et al. (2014). Trajectories
577 of the ribosome as a brownian nanomachine. *Proceedings of the National Academy of Sciences* 111,
578 17492–17497
579 Davies, K. M., Blum, T. B., and Kühlbrandt, W. (2018). Conserved in situ arrangement of complex i and
580 iii2 in mitochondrial respiratory chain supercomplexes of mammals, yeast, and plants. *Proceedings of*
581 *the National Academy of Sciences* 115, 3024–3029
582 De la Rosa-Trevín, J., Quintana, A., Del Cano, L., Zaldivar, A., Foche, I., Gutierrez, J., et al. (2016). Scipion:
583 A software framework toward integration, reproducibility and validation in 3d electron microscopy.
584 *Journal of structural biology* 195, 93–99
585 Delarue, M. and Dumas, P. (2004). On the use of low-frequency normal modes to enforce collective
586 movements in refining macromolecular structural models. *Proceedings of the National Academy of*
587 *Sciences* 101, 6957–6962
588 Durand, P., Trinquier, G., and Sanejouand, Y.-H. (1994). A new approach for determining low-frequency
589 normal modes in macromolecules. *Biopolymers: Original Research on Biomolecules* 34, 759–771
590 Elad, N., Clare, D. K., Saibil, H. R., and Orlova, E. V. (2008). Detection and separation of heterogeneity in
591 molecular complexes by statistical analysis of their two-dimensional projections. *Journal of structural*
592 *biology* 162, 108–120
593 Eltsov, M., Grewe, D., Lemercier, N., Frangakis, A., Livolant, F., and Leforestier, A. (2018). Nucleosome
594 conformational variability in solution and in interphase nuclei evidenced by cryo-electron microscopy of
595 vitreous sections. *Nucleic acids research* 46, 9189–9200
596 Förster, F., Pruggnaller, S., Seybert, A., and Frangakis, A. S. (2008). Classification of cryo-electron
597 sub-tomograms using constrained correlation. *Journal of structural biology* 161, 276–286

598 Frank, J. and Ourmazd, A. (2016). Continuous changes in structure mapped by manifold embedding of
599 single-particle data in cryo-em. *Methods* 100, 61–67
600 Fu, J., Gao, H., and Frank, J. (2007). Unsupervised classification of single particles by cluster tracking in
601 multi-dimensional space. *Journal of structural biology* 157, 226–239
602 Galaz-Montoya, J. G., Flanagan, J., Schmid, M. F., and Ludtke, S. J. (2015). Single particle tomography in
603 eman2. *Journal of structural biology* 190, 279–290
604 Guo, Q., Lehmer, C., Martínez-Sánchez, A., Rudack, T., Beck, F., Hartmann, H., et al. (2018). In situ
605 structure of neuronal c9orf72 poly-ga aggregates reveals proteasome recruitment. *Cell* 172, 696–705
606 Harastani, M., Sorzano, C. O. S., and Jonić, S. (2020). Hybrid electron microscopy normal mode analysis
607 with scipion. *Protein Science* 29, 223–236
608 Haselbach, D., Komarov, I., Agafonov, D. E., Hartmuth, K., Graf, B., Dybkov, O., et al. (2018). Structure
609 and conformational dynamics of the human spliceosomal bact complex. *Cell* 172, 454 – 464.e11
610 Himes, B. A. and Zhang, P. (2018). emclarity: software for high-resolution cryo-electron tomography and
611 subtomogram averaging. *Nature methods* 15, 955–961
612 Hutchings, J., Stancheva, V., Miller, E. A., and Zanetti, G. (2018). Subtomogram averaging of copii
613 assemblies reveals how coat organization dictates membrane shape. *Nature communications* 9, 1–8
614 Iwasaki, W., Miya, Y., Horikoshi, N., Osakabe, A., Taguchi, H., Tachiwana, H., et al. (2013). Contribution
615 of histone n-terminal tails to the structure and stability of nucleosomes. *FEBS open bio* 3, 363–369
616 Jin, Q., Sorzano, C. O. S., De La Rosa-Trevín, J. M., Bilbao-Castro, J. R., Núñez-Ramírez, R., Llorca, O.,
617 et al. (2014). Iterative elastic 3d-to-2d alignment method using normal modes for studying structural
618 dynamics of large macromolecular complexes. *Structure* 22, 496–506
619 Jonić, S. (2017). Computational methods for analyzing conformational variability of macromolecular
620 complexes from cryo-electron microscopy images. *Current Opinion in Structural Biology* 43, 114 – 121
621 Jonić, S. and Sorzano, C. (2016). Coarse-graining of volumes for modeling of structure and dynamics in
622 electron microscopy: Algorithm to automatically control accuracy of approximation. *IEEE Journal of*
623 *Selected Topics in Signal Processing* 10, 161–173
624 Kaplan, M., Ghosal, D., Subramanian, P., Oikonomou, C. M., Kjaer, A., Pirbadian, S., et al. (2019). The
625 presence and absence of periplasmic rings in bacterial flagellar motors correlates with stator type. *Elife*
626 8, e43487
627 Katsevich, E., Katsevich, A., and Singer, A. (2015). Covariance matrix estimation for the cryo-em
628 heterogeneity problem. *SIAM J Imaging Sci* 8, 126–185
629 Kovacs, J. A., Chacón, P., Cong, Y., Metwally, E., and Wriggers, W. (2003). Fast rotational matching of
630 rigid bodies by fast Fourier transform acceleration of five degrees of freedom. *Acta Crystallographica*
631 *Section D* 59, 1371–1376
632 Kovacs, J. A. and Wriggers, W. (2002). Fast rotational matching. *Acta crystallographica. Section D,*
633 *Biological crystallography* 58, 1282–1286
634 Kovtun, O., Leneva, N., Bykov, Y. S., Ariotti, N., Teasdale, R. D., Schaffer, M., et al. (2018). Structure of
635 the membrane-assembled retromer coat determined by cryo-electron tomography. *Nature* 561, 561–564

636 Kufareva, I. and Abagyan, R. (2011). Methods of protein structure comparison. In *Homology Modeling*
637 (Springer). 231–257
638 Leigh, K. E., Navarro, P. P., Scaramuzza, S., Chen, W., Zhang, Y., Castaño-Díez, D., et al. (2019).
639 Subtomogram averaging from cryo-electron tomograms. *Methods in cell biology* 152, 217–259
640 Lyumkis, D., Brilot, A. F., Theobald, D. L., and Grigorieff, N. (2013). Likelihood-based classification of
641 cryo-em images using frealign. *Journal of structural biology* 183, 377–388
642 Ma, J. (2005). Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular
643 complexes. *Structure* 13, 373–380
644 Mahalanobis, P. C. (1936). On the generalized distance in statistics (National Institute of Science of India)
645 Mahamid, J., Pfeffer, S., Schaffer, M., Villa, E., Danev, R., Cuellar, L. K., et al. (2016). Visualizing the
646 molecular sociology at the hela cell nuclear periphery. *Science* 351, 969–972
647 Mattei, S., Glass, B., Hagen, W. J. H., Kräusslich, H.-G., and Briggs, J. A. G. (2016). The structure and
648 flexibility of conical hiv-1 capsids determined within intact virions. *Science* 354, 1434–1437
649 Moebel, E. and Kervrann, C. (2020). A monte carlo framework for missing wedge restoration and noise
650 removal in cryo-electron tomography. *Journal of Structural Biology: X* 4, 100013
651 Mosalaganti, S., Kosinski, J., Albert, S., Schaffer, M., Strenkert, D., Salomé, P. A., et al. (2018). In situ
652 architecture of the algal nuclear pore complex. *Nature communications* 9, 1–8
653 Müller, C., Schlauderer, G., Reinstein, J., and Schulz, G. E. (1996). Adenylate kinase motions during
654 catalysis: an energetic counterweight balancing substrate binding. *Structure* 4, 147–156
655 Nogales-Cadenas, R., Jonic, S., Tama, F., Arteni, A., Tabas-Madrid, D., Vázquez, M., et al. (2013). 3dem
656 loupe: analysis of macromolecular dynamics using structures from electron microscopy. *Nucleic acids*
657 *research* 41, W363–W367
658 Park, D., Lara-Tejero, M., Waxham, M. N., Li, W., Hu, B., Galán, J. E., et al. (2018). Visualization of
659 the type iii secretion mediated salmonella–host cell interface using cryo-electron tomography. *Elife* 7,
660 e39514
661 Penczek, P. A., Frank, J., and Spahn, C. M. (2006). A method of focused classification, based on the
662 bootstrap 3d variance analysis, and its application to ef-g-dependent translocation. *Journal of structural*
663 *biology* 154, 184–194
664 Penczek, P. A., Kimmel, M., and Spahn, C. M. (2011). Identifying conformational states of macromolecules
665 by eigen-analysis of resampled cryo-em images. *Structure* 19, 1582–1590
666 Peng, L.-M., Ren, G., Dudarev, S., and Whelan, M. (1996). Robust parameterization of elastic and
667 absorptive electron atomic scattering factors. *Acta Crystallographica Section A: Foundations of*
668 *Crystallography* 52, 257–276
669 Pfeffer, S., Dudek, J., Schaffer, M., Ng, B. G., Albert, S., Plitzko, J. M., et al. (2017). Dissecting the
670 molecular organization of the translocon-associated protein complex. *Nature communications* 8, 1–9
671 Rapisarda, C., Cherrak, Y., Kooger, R., Schmidt, V., Pellarin, R., Logger, L., et al. (2019). In situ and
672 high-resolution cryo-em structure of a bacterial type vi secretion system membrane complex. *The EMBO*
673 *journal* 38, e100886

