



**HAL**  
open science

# ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data

Mohamad Harastani, Rémi Vuillemot, Ilyes Hamitouche, Nima Barati Moghadam, Slavica Jonic

## ► To cite this version:

Mohamad Harastani, Rémi Vuillemot, Ilyes Hamitouche, Nima Barati Moghadam, Slavica Jonic. ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. *Journal of Structural Biology*, 2022, 214 (4), pp.107906. 10.1016/j.jsb.2022.107906 . hal-03821777

**HAL Id: hal-03821777**

**<https://hal.science/hal-03821777>**

Submitted on 19 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

This is the author's version of an article accepted for publication in Journal of Structural Biology, <https://doi.org/10.1016/j.jsb.2022.107906>

**ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data**

Mohamad Harastani, Rémi Vuillemot, Ilyes Hamitouche, Nima Barati Moghadam, Slavica Jonic\*

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

\*corresponding author

**Contact details of the corresponding author:**

Slavica Jonic

IMPMC - UMR 7590 CNRS

Sorbonne Université, CC 115

4 place Jussieu, 75005 Paris, France

Phone : +33 1 44 27 72 05

Fax : +33 1 44 27 37 85

E-mail : [slavica.jonic@upmc.fr](mailto:slavica.jonic@upmc.fr)

**Short title:** ContinuousFlex

**Abstract:** ContinuousFlex is a user-friendly open-source software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy (cryo-EM) and cryo electron tomography (cryo-ET) data. In 2019, ContinuousFlex became available as a plugin for Scipion, an image processing software package extensively used in the cryo-EM field. Currently, ContinuousFlex contains software for running (1) recently published methods HEMNMA-3D, TomoFlow, and NMMD; (2) earlier published methods HEMNMA and StructMap; and (3) methods for simulating cryo-EM and cryo-ET data with conformational variability and methods for data preprocessing. It also includes external software for molecular dynamics simulation (GENESIS) and normal mode analysis (EINemo), used in some of the mentioned methods. The HEMNMA software has been presented in the past, but not the software of other methods. Besides, ContinuousFlex currently also offers a deep learning extension of HEMNMA, named DeepHEMNMA. In this article, we review these methods in the context of the ContinuousFlex package, developed to facilitate their use by the community.

### **Keywords**

ContinuousFlex; cryo-EM single particle analysis; cryo-ET subtomogram analysis; continuous conformational variability of macromolecules, software, Scipion

## 1. Introduction

Recent instrumental and software developments allowed obtaining various *in vitro* and *in situ* high-resolution macromolecular structures by single particle cryo electron microscopy (cryo-EM) (Banerjee et al., 2016; Kato et al., 2021; Li et al., 2022; Liao et al., 2013; Nakane et al., 2020) and cryo electron tomography (cryo-ET) (Himes and Zhang, 2018; Schur et al., 2016; Turoňová et al., 2020; von Kügelgen et al., 2020; Wan et al., 2017). However, the challenge is still to fully decipher the conformational variability underlying continuous conformational transitions of macromolecules (Dashti et al., 2020; Jonić, 2017; Sorzano et al., 2019), especially *in situ* (Harastani et al., 2021; Harastani et al., 2022; Turoňová et al., 2020).

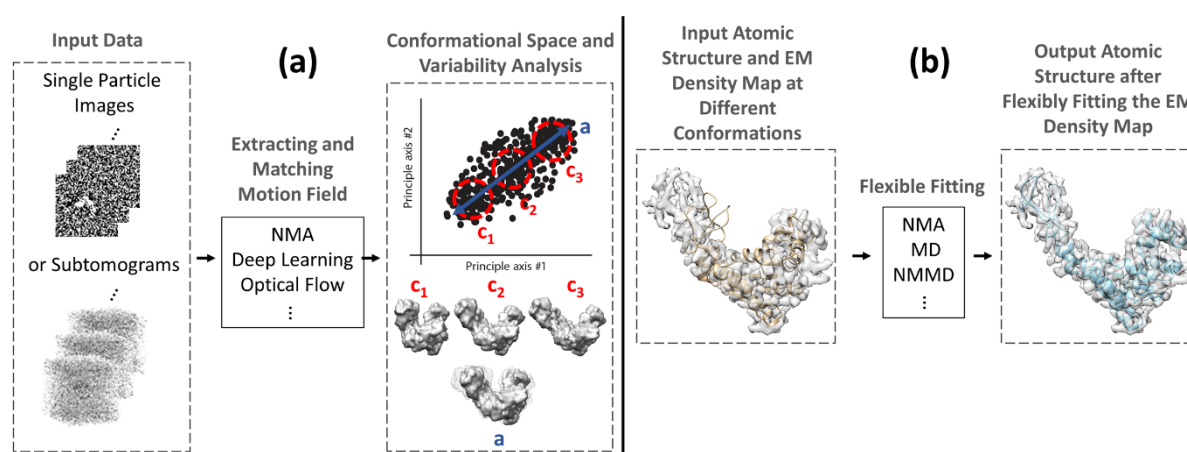
Macromolecular conformations determine, to a large extent, their functional mechanisms (Hofmann et al., 2019). To accomplish biological functions, macromolecules change their conformations by traversing many intermediate states (Dashti et al., 2020). The hypothesis that the states are samples of a continuous trajectory of the conformational transition is at the basis of the concept of continuous conformational variability (Jonić, 2017). A way to achieve a better understanding of the biological functions of the macromolecules is to determine their entire conformational landscapes by cryo-EM or cryo-ET (Chen and Ludtke, 2021; Dashti et al., 2020; Harastani et al., 2021; Harastani et al., 2022; Jin et al., 2014; Punjani and Fleet, 2021; Zhong et al., 2021). This requires solving a complex problem of combined heterogeneity of conformations, orientations, and positions of the macromolecules in the experimental data, which are highly noisy and additionally affected by the specificities of the data acquisition geometry.

ContinuousFlex is a user-friendly open-source software package primarily developed for obtaining conformational landscapes of macromolecules by an exhaustive analysis of their continuous conformational variability in cryo-EM/ET data (**Figure 1a**). Additionally, it provides methods for flexible fitting of cryo-EM maps with atomic models (**Figure 1b**). ContinuousFlex branched from Scipion (de la Rosa-Trevín et al., 2016) and its backend software Xmipp (Strelak et al., 2021) in 2019 (Harastani et al., 2020). It is currently available as a plugin of Scipion, with Xmipp hosting several of its backend data processing steps. This pluginization allowed better maintenance, faster development, and more frequent releases of bug fixes and developed methods. As a plugin of Scipion, ContinuousFlex allows reproducible research, as all the data processing steps used in experiments are automatically stored on the disk (together with their parameters) and can be reproduced at any moment using the same or modified parameters. Additionally, the project containing all the data processing steps can be directly uploaded to EMPIAR, as allowed by Scipion.

Currently, ContinuousFlex contains software for running (1) recently published methods HEMNMA-3D (3D Hybrid Electron Microscopy Normal Mode Analysis of cryo-ET subtomograms (Harastani et al., 2021)), TomoFlow (3D Tomographic optical Flow analysis of cryo-ET subtomograms (Harastani et al., 2022)), and NMMD (3D Normal Mode

and Molecular Dynamics flexible fitting of cryo-EM maps with atomic structures (Vuillemot et al., 2022)); (2) earlier published methods HEMNMA (3D-to-2D Hybrid Electron Microscopy Normal Mode Analysis of cryo-EM images (Harastani et al., 2020; Jin et al., 2014)) and StructMap (Structural Mapping of cryo-EM maps (Sanchez Sorzano et al., 2016)); and (3) methods for simulating cryo-EM and cryo-ET data with conformational variability and methods for data preprocessing used in our previous works (Harastani and Jonic, 2021; Harastani et al., 2021; Harastani et al., 2022; Vuillemot et al., 2022). It also includes external software for molecular dynamics simulation (GENESIS (Kobayashi et al., 2017)) and normal mode analysis (EINemo (Suhre and Sanejouand, 2004)), used in some of the mentioned methods. The first version of ContinuousFlex, released in 2019, contained only the software of HEMNMA (Harastani et al., 2020; Jin et al., 2014). The software of other methods has been developed and added to ContinuousFlex later. Besides, ContinuousFlex currently also offers a deep learning extension of HEMNMA, named DeepHEMNMA, the software of a very recent method (article in press (Hamitouche and Jonic, 2022)). In this article, we review these methods in the context of the ContinuousFlex package, developed to facilitate their use by the community.

The ContinuousFlex methods for exhaustive analyses of continuous conformational variability in cryo-EM and cryo-ET data are presented in **Section 2** and **Section 3**, respectively. The methods for flexible fitting of cryo-EM maps with atomic models are presented in **Section 4**. The methods for optional data synthesis and optional (pre)processing are presented in **Section 5**. ContinuousFlex dependencies are discussed in **Section 6** and basic information on installation tools and on tools for testing the methods contained in ContinuousFlex is provided in **Supplementary Material**. The discussion and conclusion are provided in **Section 7**.