674 Riedel, C., Vasishtan, D., Siebert, C. A., Whittle, C., Lehmann, M. J., Mothes, W., et al. (2017). Native
675 structure of a retroviral envelope protein and its conformational change upon interaction with the target
676 cell. *Journal of structural biology* 197, 172–180
677 Scheres, S. H. (2012). Relion: implementation of a bayesian approach to cryo-em structure determination.
678 *Journal of structural biology* 180, 519–530
679 Scheres, S. H., Melero, R., Valle, M., and Carazo, J.-M. (2009). Averaging of electron subtomograms and
680 random conical tilt reconstructions through likelihood optimization. *Structure* 17, 1563–1572
681 Schur, F. K., Obr, M., Hagen, W. J., Wan, W., Jakobi, A. J., Kirkpatrick, J. M., et al. (2016). An atomic
682 model of hiv-1 capsid-sp1 reveals structures regulating assembly and maturation. *Science* 353, 506–508
683 Sorzano, C., Jonic, S., Núñez-Ramírez, R., Boisset, N., and Carazo, J. (2007). Fast, robust, and accurate
684 determination of transmission electron microscopy contrast transfer function. *Journal of Structural*
685 *Biology* 160, 249–262
686 Sorzano, C. O., de la Rosa Trevín, J., Otón, J., Vega, J., Cuenca, J., Zaldívar-Peraza, A., et al. (2013).
687 Semiautomatic, high-throughput, high-resolution protocol for three-dimensional reconstruction of single
688 particles in electron microscopy. In *Nanoimaging* (Springer). 171–193
689 Sorzano, C. O. S., Alvarez-Cabrera, A. L., Kazemi, M., Carazo, J. M., and Jonić, S. (2016). Structmap:
690 elastic distance analysis of electron microscopy maps for studying conformational changes. *Biophysical*
691 *journal* 110, 1753–1765
692 Sorzano, C. O. S., de La Rosa-Trevín, J. M., Tama, F., and Jonić, S. (2014). Hybrid electron microscopy
693 normal mode analysis graphical interface and protocol. *Journal of structural biology* 188, 134–141
694 Sorzano, C. O. S., Jiménez, A., Mota, J., Vilas, J. L., Maluenda, D., Martínez, M., et al. (2019). Survey
695 of the analysis of continuous conformational variability of biological macromolecules by electron
696 microscopy. *Acta Crystallographica Section F* 75, 19–32
697 Stölken, M., Beck, F., Haller, T., Hegerl, R., Gutsche, I., Carazo, J.-M., et al. (2011). Maximum likelihood
698 based classification of electron tomographic data. *Journal of structural biology* 173, 77–85
699 Suhre, K., Navaza, J., and Sanejouand, Y.-H. (2006). Norma: a tool for flexible fitting of high-resolution
700 protein structures into low-resolution electron-microscopy-derived density maps. *Acta crystallographica*
701 *section D: biological crystallography* 62, 1098–1100
702 Suhre, K. and Sanejouand, y.-h. (2004). Elnemo: A normal mode web server for protein movement analysis
703 and the generation of templates for molecular replacement. *Nucleic acids research* 32, W610–4
704 Tagare, H. D., Kucukelbir, A., Sigworth, F. J., Wang, H., and Rao, M. (2015). Directly reconstructing
705 principal components of heterogeneous particles from cryo-em images. *Journal of Structural Biology*
706 191, 245 – 262
707 Tama, F. and Brooks III, C. L. (2006). Symmetry, form, and shape: guiding principles for robustness in
708 macromolecular machines. *Annu. Rev. Biophys. Biomol. Struct.* 35, 115–133
709 Tama, F., Gadea, F. X., Marques, O., and Sanejouand, Y.-H. (2000). Building-block approach for
710 determining low-frequency normal modes of macromolecules. *Proteins: Structure, Function, and*
711 *Bioinformatics* 41, 1–7

712 Tama, F., Miyashita, O., and Brooks, C. L. (2004a). Flexible multi-scale fitting of atomic structures into
713 low-resolution electron density maps with elastic network normal mode analysis. *Journal of Molecular*
714 *Biology* 337, 985 – 999
715 Tama, F., Miyashita, O., and Brooks III, C. L. (2004b). Normal mode based flexible fitting of high-
716 resolution structure into low-resolution experimental data from cryo-em. *Journal of Structural Biology*
717 147, 315 – 326. Time-Resolved Imaging of Macromolecular Processes and Interactions
718 Tama, F. and Sanejouand, Y.-H. (2001). Conformational change of proteins arising from normal mode
719 calculations. *Protein engineering* 14, 1–6
720 Tama, F., Wriggers, W., and Brooks III, C. L. (2002). Exploring global distortions of biological
721 macromolecules and assemblies from low-resolution structural information and elastic network theory.
722 *Journal of molecular biology* 321, 297–305
723 Tirion, M. M. (1996). Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.
724 *Phys. Rev. Lett.* 77, 1905–1908
725 Wan, W. and Briggs, J. (2016). Chapter thirteen - cryo-electron tomography and subtomogram averaging.
726 In *The Resolution Revolution: Recent Advances In cryoEM*, ed. R. Crowther (Academic Press), vol. 579
727 of *Methods in Enzymology*. 329 – 367
728 Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S., et al. (2017). Structure and
729 assembly of the ebola virus nucleocapsid. *Nature* 551, 394–397
730 Wang, Y., Rader, A., Bahar, I., and Jernigan, R. L. (2004). Global ribosome motions revealed with elastic
731 network model. *Journal of structural biology* 147, 302–314
732 Xu, M., Beck, M., and Alber, F. (2012). High-throughput subtomogram alignment and classification by
733 fourier space constrained fast volumetric matching. *Journal of structural biology* 178, 152–164
734 Zhai, X., Lei, D., Zhang, M., Liu, J., Wu, H., Yu, Y., et al. (2020). Lottor: An algorithm for missing-wedge
735 correction of the low-tilt tomographic 3d reconstruction of a single-molecule structure. *Scientific reports*
736 10, 1–17
737 Zhang, L. and Ren, G. (2012). High-resolution single-molecule structure revealed by electron microscopy
738 and individual particle electron tomography. *Journal of Physical Chemistry & Biophysics* 2
739 Zhou, A., Rohou, A., Schep, D. G., Bason, J. V., Montgomery, M. G., Walker, J. E., et al. (2015). Structure
740 and conformational states of the bovine mitochondrial atp synthase by cryo-em. *Elife* 4, e10180
741 Zlatanova, J., Bishop, T. C., Victor, J., Jackson, V., and van Holde, K. V. (2009). The nucleosome family:
742 dynamic and growing. *Structure* 17, 160–71

Experiment		Mode 7		Mode 8		Mode 9		Anglular [deg]		Shifting [vox]		p-value	Samples
Ref	**Dataset**	**mean**	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**		
Atomic	”Discrete”	16.51	11.87	10.91	7.64	10.70	6.66	1.33	0.77	0.19	0.09	p >0.01	871/900
Volume	”Discrete”	17.70	13.26	11.90	10.13	12.29	8.03	1.33	0.81	0.21	0.10	p >0.01	870/900

Table 1. Mean absolute error and standard deviation between the estimated and ground-truth normal-mode amplitudes along with the angular and shift distances obtained with HEMNMA-3D and ”Discrete” synthetic dataset, using an atomic structure (Atomic) and simulated EM map (Volume) as input references. The data points below the p-value of 0.01 were excluded from the error evaluation based on the Mahalanobis distance measure (few data points differing significantly from the remaining observations, which would not be selected in real-case experiments as being too isolated and far from other points). The number of points used for the error computation is shown in the last column of the table (column Samples) and the region with the kept points (p-value $>$ 0.01) is shown in Figure 6.