**Figure 1** Illustration of the methods in ContinuousFlex: (a) methods used for obtaining conformational landscapes of macromolecules by an exhaustive analysis of their continuous conformational variability in cryo-EM/ET data; (b) methods for flexible fitting of cryo-EM maps with atomic models. The methods in ContinuousFlex are based on normal mode analysis (NMA), molecular dynamics simulation (MD), combination of NMA and MD (NMMD), deep learning, or optical flow.

## 2. Methods for cryo-EM data analysis of continuous conformational variability

This section presents ContinuousFlex methods for analyzing single particle images and EM maps in terms of continuous conformational variability.

### 2.1. HEMNMA: Analysis of a set of single particle images using normal modes

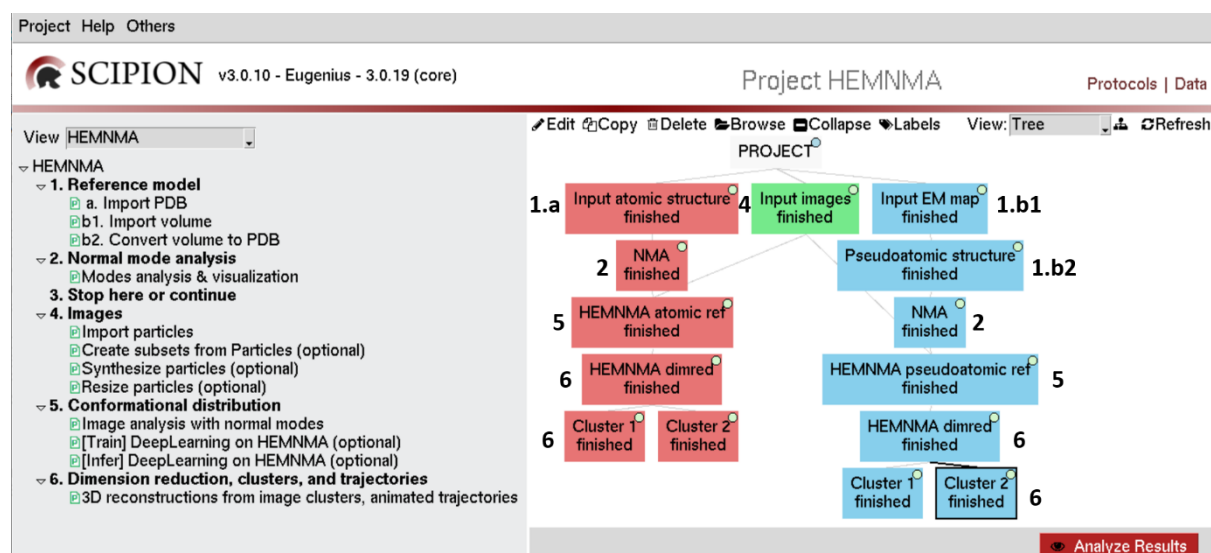
HEMNMA (Harastani et al., 2020; Jin et al., 2014) is an image analysis method for analyzing macromolecular continuous conformational variability in single-particle cryo-EM images. HEMNMA employs a reference (pseudo)atomic model for the image analysis, i.e., it requires as input either an atomic structure or an EM density map that is converted to a pseudoatomic structure (Jonić and Sorzano, 2016). An EM density map as a choice of reference in HEMNMA is particularly useful when no atomic model of the macromolecule is available, and it can be obtained with traditional alignment and reconstruction techniques from input single particle images (several of these techniques can be found in Scipion and Xmipp). HEMNMA performs normal mode analysis (NMA) on the input structure using an Elastic Network Model (Tirion, 1996). NMA is a commonly used technique for simulating molecular mechanics (an interested reader is referred to (Ma, 2005)). The NMA software available in ContinuousFlex is the NMA implementation run by EINemo (Suhre and Sanejouand, 2004) webserver. This NMA software can be run independently of HEMNMA as well, and it is also used by other methods in ContinuousFlex (HEMNMA-3D and NMMD, as explained below). NMA results in a set of predicted motion directions (vectors along which move atoms/pseudoatoms, which are called normal modes), but it does not provide information about the amplitudes of motion along the normal modes. In the case of HEMNMA, these amplitudes are determined by analyzing single-particle images with normal modes. Normal modes are ordered according to their frequency, and normal modes with low frequencies usually have higher collectivity (represent large global movements). A subset of normal modes with low frequency and high collectivity (e.g., the first ten modes from mode 7 on) can be a sufficiently good basis for modeling molecular elastic deformation. It should be noted that modes 1-6 represent rigid-body movements. A macromolecule at conformation  $A$  displaced using a linear combination of normal modes of given amplitudes will reach a conformation  $B$ . Given a reference conformation  $A$  (atomic or pseudoatomic model) of a macromolecule and a single-particle image of conformation  $B$  of the same macromolecule, the goal of HEMNMA is to find the amplitudes of the linear combination of normal modes that can elastically deform the input conformation  $A$  to match  $B$ . For each input image, HEMNMA iteratively (via a numerical optimizer (Vanden Berghen and Bersini, 2005)) displaces the atoms/pseudoatoms of the input model using normal modes, converts the modified model into a density map, and compares 2D projections of that map with the input image until the best matching is found. This iterative process determines the optimal orientation (three Euler angles), position (two in-plane translations), and normal-mode amplitudes of the input model to match each input image. The conformational parameters for the ensemble of images are then fed to a dimensionality reduction method, e.g., Principal Component Analysis (PCA) (Wold et al., 1987), to map images onto a low-dimensional space (usually 2D or 3D space for visualization purposes). Trajectories of conformational changes may be identified by exploring

interactively the most densely populated regions in the low-dimensional space, and the input (pseudo)atomic structure can be displaced along these trajectories to visualize the changes in 3D. Also, density maps can be reconstructed by an interactive grouping of images with similar conformations (close points in dense regions in the low dimensional space).

Note here that the terms “protocol” and “tree” used in this article mean a data processing step and a set of performed data processing steps, respectively (e.g., performing NMA on a structure and performing all the steps of HEMNMA on an input single particle dataset, respectively).

ContinuousFlex allows performing all the steps of HEMNMA, as follows (**Figure 2**):

- 1- Importing an atomic structure (1.a) or an EM map (1.b1). If an EM map is imported, it is converted to a pseudoatomic structure (1.b2).
- 2- Performing NMA.
- 3- Reminding the user that HEMNMA software may also be used only for NMA and visualization of normal modes.
- 4- Importing images, creating subsets of images (optional), synthesizing images (optional), and resizing images (optional). We note here that the protocol "Create subsets from Particles" (creating subsets of images) is useful when performing HEMNMA with deep learning (explained in the next paragraph). The protocol "Synthesize particles" produces synthetic images for testing the method and is not a part of data processing (refer to **Section 5.1** for details). The protocol "Resize images" can be useful for downscaling images of large sizes.
- 5- Performing rigid-body and normal-mode-based elastic alignment of the input model with each single-particle image. At this step, HEMNMA calculates normal-mode amplitudes, three Euler angles, and two in-plane shifts for each single-particle image. Two optional protocols available at this step are for combining HEMNMA with deep learning, which is explained in the next paragraph.
- 6- Obtaining and analyzing a low-dimensional conformational space using the normal-mode amplitudes, Euler angles, and shifts estimated for different single-particle images at step 5. This step allows obtaining 3D reconstructions and molecular movies from the conformational space.



**Figure 2** Graphical interface of HEMNMA in ContinuousFlex. Green box: input dataset of single particle images. Red branch: HEMNMA processing with an atomic structure as the reference. Blue branch: HEMNMA processing with an EM map as the reference. None of the tools marked as optional (menu on the left) were used in this figure. The numbers of the steps in the menu on the left are indicated in the tree on the right.