Experiment		Mode 7		Mode 8		Mode 9		Angular [deg]		Shifting [vox]		p-value	Samples
Ref	**Dataset**	**mean**	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**		
Atomic	”Continuous”	20.12	11.30	12.78	11.24	12.74	7.71	1.31	0.79	0.19	0.10	P >0.001	960/1000
Volume	”Continuous”	21.94	12.59	14.03	9.92	15.68	10.04	1.34	0.80	0.21	0.10	P >0.001	957/1000

Table 2. Mean absolute error and standard deviation between the estimated and ground-truth normal-mode amplitudes along with the angular and shift distances obtained with HEMNMA-3D and ”Continuous” synthetic dataset, using an atomic structure (Atomic) and simulated EM map (Volume) as input references. The data points below the p-value of 0.001 were excluded from the error evaluation based on the Mahalanobis distance measure (few data points differing significantly from the remaining observations, which would not be selected in real-case experiments as being too isolated and far from other points). The number of points used for the error computation is shown in the last column of the table (column Samples) and the region where p-value $>$ 0.001 is shown in Figure 8.

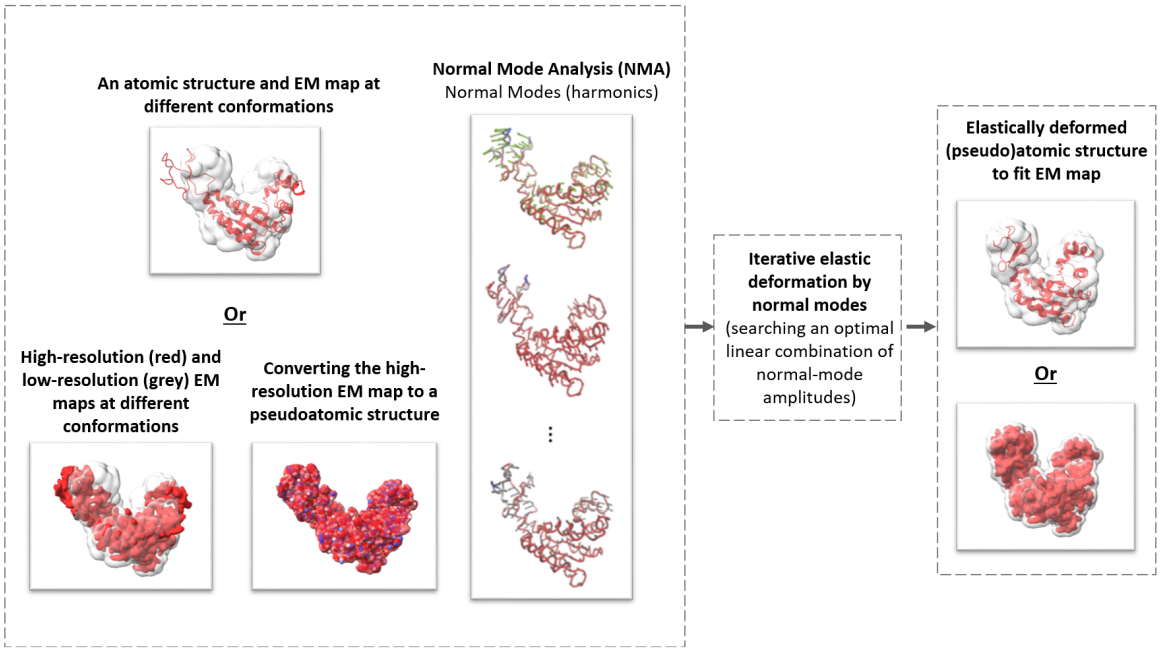

Figure 1. The general scheme of elastic deforming of a reference structure (atomic or pseudoatomic) using normal modes to fit a density map (e.g., an EM map or a subtomogram average).

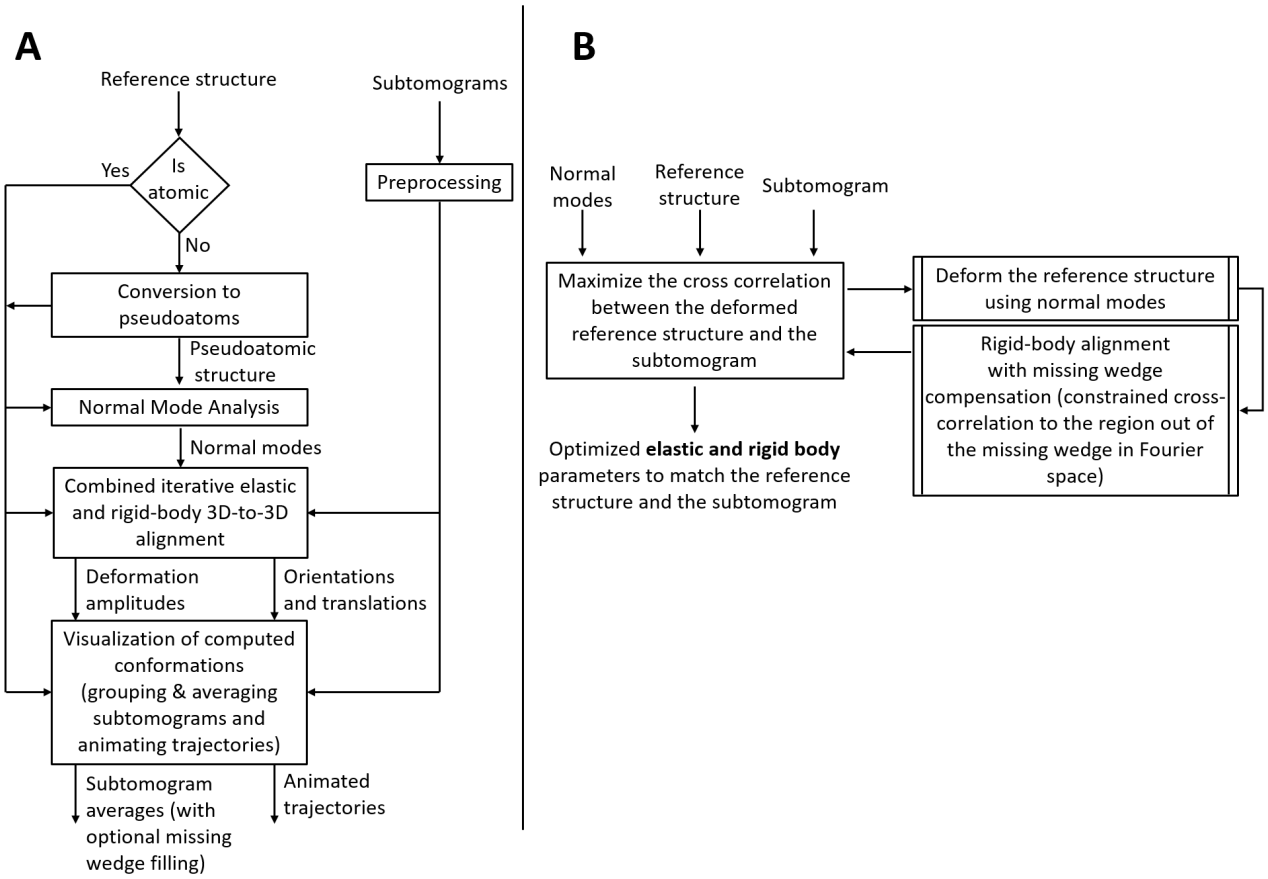

Figure 2. Flowchart of HEMNMA-3D. **(A)** Workflow. **(B)** Combined iterative elastic and rigid-body 3D-to-3D alignment step (the core module of HEMNMA-3D).