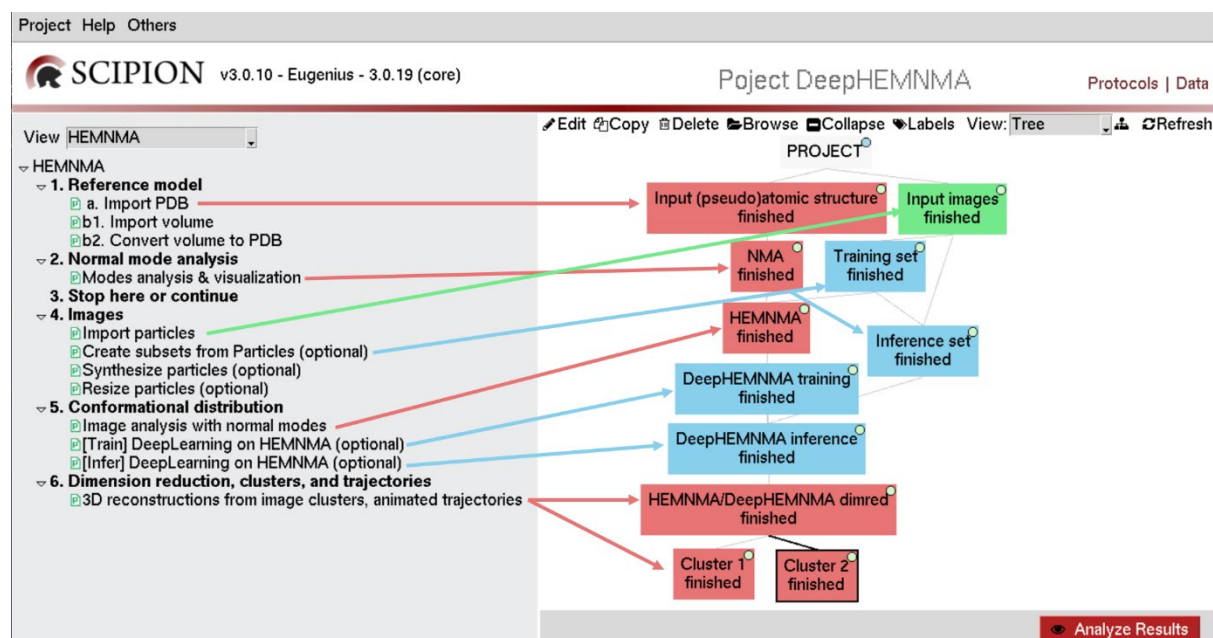
Step 5 is the most time-consuming step of HEMNMA, and it is Message Passing Interface (MPI) parallelized, which allows the processing of multiple images simultaneously. Nevertheless, the computational demand is overwhelming for cryo-EM single particle datasets containing hundreds of thousands to millions of images nowadays. Therefore, we have recently combined HEMNMA with deep learning to accelerate it and this method is referred to as DeepHEMNMA (Hamitouche and Jonić, 2021; Hamitouche and Jonic, 2022). The deep learning neural network of DeepHEMNMA is based on ResNet-34, which is a residual convolutional neural network with 34 layers (He et al., 2016), trained on HEMNMA estimations of rigid-body and elastic alignment parameters for single particle images (angles, shifts, and normal-mode amplitudes) on a small subset (training set) of an input single particle dataset. Once trained, DeepHEMNMA infers angles, shifts, and normal-mode amplitudes from the remaining images of the dataset (inference set). Then, the results from HEMNMA (on the training set) and DeepHEMNMA (on the inference set) are combined and analyzed in terms of animations and 3D reconstructions in the low-dimensional conformational space, as when using HEMNMA alone. HEMNMA and DeepHEMNMA give very similar results but the estimated total number of computing hours needed by DeepHEMNMA for obtaining normal-mode amplitudes, angles, and shifts is several tens of times smaller compared to the number of computing hours needed by HEMNMA (e.g., of the order of 40 times for 1 million synthetic images of size  $128 \times 128$  pixels and 3 normal modes, (Hamitouche and Jonic, 2022)).

DeepHEMNMA is incorporated in ContinuousFlex via optional steps of HEMNMA, as shown in **Figure 3**. These optional steps allow:

- Splitting an input single-particle dataset into training and inference sets,
- Performing HEMNMA on the training set (up to step 5 or 6 in **Figure 2**),



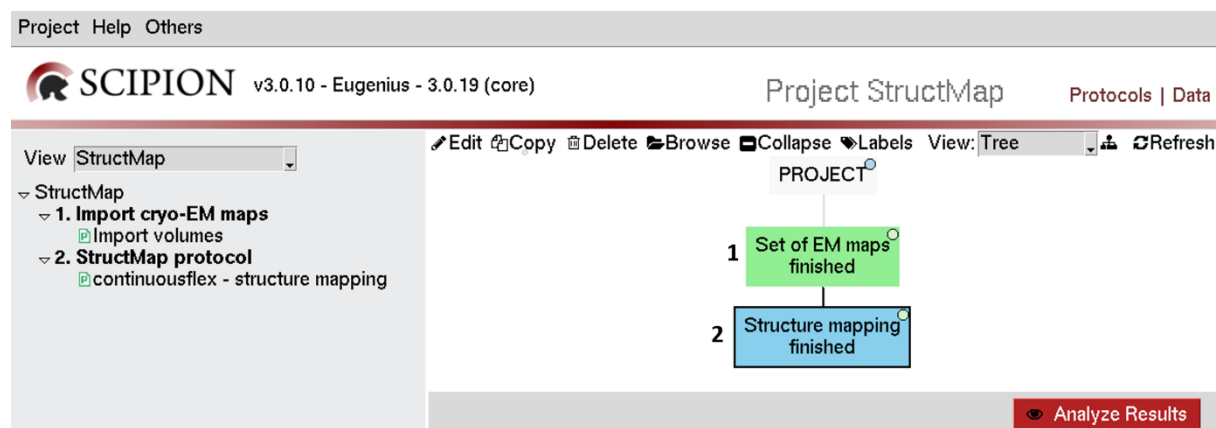
- Training the network on the results of HEMNMA,
- Using the trained network to produce results on the inference set, and
- Obtaining and analyzing the conformational space for the complete input dataset by combining the results of HEMNMA and inference.



**Figure 3** Graphical interface of HEMNMA in ContinuousFlex showing the use of deep learning in combination with HEMNMA for faster data processing (DeepHEMNMA method). Green box: input dataset of single particle images. Red boxes: Typical HEMNMA steps (also shown in **Figure 2**). Blue boxes: Additional steps required to accelerate HEMNMA with deep learning.

## 2.2. StructMap: Analysis of a set of EM maps using normal modes

StructMap (Sanchez Sorzano et al., 2016) is a method for analyzing conformational differences between EM maps obtained via 3D reconstructions from different image datasets or different classes of the same image dataset. It projects a set of input EM maps to a lower-dimensional space based on their conformational distances. This lower-dimensional space allows visualization of the differences. For a set of EM maps, StructMap performs pairwise EM-map fitting based on the conversion of cryo-EM maps to pseudoatomic structures (Jonić and Sorzano, 2016) and normal modes of the pseudoatomic structures. StructMap uses the pairwise fitting scores between the given EM maps (correlation coefficients) to find a distance matrix. This distance matrix is mapped onto a lower-dimensional space via multidimensional scaling (Cox and Cox, 2008), allowing visualization of the distances.



**Figure 4** Graphical interface of StructMap in ContinuousFlex. Green box: input set of EM maps. Blue box: StructMap processing. The "Analyze Results" button allows visualizing the low-dimensional space of the distances between the EM maps. The numbers of the steps in the menu on the left are indicated in the tree on the right.

The previous dependency of StructMap software on Matlab Statistical Toolbox (Sanchez Sorzano et al., 2016) has been removed, and StructMap is now entirely integrated into ContinuousFlex. The workflow consists of importing a set of EM maps and applying StructMap, as shown in **Figure 4**. The "Analyze Results" button allows visualizing the low-dimensional space of the distances between the EM maps.

### 3. Methods for cryo-ET data analysis of continuous conformational variability

This section presents ContinuousFlex methods for analyzing cryo-ET subtomograms in terms of continuous conformational variability. The input subtomograms are assumed to come from the tomograms corrected for the Contrast Transfer Function (CTF) of the electron microscope. For instance, the nucleosome subtomograms processed with TomoFlow and HEMNMA-3D methods in (Harastani et al., 2021; Harastani et al., 2022) had been extracted from the tomogram corrected for the CTF using the method proposed in (Kunz and Frangakis, 2017).

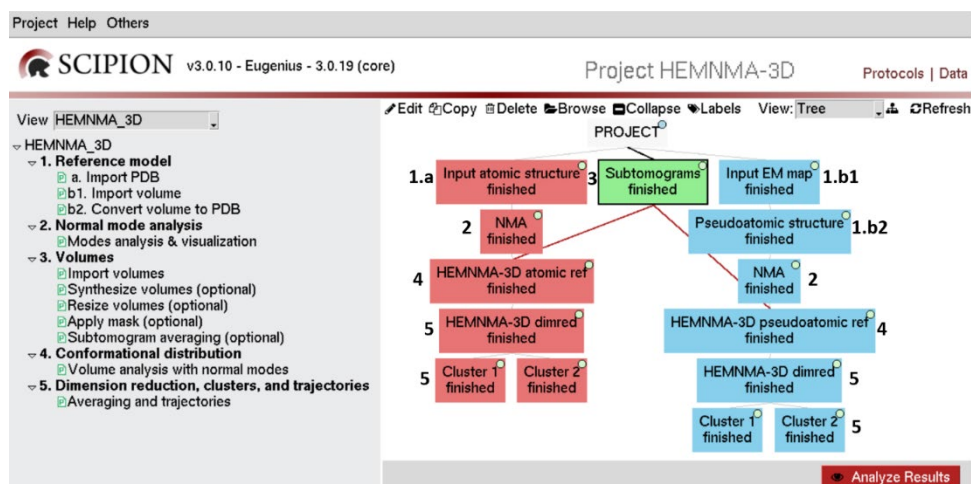
#### 3.1. HEMNMA-3D: Analysis of a set of subtomograms using normal modes

HEMNMA-3D (Harastani et al., 2021) is an extension of HEMNMA (explained in **Section 2.1**) to analyze cryo-ET subtomograms. We explain in the following the key differences between HEMNMA and HEMNMA-3D. HEMNMA-3D is a volumetric image analysis method for macromolecular continuous conformational variability in cryo-ET subtomograms. As HEMNMA, HEMNMA-3D employs a reference structure with its normal modes to analyze the input data. As in HEMNMA, this reference can be an atomic structure or an EM density map converted to a pseudo-atomic structure. For instance, the reference can be a pseudoatomic structure from the global subtomogram average found using traditional subtomogram alignment and averaging techniques (one such technique is available in ContinuousFlex). For each input subtomogram, HEMNMA-3D iteratively modifies the reference using normal modes, converts the modified reference into a density map, performs its rigid-body alignment with the subtomogram via Constrained Cross-Correlation (meaning in the presence of a missing wedge mask to compensate for missing wedge artifacts) (Chen et al., 2013), until the optimal orientation (three Euler angles), 3D translation,

and normal-mode amplitudes are found. The conformational parameters for the ensemble of subtomograms are then fed to a dimensionality reduction method to map them onto a low-dimensional space. As in HEMNMA, conformational changes may be animated in the low-dimensional space interactively (along the densest regions). Also, new subtomogram averages can be computed from subtomograms with similar conformations interactively (close points in the conformational space). HEMNMA-3D deals with the problem of missing wedge by using a missing-wedge mask in the Fourier space to exclude the missing-wedge-related Fourier coefficients from the cross-correlation evaluation (used for the alignment of the subtomograms with the reference to calculate the conformational and rigid-body parameters). The missing-wedge is expected to be attenuated in the subtomogram averages calculated by low-dimensional-space grouping of aligned similar conformations because of the averaging of different missing-wedge directions in the aligned subtomograms. The HEMNMA-3D performance for different values of Signal-to-Noise Ratio (SNR between 0.005 and 0.4 applied on synthesized tilt-series particle images) and CTF defocus of  $-1\ \mu\text{m}$  was shown in the HEMNMA-3D introductory article (see the main text and the supplementary material of (Harastani et al., 2021)).

ContinuousFlex allows performing all the steps of HEMNMA-3D, as follows (**Figure 5**):

- 1- Importing an atomic structure (1.a) or an EM map (1.b1). If an EM map is imported, it is converted to a pseudoatomic structure (1.b2).
- 2- Performing NMA.
- 3- Importing volumes (subtomograms), synthesizing volumes (optional), resizing volumes (optional), masking volumes (optional), and performing traditional subtomogram averaging (optional). We note here that the protocol "Synthesize volumes" can be used to create data for testing the method and is not a part of data processing, whereas the protocol "Subtomogram averaging" can be used to find the global subtomogram average of the input set (which can serve as a starting reference for HEMNMA-3D). The protocols "Resize volumes" and "Apply mask" are optional data preprocessing techniques.
- 4- Performing rigid-body and normal-mode-based elastic alignment of the reference model with each subtomogram. At this step, HEMNMA-3D calculates normal-mode amplitudes, three Euler angles, and a 3D shift for each subtomogram.
- 5- Obtaining and analyzing a low-dimensional conformational space using the normal-mode amplitudes, Euler angles, and shifts estimated for all subtomograms at step 4. This step allows obtaining movies of conformational transitions and subtomogram averages of similar conformations from the conformational space.



**Figure 5** Graphical interface of HEMNMA-3D in ContinuousFlex. Green box: input subtomograms. Red branch: HEMNMA-3D processing with an atomic structure as the reference. Blue branch: HEMNMA-3D processing with an EM map as the reference. None of the tools marked as optional (menu on the left) were used in this figure. The numbers of the steps in the menu on the left are indicated in the tree on the right.

Step 4 of HEMNMA-3D software is MPI parallelized, allowing multiple volumes to be processed simultaneously. Unlike cryo-EM datasets, the size of cryo-ET datasets is still limited to a few thousand subtomograms, which can be processed with HEMNMA-3D in a few hours on supercomputer centers. Nevertheless, combining HEMNMA-3D with deep learning for faster data processing is an interesting future work.

### 3.2. TomoFlow: Analysis of a set of subtomograms using optical flow

TomoFlow (Harastani et al., 2022) is a method for analyzing macromolecular continuous conformational variability in cryo-ET subtomograms based on the approach of 3D dense optical flow (OF) (Farnebäck, 2003; Shah and Xuezi, 2021). TomoFlow maps a set of subtomograms onto a low-dimensional space based on a novel mathematical model that uses OF properties.

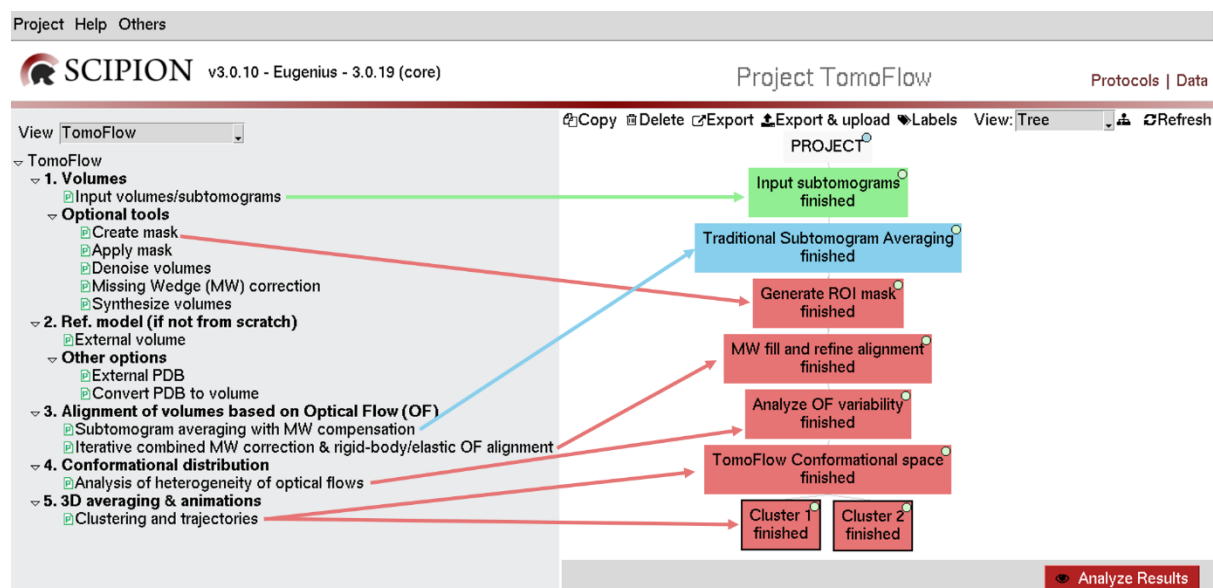
TomoFlow first obtains rigid-body alignment of subtomograms by traditional subtomogram averaging approaches. Then, it refines this rigid-body alignment using an iterative missing wedge correction (Tang et al., 2007), optical flow warping, and fast rotational matching (Chen et al., 2013). For each subtomogram with the refined rigid-body alignment, it finds an OF against the global subtomogram average, which represents an estimation of the conformational transition between the subtomogram and the global subtomogram average. The set of OFs obtained for all subtomograms is used to generate a Gram matrix (a matrix of their pairwise inner products), which is mapped onto a low-dimensional space via PCA. As HEMNMA-3D (explained earlier in **Section 3.1**), the low-dimensional space of TomoFlow allows obtaining movies of macromolecular motion along different directions in the conformational space and obtaining subtomogram averages of similar conformations by interactive grouping of close points in this space (close points represent conformationally similar subtomograms).

It is worth noting that TomoFlow, unlike HEMNMA-3D, does not depend on matching subtomograms with simulated movements of an input reference structure (e.g., based on NMA), but rather, it extracts motions and conformations directly from the data, which makes it easier to use and less prone to results misinterpretation. However, TomoFlow is prone to conformational smoothing due to the smoothing effects introduced during the OF calculation, and the used missing wedge correction based on filling the missing wedge region in the subtomogram with the corresponding region of the global subtomogram average (Tang et al., 2007). Nevertheless, it has been shown that TomoFlow and HEMNMA-3D can detect similar conformations (e.g., breathing and gapping motions of nucleosomes *in situ* (Harastani et al., 2021; Harastani et al., 2022)), and an interested user can try both methods on their dataset. It should be noted that the missing wedge correction methods that introduce data smoothing via denoising (e.g., isoNET (Liu et al., 2021)) may be detrimental to TomoFlow performance because of the data smoothing already introduced by the OF approach of TomoFlow.