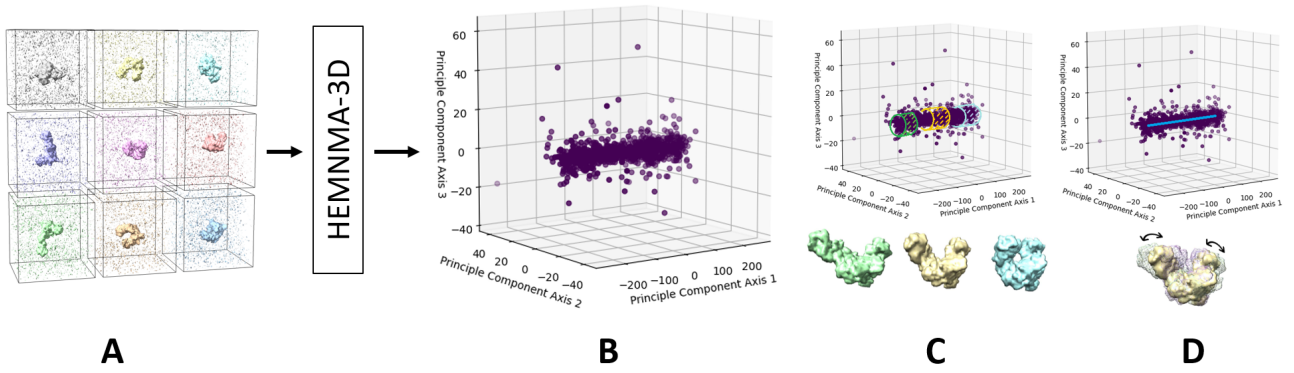

Figure 3. A graphical summary of the dataflow of HEMNMA-3D. **(A)** Input subtomograms containing the same biomolecule but at different orientations, positions and conformations (here represented with a low level of noise for illustration). **(B)** Input subtomograms projected onto a low-dimensional "space of conformations", describing and visualizing the biomolecular conformational variability contained in the subtomograms. **(C)** Grouping of close points (subtomograms with similar biomolecular conformations) and averaging of subtomograms in these groups. **(D)** Animating biomolecular motion along trajectories identified in the densest regions.

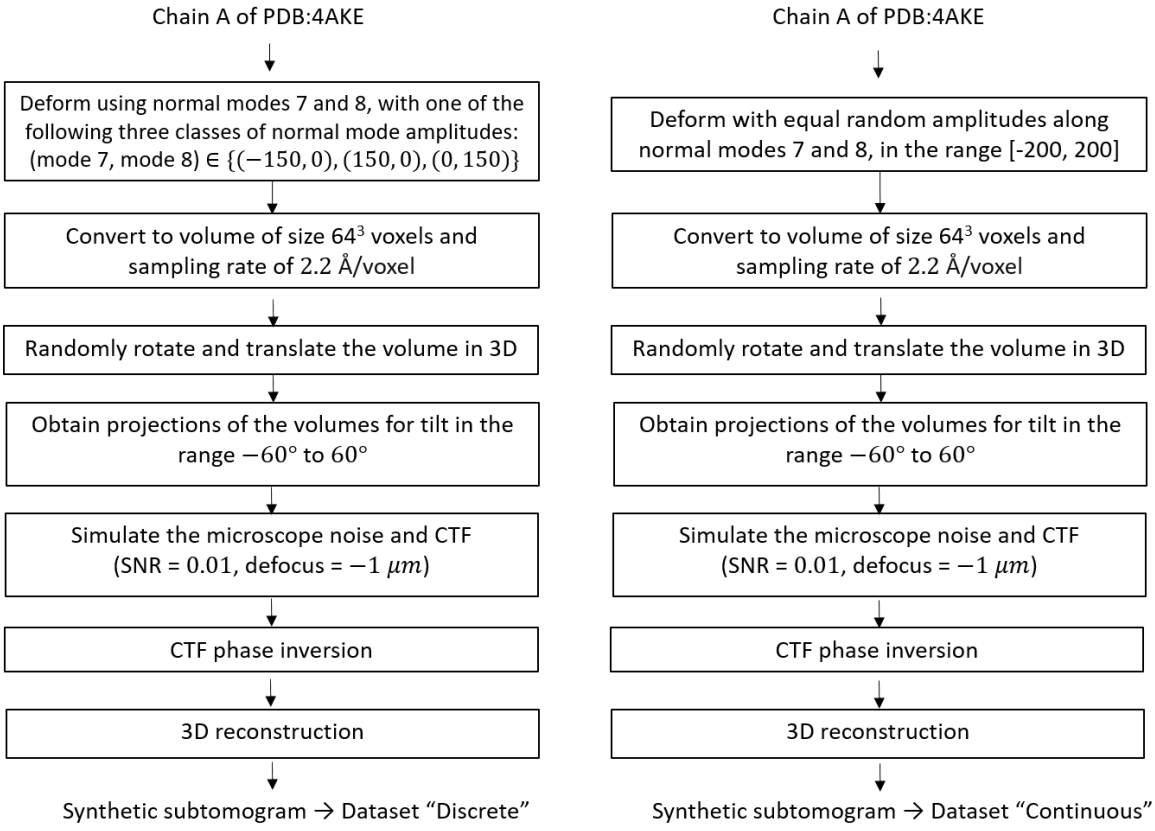

Figure 4. Flowcharts of synthesis of the datasets used for testing and validating HEMNMA-3D, namely "Discrete" dataset (left) and "Continuous" dataset (right).

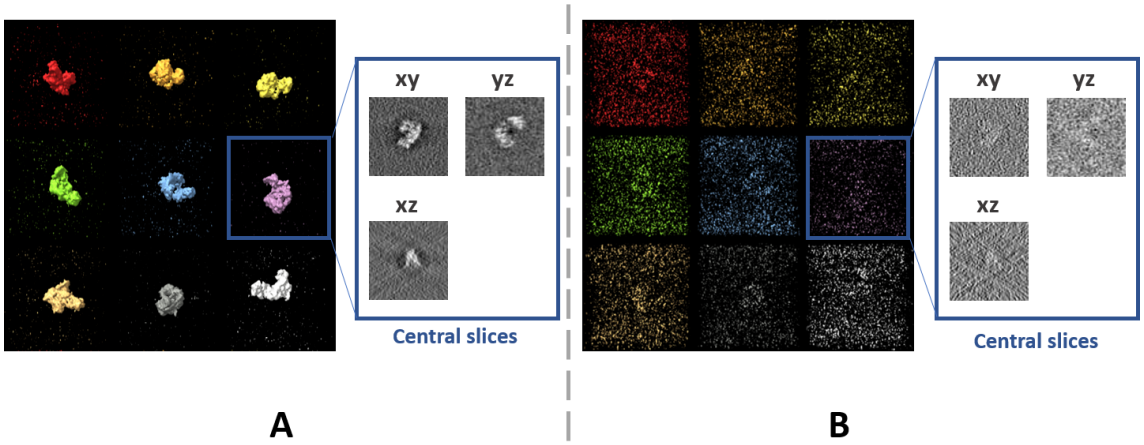

Figure 5. Examples of synthetic subtomograms containing the same molecule but at different orientations, positions and conformations, for two different noise levels. **(A)** low level of noise (SNR=0.5). **(B)** high level of noise (SNR=0.01).

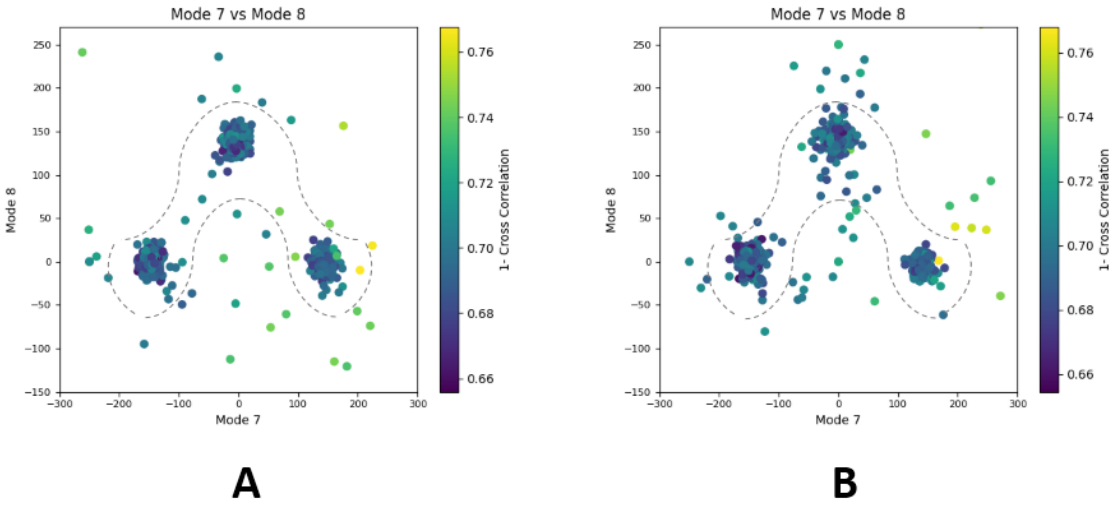

Figure 6. Plots showing the output of the 3D-to-3D elastic and rigid-body alignment module of HEMNMA-3D with "Discrete" dataset (synthetic subtomograms are simulating discrete conformational heterogeneity). **(A)** Use of the atomic structure (chain A of PDB:4AKE) and its normal modes to estimate the conformational parameters (normal-mode amplitudes) and rigid-body parameters (orientation and shift) of the molecules in the input synthetic subtomograms. **(B)** Use of a pseudoatomic structure (from a simulated density map) and its normal modes to estimate the conformational and rigid-body parameters of the molecules in the input synthetic subtomograms. The goal was the retrieval of the ground-truth relationship between the amplitudes along normal modes 7 and 8; ideally, all data should lay in one of the following three clusters of normal-mode amplitudes: (mode 7, mode 8) $\in$ {(-150, 0), (150, 0), (0, 150)}; each point in the plot represents a subtomogram and close points represent similar conformations. Note that the dashed curves enclose the data points where p-value $>$ 0.01 in Table 1. See the text for more details on this experiment.