ContinuousFlex allows performing all the steps of TomoFlow, as follows (**Figure 6**):

- 1- Importing volumes (subtomograms) and some optional tools for creating and applying masks, denoising volumes, missing wedge correction, and synthesizing volumes. The optional tools for missing wedge correction and volume synthesis are explained in **Section 5**. The mask tools are useful for creating a mask based on the global subtomogram average and applying it to aligned subtomograms to eliminate the background noise.
- 2- Importing an external reference (optional) as an EM map or an atomic structure converted to an EM map to perform subtomogram alignment. This is only recommended when reference-free subtomogram alignment fails.
- 3- Performing or importing subtomogram alignment and averaging (explained in **Section 5**), generating a mask for the region of interest (used from Optional tools in Step 1), refining the rigid-body alignment, and filling the missing wedge.
- 4- Finding the OF between the refined subtomogram average and the missing wedge corrected and aligned subtomograms, then using the OFs to construct a Gram matrix.
- 5- Obtaining the conformational space (PCA on the Gram matrix) and its interactive analysis by generating movies of conformational transitions along different directions in the space and calculating subtomogram averages of similar conformations.

TomoFlow software is MPI parallelized for rigid-body alignment, and it uses GPU processing for OF calculation.



**Figure 6** Graphical interface of TomoFlow in ContinuousFlex. Green box: input subtomograms. Blue box: traditional subtomogram averaging (which can also be done using other software packages and imported to Scipion). Red branch: other steps specific to TomoFlow.

## 4. Methods for flexible fitting of cryo-EM maps with atomic models

This section presents the methods available in ContinuousFlex for flexible fitting of cryo-EM maps with atomic models, which is usually done to interpret cryo-EM maps in terms of available atomic coordinates assuming that the cryo-EM map was reconstructed from a globally conformationally homogeneous set of particles. The macromolecular conformations in the atomic model and the cryo-EM map can be regarded as two samples of a continuous conformational transition of the macromolecule. In other words, continuous conformational variability and flexible fitting of cryo-EM maps with atomic models are closely related.

### 4.1. EMfit: Fitting based solely on MD simulation

When considering the problem of flexible fitting of an atomic model into a cryo-EM map, a well-established method for simulating molecular mechanics is Molecular Dynamics (MD) (Miyashita et al., 2017; Orzechowski and Tama, 2008; Trabuco et al., 2008). MD allows exploring the conformational space of macromolecules by simulating physical interactions between atoms. Atomic displacements in MD are simulated using a set of semi-empirical potentials (or force fields), producing biochemically reliable structures. In MD-based flexible fitting methods, the set of potentials includes an additional "biasing" potential that incorporates information from the cryo-EM map (e.g., through the correlation coefficient) and guides the simulation to the target conformation (Orzechowski and Tama, 2008).

In ContinuousFlex, we incorporated a version of the MD simulation software GENESIS (Kobayashi et al., 2017), which features the possibility to use several force fields, from all-atom to coarse-grained models, equipped with high parallelization efficiency. GENESIS comes with the flexible fitting method EMfit that is based on MD simulation



and can use the Replica Exchange Umbrella Sampling (REUS) procedure for refining the biasing-potential force constant during the fitting (Miyashita et al., 2017; Orzechowski and Tama, 2008). The use of GENESIS for MD-based flexible fitting can be found in the literature (Kulik et al., 2021; Miyashita et al., 2017; Mori et al., 2021; Mori et al., 2019; Vuillemot et al., 2022). The force fields available in ContinuousFlex, via GENESIS, are CHARMM (Huang and MacKerell Jr, 2013) and two Go models (all-atom and C $\alpha$ -atom based) (Karanicolas and Brooks, 2003). CHARMM relies on atomistic models used in a wide range of molecular simulations, whereas Go models are coarse-grained models useful for simulating long-term dynamics. Also, GENESIS in ContinuousFlex allows performing energy minimization, which is recommended to be done before MD simulations.

We used the implementation of EMfit in GENESIS version 1.4 as the basis for the development of the NMMD flexible fitting method, which is discussed next (**Section 4.2**). For this purpose, GENESIS version 1.4 has been modified and ContinuousFlex provides this modified version of GENESIS. Also, ContinuousFlex allows running EMfit and NMMD separately, for their comparison, as presented in the next section (**Section 4.2, Figure 7**). A typical workflow using EMfit in ContinuousFlex has the following steps: importing an atomic structure and a cryo-EM density map, performing energy minimization using GENESIS, and performing flexible fitting with MD simulation (EMfit) using GENESIS.

#### **4.2. NMMD: Fitting using a combination of NMA and MD simulation**

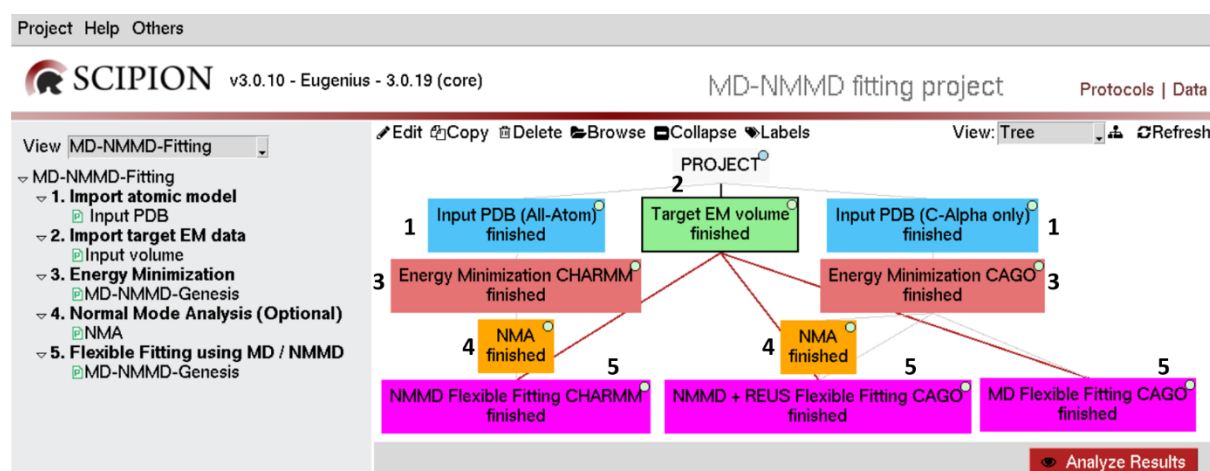
ContinuousFlex also provides the NMMD method (Vuillemot et al., 2022), which efficiently performs flexible fitting of atomic models into cryo-EM maps using a combination of MD simulations and NMA.

While MD produces accurate atomic motions and reliable conformations, it requires integrating a large number of degrees of freedom, generally three Cartesian coordinates per atom, which results in computationally expensive simulations. On the other hand, NMA simulates molecular mechanics with a few degrees of freedom (generally a small subset of normal modes), which considerably reduces the computational cost. Nevertheless, NMA limits the possibilities of motions, making it inadequate when fitting local dynamics but well suited for simulating global displacements. Besides, NMA can create distortions of the structure for large normal mode amplitudes, making it less reliable than MD in terms of structural quality.

NMMD combines the advantages of MD and NMA into a single simulation method. In NMMD, the atomic displacements are obtained by an MD simulation that incorporates an additional NMA-based displacement. MD and NMA parameters are integrated simultaneously to accelerate the simulation (through normal-mode-based motion) and ensure good local dynamics simulation (through MD-based motion). Moreover, MD-based displacements compensate for the distortions produced by NMA-based displacements, ensuring that the properties of the structure remain unchanged. When compared to MD simulation, NMMD proved to be 40 % faster (on

average) while producing a comparable, if not better, sampling efficiency and structure quality (Vuillemot et al., 2022), using different synthetic EM maps and experimental cryo-EM maps of p97 ATPase (Banerjee et al., 2016) and ABC exporter (Hofmann et al., 2019).

As mentioned in **Section 4.1**, the implementation of NMMD in GENESIS is based on the GENESIS implementation of EMfit (fitting using MD only). As EMfit, NMMD can thus be run with REUS using CHARMM or Go-model force fields, which are provided by GENESIS. As in other normal-mode-based methods in ContinuousFlex, NMA in NMMD is performed using Elnemo.



**Figure 7** ContinuousFlex graphical interface of the methods for flexible fitting of atomic structures into cryo-EM maps, using MD only (EMfit method) and using a combination of MD and NMA (NMMD method). Green box: input EM map. Blue boxes: two options for an atomic input structure (with all atoms or with C $\alpha$  atoms only). Red boxes: minimization steps (prior to flexible fitting). Orange boxes: NMA (marked as optional in the menu on the left as it is only used with NMMD). Magenta boxes: flexible fitting using MD only (EMfit) or using NMMD, with different options. The numbers of the steps in the menu on the left are indicated in the tree on the right. Only the examples of using CHARMM and C $\alpha$ -atom Go model (CAGO) are shown, but the interface allows using all-atom Go model (AAGO) as well.

A typical workflow using EMfit (MD only) or NMMD (MD combined with NMA) has the following steps (**Figure 7**):

- 1- Importing an atomic structure.
- 2- Importing a cryo-EM density map.
- 3- Performing energy minimization using GENESIS.
- 4- For NMMD: computing normal modes of the structure at minimum energy.
- 5- Performing flexible fitting with EMfit (denoted by MD in **Figure 7**) or NMMD using GENESIS. The simulation can be handled by CHARMM or one of the two Go models (all-atom or C $\alpha$ -atom Go model), and coupled with REUS. Note that **Figure 7** only shows the examples of using CHARMM and C $\alpha$ -atom Go model (denoted by CAGO), but it is also possible to use all-atom Go model (AAGO).



## 5. Methods for data synthesis and (pre)processing (optional)

This section lists ContinuousFlex tools that can be used to simulate conformationally heterogeneous cryo-EM/ET datasets for methods testing purposes and optional data (pre)processing tools other than those explained in the previous sections.

### 5.1. Optional data synthesis

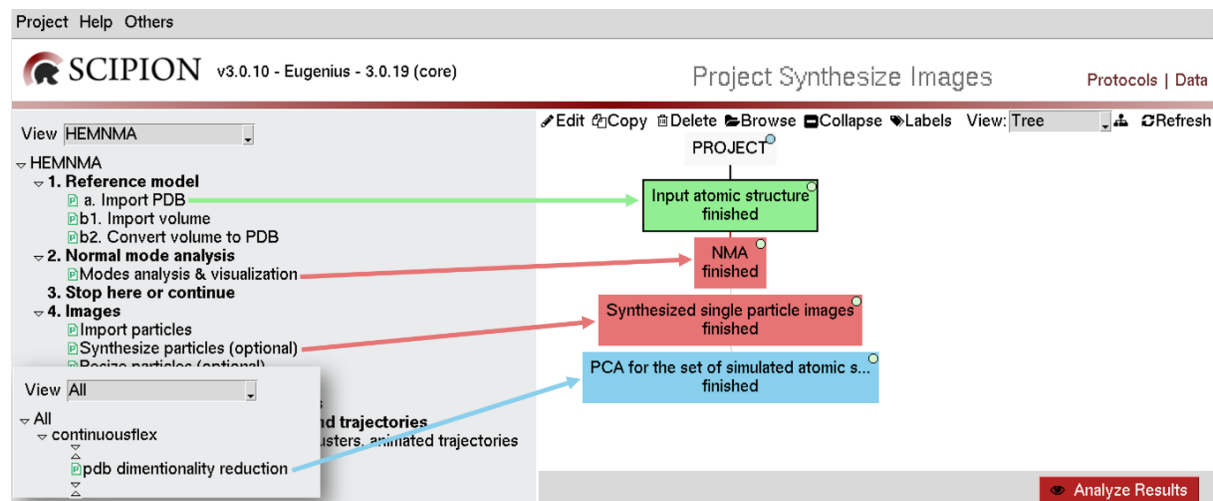
ContinuousFlex contains tools for synthesizing single particle images and subtomograms, primarily designed to simulate macromolecular conformational variability for testing methods (e.g., tests of methods under development, automated tests of the methods provided in ContinuousFlex, etc.). Two similar protocols exist, one of which is for single particle images and the other for subtomograms. These protocols simulate macromolecular motions by elastically deforming an input atomic or pseudoatomic structure (pseudoatoms from an EM map) using normal modes or they import a set of varying conformations from other sources (e.g., from an MD trajectory). Then, each of the obtained multiple conformations is converted to a density map, which is then rotated and shifted randomly. Optionally, the obtained density maps can be lowpass filtered, if simulating blurring effects is desired (e.g., caused by radiation dose accumulation). For image simulation, an obtained density map is projected once. For subtomogram simulation, the density map is projected several times to obtain a tilt-series of projection images. Then, microscope conditions specified by the CTF and noise are simulated on the obtained image or tilt-series images, using the physical image formation model in which a part of noise is applied before and the other part after the modulation by the CTF (Sorzano et al., 2007). In subtomogram simulation, a 3D reconstruction is obtained from the tilt-series. The simulated macromolecular trajectories can be visualized by applying PCA on the set of simulated (pseudo)atomic structures using a specific tool.

The user is given control over the different steps used in the simulation, including the possibility to impose a specific relationship between the amplitudes of the different normal modes used for data synthesis (if NMA is used for the data synthesis), pixel/voxel size, range and type of distribution (random uniform or Gaussian) of shifts and Euler angles that determine the position and orientation of the macromolecule in 3D, lowpass filter parameters, tilt-series angular range and angular step, CTF parameters, signal-to-noise ratio, and the 3D reconstruction algorithm.

It is worth mentioning that these protocols additionally allow simulating data with macromolecules homogeneous in conformation but heterogeneous in orientation and position in 3D (variability of rigid-body parameters only), which can help test different concepts and algorithms.

To synthesize single particle images, one can use the graphical interface of HEMNMA, as shown in **Figure 8**. A very similar approach would be used to synthesize subtomograms, but using the graphical interface of HEMNMA-3D. **Figure 8** shows an example of using the interface to synthesize single particle images with normal modes of an

atomic structure, but the images can also be synthesized without using normal modes (e.g., using an input set of atomic structures sampled from a trajectory simulated by MD) or they can be synthesized using normal modes of a pseudoatomic representation of an input cryo-EM map (not illustrated in the tree in **Figure 8**). Finally, if the purpose is only to simulate the variability of rigid-body parameters, a single atomic structure or an EM map is enough as input. Several works have used this data synthesis protocol (Harastani and Jonic, 2021; Harastani et al., 2021; Harastani et al., 2022).



**Figure 8** Protocol for synthesizing single particle images with macromolecular conformational variability using normal modes (using the graphical interface of HEMNMA). Green box: input atomic structure (note that an input cryo-EM map can also be used). Red boxes: calculation and use of normal modes to synthesize images. Blue box: obtaining and visualizing the simulated molecular trajectory based on mapping the synthesized atomic structures to a low-dimensional space using PCA (the inset shows that the View list should be switched to All to access this protocol). Additionally, images can be synthesized from an imported set of conformations (e.g., from an MD trajectory). A similar protocol exists for synthesizing subtomograms, but using the graphical interface of HEMNMA-3D.

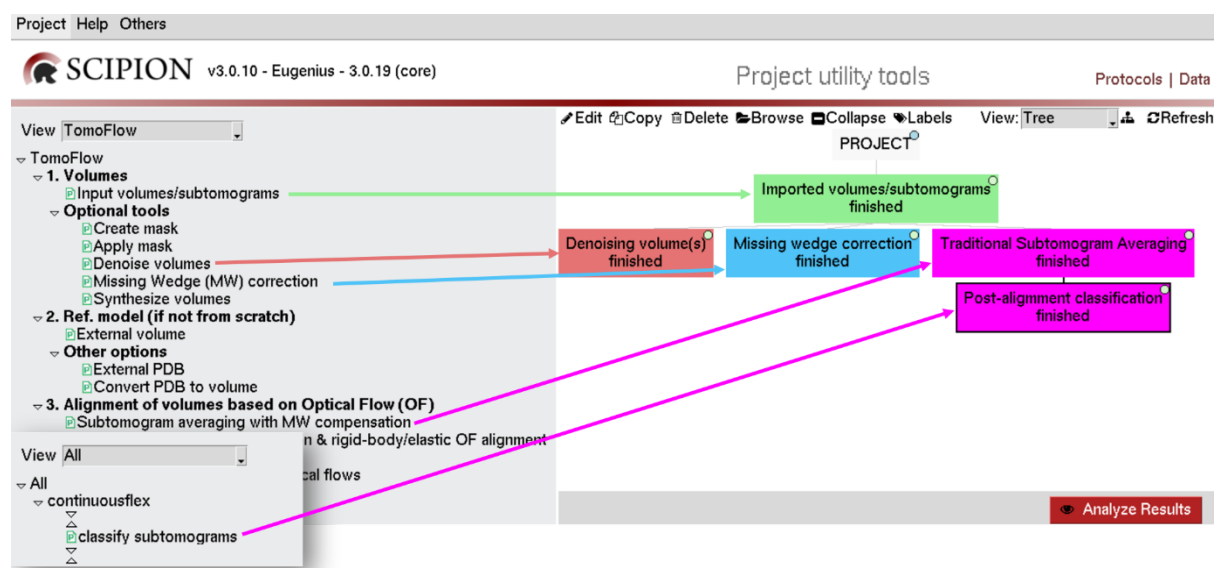
## 5.2. Optional subtomogram (pre)processing

In this subsection, we discuss additional tools for subtomogram (pre)processing that are available in ContinuousFlex. For instance, optional subtomogram averaging and classification, missing wedge correction, and volume denoising tools can be accessed using the graphical interface of TomoFlow (**Figure 9**).

The Subtomogram Averaging protocol is mandatory in TomoFlow (**Section 3.2**), but can also be used ahead of optional traditional classification. The Subtomogram Averaging protocol has two modes of operation. The first is the mode as a wrapper for an iterative subtomogram alignment using FRM (Chen et al., 2013) and averaging, but allowing fast MPI-parallelized processing, which is an important difference with respect to the original FRM implementation. The second is the mode that allows importing results of the subtomogram averaging performed with other software packages, in particular Dynamo (Navarro et al., 2018) and Artiatomi (<https://github.com/uermel/Artiatomi>). The Subtomogram Classification protocol allows a classification of the subtomograms that were rigid-body aligned using the Subtomogram Averaging protocol (post-alignment classification), and is based on a re-implementation of the work in (Förster et al., 2008).