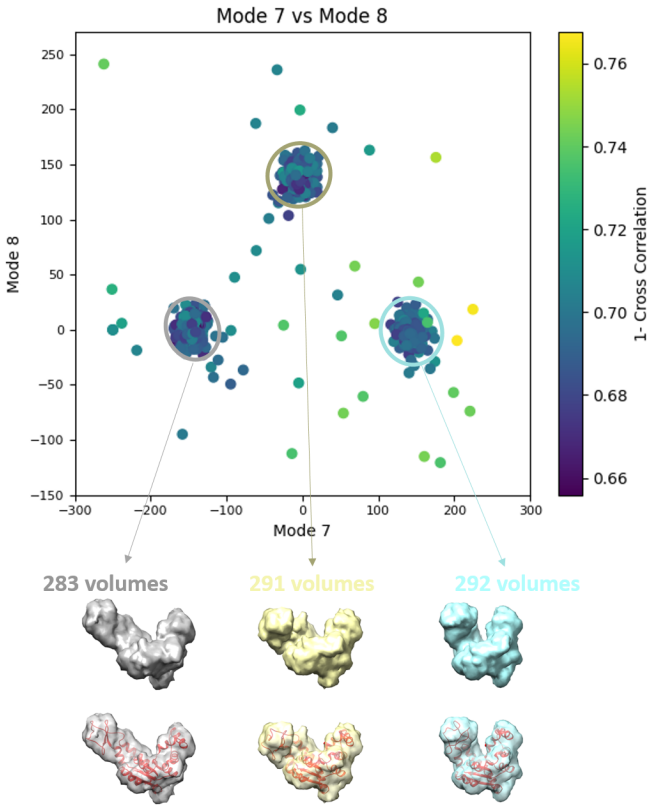

Figure 7. Averages of the three groups (enclosed by ellipses) of subtomograms identified from the output of the 3D-to-3D elastic and rigid-body alignment module of HEMNMA-3D with ”Discrete” dataset (shown in Figure 6(A)), using the atomic structure (chain A of PDB:4AKE) and its normal modes to estimate the conformational parameters (normal-mode amplitudes) and rigid-body parameters (orientation and shift) of the molecules in the input synthetic subtomograms. Subtomograms are represented by points and close points represent similar conformations. The numbers of volumes written above the shown subtomogram averages are the numbers of synthetic subtomograms used for computing these subtomogram averages (the numbers of points enclosed by the corresponding ellipses). On the bottom, the subtomogram averages are shown at 50% transparency along with the corresponding ground-truth deformed atomic structure (in red).

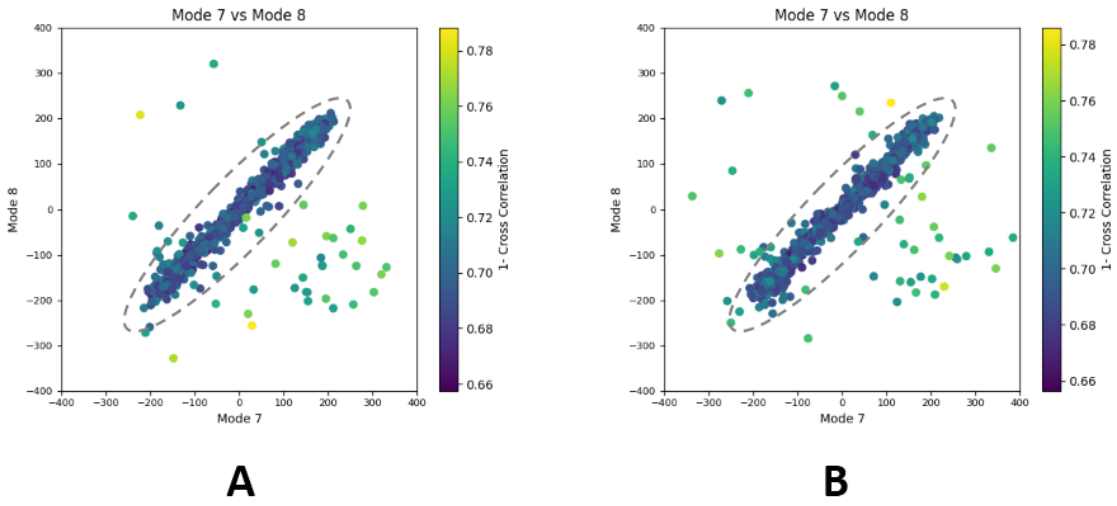

Figure 8. Plots showing the output of the 3D-to-3D elastic and rigid-body alignment module of HEMNMA-3D with ”Continuous” dataset (synthetic subtomograms are simulating continuous conformational heterogeneity). **(A)** Use of the atomic structure (chain A of PDB:4AKE) and its normal modes to estimate the conformational parameters (normal-mode amplitudes) and rigid-body parameters (orientation and shift) of the molecules in the input synthetic subtomograms. **(B)** Use of a pseudoatomic structure (from a simulated density map) and its normal modes to estimate the conformational and rigid-body parameters of the molecules in the input synthetic subtomograms. The goal was the retrieval of the ground-truth relationship between the amplitudes along normal modes 7 and 8 (ideally linear relationship, with equal amplitudes of normal modes 7 and 8); each point in the plot represents a subtomogram and close points represent similar conformations. Note that the dashed ellipses enclose the data points where p-value $>$ 0.001 in Table 2. See the text for more details on this experiment.

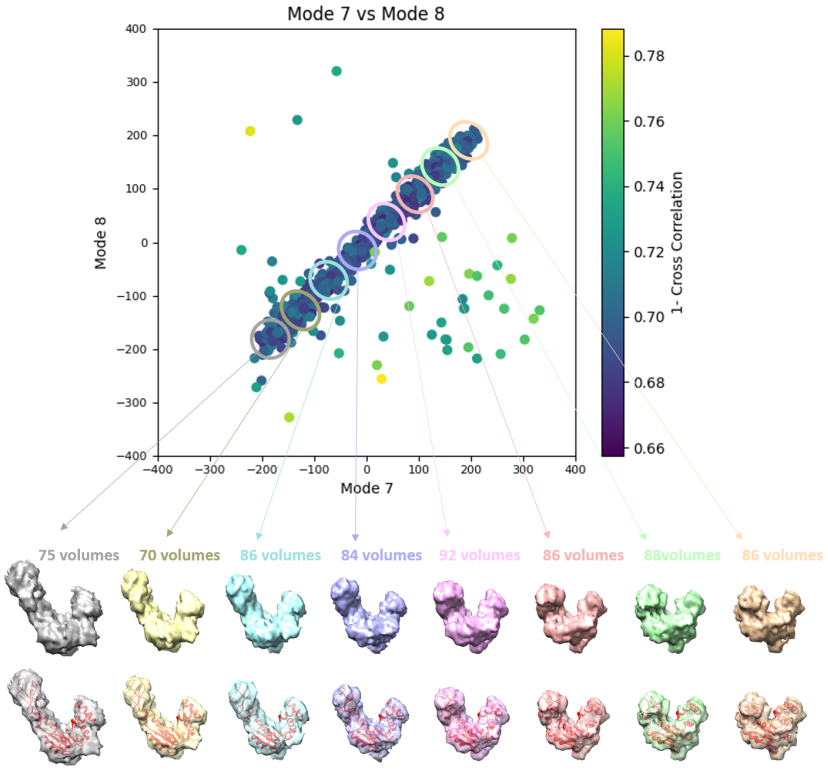

Figure 9. Averages of eight groups (enclosed by ellipses) of subtomograms identified from the output of the 3D-to-3D elastic and rigid-body alignment module of HEMNMA-3D with "Continuous" dataset (shown in Figure 8(A)), using the atomic structure (chain A of PDB:4AKE) and its normal modes to estimate the conformational parameters (normal-mode amplitudes) and rigid-body parameters (orientation and shift) of the molecules in the input synthetic subtomograms. Subtomograms are represented by points and close points represent similar conformations. The numbers of volumes written above the shown subtomogram averages are the numbers of synthetic subtomograms used for computing these subtomogram averages (the numbers of points enclosed by the corresponding ellipses). On the bottom, the subtomogram averages are shown at 50% transparency along with the corresponding theoretical centroid deformed atomic structure (in red).

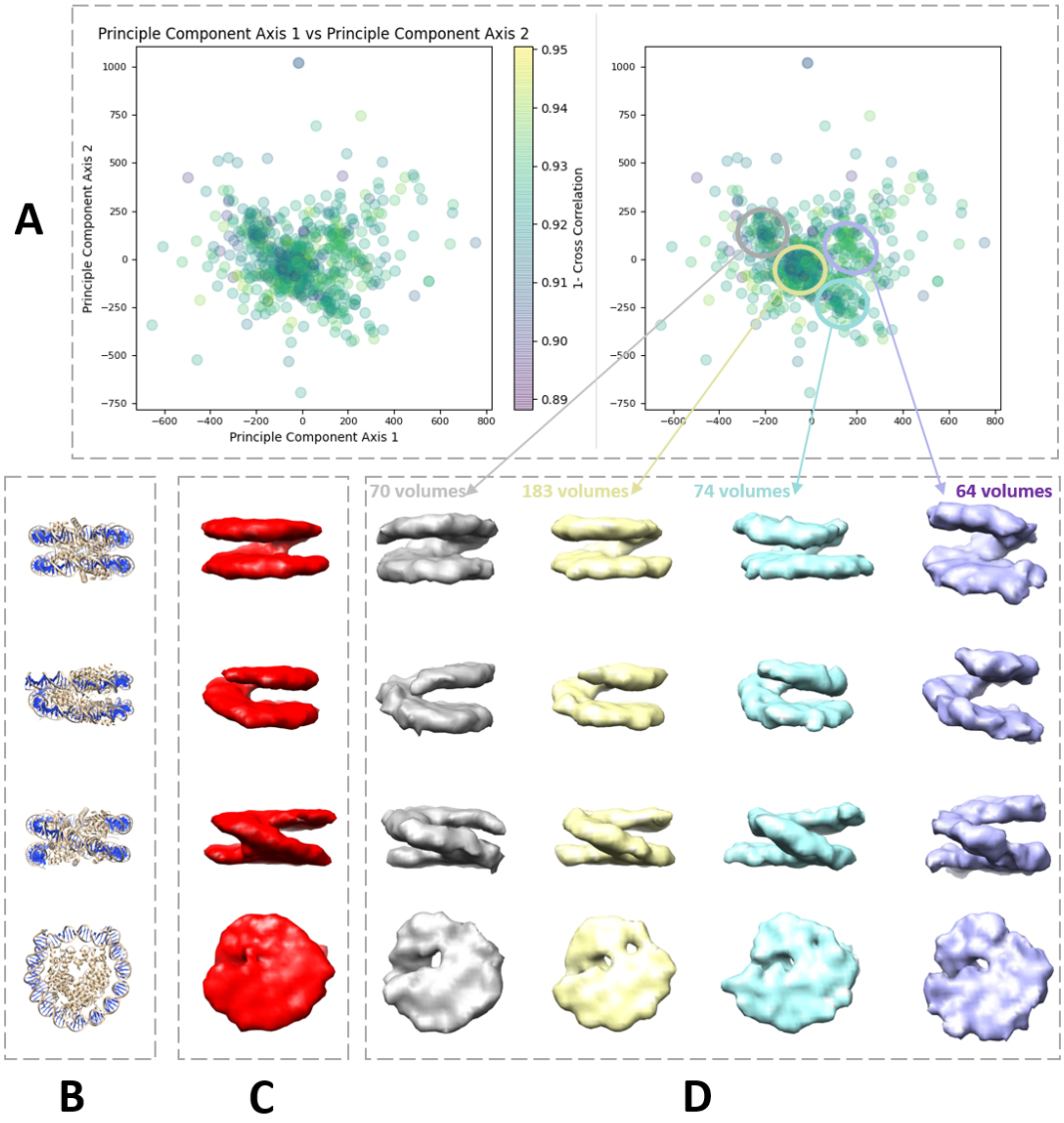

Figure 10. Illustration of HEMNMA-3D use with *in situ* cryo-ET nucleosome dataset. **(A)** Space of conformations resulting from projecting the estimated amplitudes of six normal modes onto a two-dimensional space using PCA. **(B)** Nucleosome atomic structure PDB:3w98, for comparison purposes. **(C)** Nucleosome subtomogram average (around 2 nm resolution) used as the input reference density map for HEMNMA-3D, obtained by classical subtomogram averaging, without taking into account conformational heterogeneity (for more information on how this global initial subtomogram average was obtained, see Section 1 of the supplementary material (Nucleosome data preparation and acquisition)). **(D)** Four subtomogram averages from four densest regions in the space of conformations (regions encircled with ellipses) showing different nucleosome conformations, mainly, different gap distances between the nucleosome gyres. The numbers of volumes written above the subtomogram averages shown in (D) are the numbers of *in situ* cryo-ET subtomograms used for computing these subtomogram averages (the numbers of points enclosed by the corresponding ellipses).

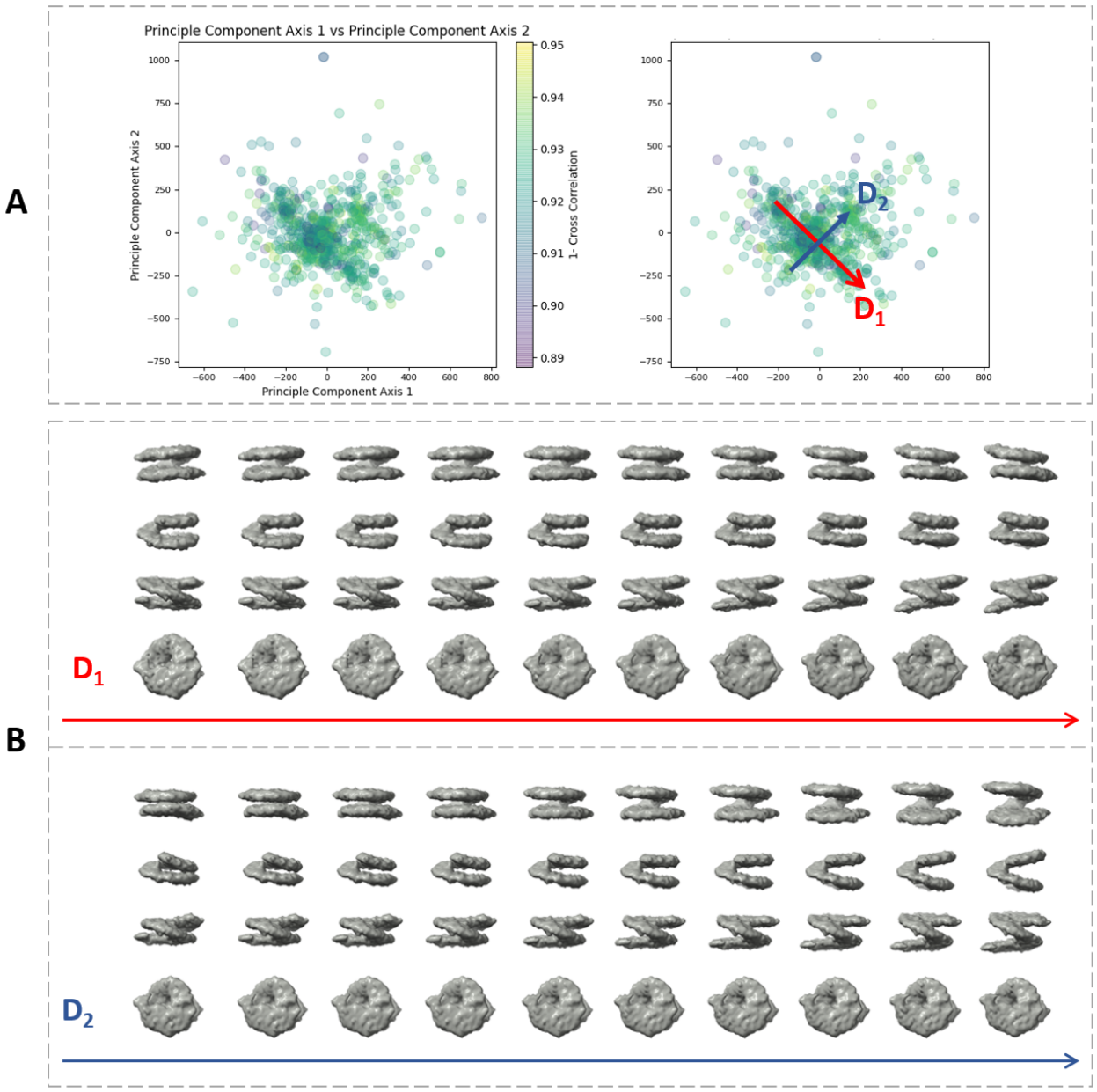

Figure 11. Displacement of the reference density map along two directions D1 and D2 in the space of conformations obtained (Figure 10) with HEMNMA-3D with *in situ* cryo-ET nucleosome dataset. **(A)** Space of conformations (left) as shown in Figure 10 and two directions D1 and D2 used to displace the reference density map (Figure 10(C)) in this space (right). **(B)** Displacement of the reference density map along the D1 and D2 directions (10 frames of the corresponding trajectory are shown row-wise.)