The default Missing Wedge Correction protocol in ContinuousFlex is based on filling the missing wedge region in the subtomogram with the corresponding region of the global subtomogram average, and is mandatory in TomoFlow (**Section 3.2**). ContinuousFlex also contains an optional Missing Wedge Correction protocol that is based on Monte Carlo Framework (Moebel and Kervrann, 2020).

Finally, the Volume Denoising protocol in ContinuousFlex allows denoising subtomograms using the method BM4D (Maggioni et al., 2013).



**Figure 9** Graphical interface for accessing optional data (pre)processing tools in ContinuousFlex. Green box: input set of subtomograms. Red box: Volume Denoising protocol using BM4D. Blue box: Missing Wedge Correction protocol using the Monte Carlo framework. Magenta boxes: traditional Subtomogram Averaging protocol using FRM followed by Post-alignment Classification protocol (the inset shows that the View list should be switched to All to access the Post-alignment Classification protocol).

## 6. Dependencies

For information on ContinuousFlex installation using Scipion plugin manager and the tools provided for testing each of the ContinuousFlex methods, the reader is referred to **Supplementary Material**.

To get the full functionality of ContinuousFlex, it is required to additionally install Xmipp, ChimeraX (Goddard et al., 2018), and VMD (Humphrey et al., 1996). Indeed, the backend code of several methods of ContinuousFlex are contained in Xmipp, and several visualization steps of ContinuousFlex are done using ChimeraX and VMD. Xmipp and ChimeraX can be installed using Scipion plugin manager, as they are already contained in Scipion. VMD can be installed by following the VMD online documentation (Humphrey et al., 1996).

The backend code of the optional Monte-Carlo missing wedge correction and the optional BM4D volume denoising is in MATLAB and requires the MATLAB Image Processing Toolbox. If these two methods are not planned to be used, MATLAB is not required. The default ContinuousFlex method for missing wedge correction (filling the missing wedge region with the one of the subtomogram average) does not require MATLAB.

## 7. Space requirement and performance

Beside the disk space for keeping the input data (e.g., single particle images or subtomograms), running the methods of ContinuousFlex requires little additional disk space. The data analysis is done iteratively, but in the majority of cases, the intermediate results produced in one iteration are overwritten in the next iteration. Besides, in the majority of cases, the intermediate results produced in an iteration are not image or volumetric data but text files. Only TomoFlow keeps volumetric data on the disk (3D optical flows and missing-wedge corrected subtomograms; up to 6 times the size of the input subtomograms). An improvement of TomoFlow space requirement will be addressed in future releases.

The performance of each method, on the hardware accessible to us, was discussed in the corresponding articles where the methods were introduced. Here, we provide a brief summary. HEMNMA, HEMNMA-3D, and NMMD require CPU processing, whereas DeepHEMNMA and TomoFlow require both CPU and GPU processing. It should be noted that only the most time-consuming steps of the methods are parallelized. For instance, the data analysis (particle images, subtomograms, or cryo-EM maps) is parallelized, but the normal mode calculation is not. The performance of the different methods is discussed only for these most time-consuming steps. The other steps include normal mode calculation, interactive normal mode analysis (selection of normal modes for data analysis), and interactive low-dimensional conformational space analysis (grouping similar conformations and calculating 3D reconstructions or subtomogram averages in this space).

The subtomogram analysis (HEMNMA-3D and TomoFlow) and the cryo-EM map fitting (NMMD) can be performed on local workstations, such as a workstation with one NVIDIA Quadro RTX 5000 GPU card (16 Gb memory), two 2.2 GHz Intel Xeon Silver 4214 CPU processors, and 64 GB RAM (use of 1 thread per CPU core, meaning the maximum of 24 CPU threads on this workstation). For instance, 3 iterations of TomoFlow with a set of 1000 subtomograms of size 64 x 64 x 64 voxels took around 36 hours on the GPU card and 20 CPU threads of such workstation. With the same dataset and 20 CPU threads on the same workstation, HEMNMA-3D took around 8 hours using 3 normal modes and 25 hours using 6 normal modes. On the same hardware, NMMD took between 10 minutes and 30 hours for fitting cryo-EM maps (size 100 x 100 x 100 voxels or 128 x 128 x 128 voxels) of complexes of the molecular weights between 25 kDa and 600 kDa, respectively, using 16 replicas and a single thread per replica.

The single particle image analysis (HEMNMA and DeepHEMNMA) requires larger CPU and GPU clusters or supercomputers, where the number of CPU threads and GPU cards can be chosen to balance the number of images that should be processed. For instance, processing a set of 20,000 images with HEMNMA and 3 normal modes took around 16 hours for images of size 256 x 256 pixels and around 8 hours for images of size 128 x 128 pixels on 160 CPU threads distributed over 6 nodes with two 2.6 GHz Intel Xeon E5-2690V4 CPU processors and 64 GB RAM per node (use of 1 thread per CPU core, meaning the maximum of 28 CPU threads per node). Training of the neural network of DeepHEMNMA to learn 3 conformational parameters (the amplitudes of 3 normal modes) and 5 rigid-body parameters (3 Euler angles and 2 shifts) using a set of 14,000 images took around 85 hours for images of size 256 x 256 pixels and around 60 hours for images of size 128 x 128 pixels on a node with 4 NVIDIA V100 GPU cards (16 Gb memory per GPU). DeepHEMNMA inference of the conformational and rigid-body parameters from a set of 50,000 images using the trained neural network took around 22 minutes for images of size 256 x 256 pixels and 15 minutes for images of size 128 x 128 pixels on 1 NVIDIA V100 GPU card.

## 8. Discussion and Conclusion

This article presented the current state of ContinuousFlex that contains cryo-EM/ET data processing methods for analyzing continuous conformational variability of macromolecules, which is one of the most challenging problems in cryo-EM/ET. ContinuousFlex is available as a plugin of Scipion and allows reproducible research like other Scipion plugins. ContinuousFlex methods combine different techniques, including NMA and MD simulation, image processing, deep learning, computer vision, statistical analysis, and data visualization techniques. Each of these techniques alone has its strengths and weaknesses. Hence, ContinuousFlex encourages the development of hybrid methods that combine these different techniques. Finally, the rapid advancement in cryo-EM/ET is constantly brought to ContinuousFlex for its further improvement and development.

The details and full performance benchmarks of the methods for analyzing continuous conformational variability of macromolecules that are currently available in ContinuousFlex (HEMNMA, DeepHEMNMA, StructMap, HEMNMA-3D, TomoFlow, NMMD) have been published in separate articles (DeepHEMNMA is the most recent method, currently in press (Hamitouche and Jonic, 2022)). ContinuousFlex is provided to facilitate the use of the different methods by the community. The work of the researchers willing to use any of these methods is facilitated through the corresponding intuitive graphical user interface. The availability of the different methods at the same place (i.e., in ContinuousFlex) allows combining them more easily. This is particularly interesting for their use in integrative structural biology studies. Indeed, ContinuousFlex allows comparing the continuous conformational variability of molecular complexes *in vitro* and *in situ*, by analyzing single particle cryo-EM and cryo-ET data, and interpreting cryo-EM maps and cryo subtomogram averages in terms of atomic coordinates via flexible fitted models. Such an integrative study is out of scope of this article and it is planned for future work.

The majority of the ContinuousFlex methods perform better for collective, large-scale motions than for localized motions. This is particularly true for the methods based on normal mode analysis, such as HEMNMA, HEMNMA-3D, and DeepHEMNMA, because usually only low-frequency normal modes (collective, large-scale motions) are selected for data analysis in these methods to avoid unreasonably long computing time that would be required with all modes. This is important for small complexes, such as Pol  $\alpha$  – B complex, but also for large complexes, such as TBSV virus (Jin et al., 2014). Yet, in some cases, localized motions of smaller parts of complexes can be detected, like the motion of tRNA bound to 80S ribosome (article in press (Hamitouche and Jonic, 2022)). TomoFlow (not based on normal modes but on optical flow) involves a large degree of data smoothing, which may hinder the detection of motions of very small parts of complexes. However, global motions of very small complexes, such as nucleosomes, can be successfully detected with the methods based on either optical flow or normal modes (Harastani et al., 2021; Harastani et al., 2022). Beside normal modes, NMMD method involves MD simulation, which allows the users of NMMD to analyze data in terms of more localized motions. NMMD was tested with complexes of the molecular weight between 26 kDa and 542 kDa (Vuillemot et al., 2022).

The larger the datasets and the better the data quality, the higher the expected resolution. In particular, the dataset size and quality are critical for reaching higher resolutions of the subtomogram data analysis with HEMNMA-3D and TomoFlow. We obtained subtomogram averages of 2 nm resolution with these methods using around 600 subtomograms of nucleosomes *in situ* (Harastani et al., 2021; Harastani et al., 2022). A higher resolution would be expected with a larger number of such small complexes *in situ*.