frontiers

This is the Supplementary Material of an article accepted for publication in Front. Mol. Biosci.
www.frontiersin.org/articles/10.3389/fmolb.2021.663121

HEMNMA-3D: Cryo Electron Tomography Method Based on Normal Mode Analysis to Study Continuous Conformational Variability of Macromolecular Complexes

Mohamad Harastani [1] **Mikhail Eltsov** [2] **Amélie Leforestier** [3] **Slavica Jonic** [1,*]

[1] *IMPMC - UMR 7590 CNRS, Sorbonne Université, Muséum National d'Histoire Naturelle, Paris, France*
[2] *Department of Integrated Structural Biology, Institute of Genetics and Molecular and Cellular Biology, Illkirch, France*
[3] *Laboratoire de Physique des Solides (LPS), UMR 8502 CNRS, Université Paris-Saclay, Orsay, France*

Correspondence*:
Slavica Jonic, IMPMC - UMR CNRS 7590, Sorbonne Univerisité, 4 Place Jussieu, 75005 Paris, France
slavica.jonic@upmc.fr

1 NUCLEOSOME DATA PREPARATION AND ACQUISITION

The Drosophila embryo cryo-sample preparation, vitreous sectioning, tilt series acquisition and tomogram reconstruction were performed as described in Eltsov et al. (2018). Briefly, Drosophila embryos (Bloomington Stock number 30564) were frozen by high-pressure freezing using an HPM 010 machine (ABRA Fluid AG). The vitreous sections were obtained using a 25° diamond knife (Diatome) and Leica FC6/UC6 ultramicrotome (Leica Microsystems). Sections were collected on C-flat CF-2/1 grids (Electron Microscopy Sciences) and transferred into a Titan Krios (FEI, Thermofischer, Eindhoven, The Netherlands) operated at 300 kV equipped with a GATAN GIF Quantum SE post-column energy filter and K2 Summit direct electron detector (Gatan, Pleasanton, USA). Tilt series were recorded using Serial EM software (Mastronarde (2005)) at a nominal magnification of 64000 × (2.2 Å/pixel), and a target defocus of -3.5 μm. The dose-symmetric recording scheme (Hagen et al. (2017)) was applied within an angular range from 60° to +60°, with a starting angle 0° and an angular increment of 2°. The electron dose was set to 1.5 $e^-/Å^2$ for individual tilt images, corresponding to the total dose of 91.5 $e^-/Å^2$ for the complete tilt series. A marker-less tilt series alignment was done in IMOD (Kremer et al. (1996)), three dimensional CTF correction and weighted backprojection with the voxel size of 4.4 Å were performed using EmSART script (Kunz and Frangakis (2014, 2017)) provided by Achilleas Frangakis. A slice of the experimental nucleosome tomographic data is shown in Figure S1. The reconstructed volumes were denoised using 3D non-linear anisotropic diffusion filter of Etomo of IMOD (k = 1, 15 iterations). Nucleosomes were manually

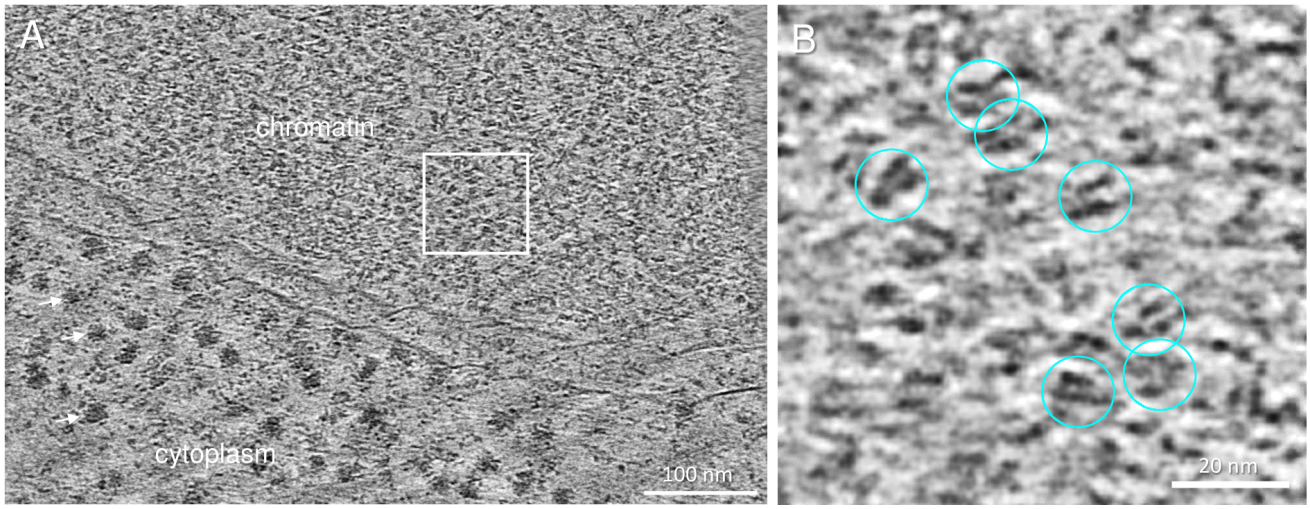

Figure S1. A slice of the experimental nucleosome tomographic data: (A) A 5-nm thick slice through a tomographic reconstruction showing an area of compact chromatin at a nuclear periphery (chromatin) that is easily distinguished from cytoplasm filled with ribosomes (arrows). (B) An enlargement of the chromatin area outlined with a white square in (A). Circles indicate positions of nucleosomes selected for subtomogram extraction.

picked in IMOD. Then, 64^3 voxel subtomograms (voxel size of 4.4 Å) were extracted from the original non-denoised volumes. To refine manually-picked nucleosome coordinates, subtomogram alignment and averaging were performed with SubTomogramAveraging script using a sum of the randomly rotated subtomograms as an initial reference. Alignment of subtomograms was performed in two steps. Initially, a bandpass filter was applied with a low cutoff frequency of 3 reciprocal-space pixels ($\approx$ 1/94 Å^{-1}), a high cutoff frequency of 8 reciprocal-space pixels ($\approx$ 1/35 Å^{-1}), and a Gaussian edge smoothing with a standard deviation of 3 reciprocal-space pixels. Ten iterations of an unconstrained rotational search (three rotational degrees of freedom) were performed with an angular sampling step of 10°, and a translational search (three translational degrees of freedom) was performed within a radius of 5 real-space pixels (22 Å). In the second step, a bandpass filter was applied with low and high cutoff frequencies of 3 reciprocal-space pixels ($\approx$ 1/94 Å^{-1}) and 15 reciprocal-space pixels ($\approx$ 1/19 Å^{-1}), respectively, and a Gaussian edge smoothing with a standard deviation of 3 reciprocal-space pixels. At that step, 20 iterations of the rotational search were performed with an angular sampling step of 2°, constrained to 20° around the orientation found in the previous step, and the translational search radius was reduced to 3 real-space pixels (13.2 Å). The cross-correlation between the last several iteration averages ($\approx$ 0.994) indicated the stabilization of the subtomogram alignment. A new set of subtomograms of the same dimensions and voxel size was extracted at the refined nucleosome positions and exported into HEMNMA-3D. The nucleosome data used in this article have been deposited in EMPIAR and EMDB databases under the accession codes EMPIAR-10679 and EMD-12699, respectively.

2 ADDITIONAL SYNTHETIC DATA TESTS WITH DIFFERENT NOISE LEVELS

Additional tests were performed on HEMNMA-3D using synthesized datasets at different noise levels of conformationally heterogeneous subtomograms that mimic continuous conformational variability. The noise levels were chosen as A) without noise, B) SNR = 0.4, C) SNR = 0.1, D) SNR = 0.04, E) SNR = 0.01 and F) SNR = 0.005. Each dataset comprises 200 synthetic subtomograms representing a continuum of conformations of the same PDB:4AKE structure. The flowchart for the data generation procedure is shown

in the article manuscript Figure 4. Here, the amplitude value of normal mode 7 is chosen randomly in the range {-300, 300} and the amplitude value mode 8 as half of the value of mode 7. More details on the synthesis of the data will be found in the main article manuscript Section 3.1. The goal in this experiment is to find a solution for the inverse problem of finding the conformation of the structure in each subtomogram using the combined elastic and rigid-body alignment of a reference model with the subtomograms in the different-in-noise datasets. Figure S2 presents the estimated amplitudes of modes 7 and 8 (the estimated amplitude of mode 9 is close to 0 and is not shown in the plots). Table S1 shows the mean absolute error and the standard deviation between the estimated and ground-truth normal-mode amplitudes.

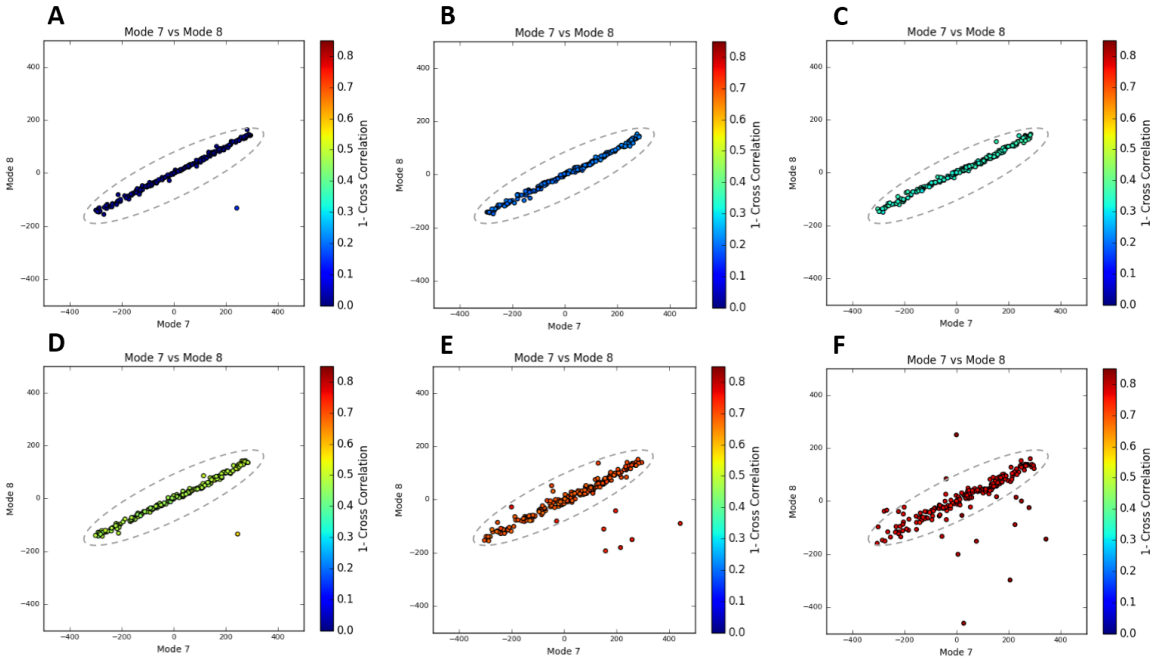

Figure S2. Plots showing the output of the 3D-to-3D elastic and rigid-body alignment module of HEMNMA-3D with synthetic datasets at different noise levels (synthetic subtomograms are simulating continuous conformational heterogeneity), using the atomic structure (chain A of PDB:4AKE) and its normal modes to estimate the conformational parameters (normal-mode amplitudes) and rigid-body parameters (orientation and shift) of the molecules in the input synthetic subtomograms. **(A)** Without noise, **(B)** SNR = 0.4, **(C)** SNR = 0.1, **(D)** SNR = 0.04, **(E)** SNR = 0.01, **(F)** SNR = 0.005. The goal was to retrieve the ground-truth relationship between the amplitudes along normal modes 7 and 8 (ideally linear relationship, with the amplitude of normal modes 8 equals to half the amplitude of mode 7); each point in the plot represents a subtomogram, and close points represent similar conformations. Note that the dashed ellipses contain the data points where the p-value is specified in Table S1.

Noise and CTF		Mode 7		Mode 8		Mode 9		p-value	Samples
Defocus [µm]	SNR	mean	std	mean	std	mean	std		
No Noise		6.38	4.47	3.37	3.51	3.70	3.85	P $>10^{-9}$	199/200
-1	0.4	8.23	6.35	7.49	5.29	5.64	4.02	P >0	200/200
-1	0.1	8.26	6.34	8.10	5.83	5.85	4.26	P >0	200/200
-1	0.04	11.14	7.20	8.11	5.89	6.98	4.83	P $>10^{-9}$	199/200
-1	0.01	26.68	9.00	13.11	8.75	16.59	9.26	p >0.01	190/200
-1	0.005	35.86	14.20	20.87	11.02	19.40	9.53	p >0.01	187/200

Table S1. Mean absolute error and standard deviation between the estimated and ground-truth normal-mode amplitudes obtained with HEMNMA-3D synthetic datasets for different noise levels, using an atomic structure as an input reference. The corresponding region for the p-value is shown in Figure S2.

3 ADDITIONAL VISUALIZATION ON THE RESULTS WITH *IN SITU* CRYO-ET NUCLEOSOME DATA

Figure S3 represents an overlapping comparison between the atomic nucleosome structure PDB:3w98 and the four group averages obtained from HEMNMA-3D for the nucleosome dataset *in situ*.

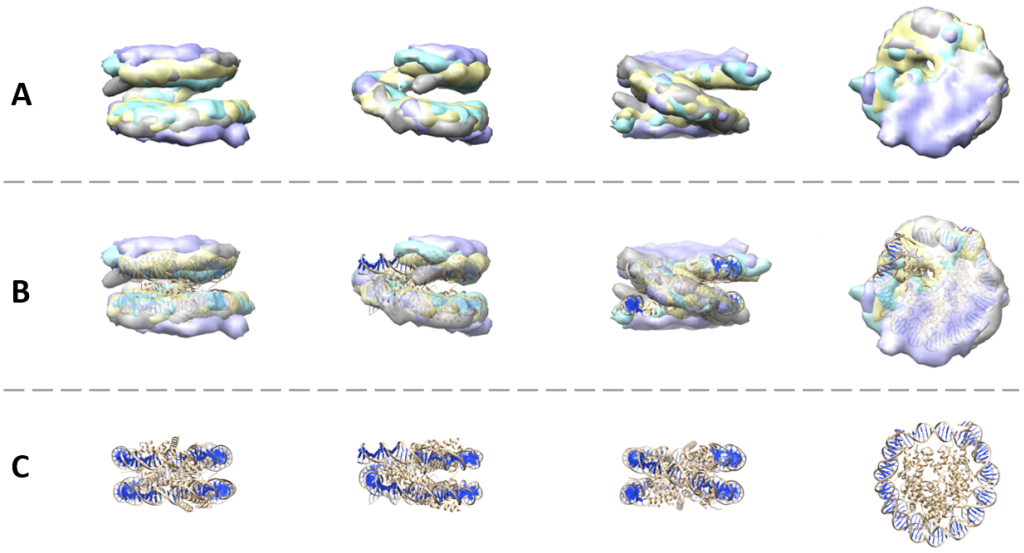

Figure S3. Comparison of the atomic nucleosome structure PDB:3w98 with the four *in situ* nucleosome subtomogram averages obtained with HEMNMA-3D (the experiment shown in Figure 10 in the main article manuscript, which used a preliminary nucleosome subtomogram average as input reference density map for HEMNMA-3D). **(A)** Four views of the four subtomogram averages overlapped. **(B)** Four views of the four averages overlapped at 50% transparency with PDB:3w98. **(C)** Four views of PDB:3w98.

REFERENCES

Eltsov, M., Grewe, D., Lemercier, N., Frangakis, A., Livolant, F., and Leforestier, A. (2018). Nucleosome conformational variability in solution and in interphase nuclei evidenced by cryo-electron microscopy of vitreous sections. *Nucleic acids research* 46, 9189–9200

Hagen, W. J., Wan, W., and Briggs, J. A. (2017). Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *Journal of structural biology* 197, 191–198

Kremer, J. R., Mastronarde, D. N., and McIntosh, J. R. (1996). Computer visualization of three-dimensional image data using imod. *Journal of structural biology* 116, 71–76

Kunz, M. and Frangakis, A. S. (2014). Super-sampling sart with ordered subsets. *Journal of structural biology* 188, 107–115

Kunz, M. and Frangakis, A. S. (2017). Three-dimensional ctf correction improves the resolution of electron tomograms. *Journal of structural biology* 197, 114–122

Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology* 152, 36–51