The article shows the recommended strategy for using each of the different methods (the order of the steps that should be performed for each method). The shortcomings, if important to note, are mentioned for each method in its corresponding sub-section of Sections 2-5 (e.g., smoothing effects in TomoFlow, speed of HEMNMA, choice of normal modes, etc.). For more details, the reader is addressed to the original articles of the corresponding methods.

This article is expected to increase the visibility of various tools of ContinuousFlex and boost their use by the cryo-EM and cryo-ET community.

## Acknowledgments

We acknowledge the support of the French National Research Agency - ANR (ANR-20-CE11-0020-03 and ANR-19-CE11-0008-01 to S.J.), the Sorbonne University (2019 "Interface pour le Vivant" PhD scholarship grant to M.H.); CNRS PRC 2889 to S.J.; and the access to the HPC resources of CINES and IDRIS granted by GENCI (A0100710998R, A0100710998, A0070710998, AP010712190, AD011012188 to S.J.).

## Authors contributions

**Mohamad Harastani:** Conceptualization, Methodology, Software, Validation, Writing – original draft. **Rémi Vuillemot:** Conceptualization, Methodology, Software, Validation, Writing – original draft. **Ilyes Hamitouche:** Conceptualization, Methodology, Software, Validation, Writing – original draft. **Nima Barati Moghadam:** Software, Validation. **Slavica Jonic:** Conceptualization, Methodology, Software, Validation, Supervision, Writing – review & editing, Funding acquisition.

## Data availability statement

The original contributions presented in the study are included in the article/supplementary files. The code of ContinuousFlex is publicly available on Github (<https://github.com/scipion-em/scipion-em-continuousflex>) and is also part of the ContinuousFlex plugin of Scipion. All questions regarding the software or data availability can be addressed to the corresponding author.

## References

- Banerjee, S., Bartesaghi, A., Merk, A., Rao, P., Bulfer, S.L., Yan, Y., Green, N., Mroczkowski, B., Neitz, R.J., Wipf, P., Falconieri, V., Deshaies, R.J., Milne, J.L., Huryn, D., Arkin, M., Subramaniam, S., 2016. 2.3 A resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* 351, 871-875.
- Chen, M., Ludtke, S.J., 2021. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. *Nature Methods* 18, 930-936.
- Chen, Y., Pfeffer, S., Hrabe, T., Schuller, J.M., Förster, F., 2013. Fast and accurate reference-free alignment of subtomograms. *Journal of Structural Biology* 182, 235-245.
- Cox, M.A.A., Cox, T.F., 2008. Multidimensional Scaling, p. 315-347, in: C.-h. Chen, et al., (Eds.), *Handbook of Data Visualization*, Springer Berlin Heidelberg, Berlin, Heidelberg.
- Dashti, A., Mashayekhi, G., Shekhar, M., Ben Hail, D., Salah, S., Schwander, P., des Georges, A., Singharoy, A., Frank, J., Ourmazd, A., 2020. Retrieving functional pathways of biomolecules from single-particle snapshots. *Nat Commun* 11, 4734.
- de la Rosa-Trevín, J.M., Quintana, A., del Cano, L., Zaldívar, A., Foche, I., Gutiérrez, J., Gómez-Blanco, J., Burguet-Castell, J., Cuenca-Alba, J., Abrishami, V., Vargas, J., Otón, J., Sharov, G., Vilas, J.L., Navas, J., Conesa, P., Kazemi, M., Marabini, R., Sorzano, C.O.S., Carazo, J.M., 2016. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *Journal of Structural Biology* 195, 93-99.
- Farneback, G. 2003. Two-Frame Motion Estimation Based on Polynomial Expansion, pp. 363-370, *In* J. Bigun and T. Gustavsson, (eds.) *Image Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Förster, F., Pruggnaller, S., Seybert, A., Frangakis, A.S., 2008. Classification of cryo-electron sub-tomograms using constrained correlation. *Journal of Structural Biology* 161, 276-286.



- Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E., 2018. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Science* 27, 14-25.
- Hamitouche, I., Jonić, S. 2021. Deep learning of elastic 3D shapes for cryo electron microscopy analysis of continuous conformational changes of biomolecules, pp. 1251-1255 2021 29th European Signal Processing Conference (EUSIPCO).
- Hamitouche, I., Jonic, S., 2022. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. *Front Mol Biosci* 9, 965645. (in press)
- Harastani, M., Jonic, S., 2021. Methods for analyzing continuous conformational variability of biomolecules in cryo electron subtomograms: HEMNMA-3D vs. traditional classification. *bioRxiv* DOI: 10.1101/2021.10.14.464366.
- Harastani, M., Sorzano, C.O.S., Jonić, S., 2020. Hybrid Electron Microscopy Normal Mode Analysis with Scipion. *Protein Science* 29, 223-236.
- Harastani, M., Eltsov, M., Leforestier, A., Jonic, S., 2021. HEMNMA-3D: Cryo Electron Tomography Method Based on Normal Mode Analysis to Study Continuous Conformational Variability of Macromolecular Complexes. *Frontiers in molecular biosciences* 8, 663121.
- Harastani, M., Eltsov, M., Leforestier, A., Jonic, S., 2022. TomoFlow: Analysis of Continuous Conformational Variability of Macromolecules in Cryogenic Subtomograms based on 3D Dense Optical Flow. *J Mol Biol* 434, 167381.
- He, K., Zhang, X., Ren, S., Sun, J. 2016. Deep residual learning for image recognition, pp. 770-778 Proceedings of the IEEE conference on computer vision and pattern recognition.
- Himes, B.A., Zhang, P., 2018. emClarity: software for high-resolution cryo-electron tomography and subtomogram averaging. *Nat Methods* 15, 955-961.
- Hofmann, S., Janulienė, D., Mehdipour, A.R., Thomas, C., Stefan, E., Brüchert, S., Kuhn, B.T., Geertsma, E.R., Hummer, G., Tampé, R., Moeller, A., 2019. Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* 571, 580-583.
- Huang, J., MacKerell Jr, A.D., 2013. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of computational chemistry* 34, 2135-2145.
- Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 14, 33-38.
- Jin, Q., Sorzano, Carlos Oscar S., de la Rosa-Trevín, José M., Bilbao-Castro, José R., Núñez-Ramírez, R., Llorca, O., Tama, F., Jonic, S., 2014. Iterative Elastic 3D-to-2D Alignment Method Using Normal Modes for Studying Structural Dynamics of Large Macromolecular Complexes. *Structure* 22, 496-506.
- Jonić, S., 2017. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. *Curr Opin Struct Biol* 43, 114-121.



- Jonić, S., Sorzano, C.Ó.S., 2016. Coarse-Graining of Volumes for Modeling of Structure and Dynamics in Electron Microscopy: Algorithm to Automatically Control Accuracy of Approximation. *IEEE Journal of Selected Topics in Signal Processing* 10, 161-173.
- Karanicolas, J., Brooks, C.L., 2003. Improved Gō-like Models Demonstrate the Robustness of Protein Folding Mechanisms Towards Non-native Interactions. *Journal of Molecular Biology* 334, 309-325.
- Kato, K., Miyazaki, N., Hamaguchi, T., Nakajima, Y., Akita, F., Yonekura, K., Shen, J.-R., 2021. High-resolution cryo-EM structure of photosystem II reveals damage from high-dose electron beams. *Communications Biology* 4, 382.
- Kobayashi, C., Jung, J., Matsunaga, Y., Mori, T., Ando, T., Tamura, K., Kamiya, M., Sugita, Y., 2017. GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *Journal of Computational Chemistry* 38, 2193-2206.
- Kulik, M., Mori, T., Sugita, Y., 2021. Multi-Scale Flexible Fitting of Proteins to Cryo-EM Density Maps at Medium Resolution. *Front Mol Biosci* 8, 631854.
- Kunz, M., Frangakis, A.S., 2017. Three-dimensional CTF correction improves the resolution of electron tomograms. *Journal of Structural Biology* 197, 114-122.
- Li, Z., Chen, S., Zhao, L., Huang, G., Pi, X., Sun, S., Wang, P., Sui, S.-F., 2022. Near-atomic structure of the inner ring of the *Saccharomyces cerevisiae* nuclear pore complex. *Cell research* 32, 437-450.
- Liao, M., Cao, E., Julius, D., Cheng, Y., 2013. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* 504, 107-112.
- Liu, Y.-T., Zhang, H., Wang, H., Tao, C.-L., Bi, G.-Q., Zhou, Z.H., 2021. Isotropic Reconstruction of Electron Tomograms with Deep Learning. *bioRxiv*, 2021.2007.2017.452128.
- Ma, J., 2005. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure* 13, 373-380.
- Maggioni, M., Katkovnik, V., Egiazarian, K., Foi, A., 2013. Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction. *IEEE Transactions on Image Processing* 22, 119-133.
- Miyashita, O., Kobayashi, C., Mori, T., Sugita, Y., Tama, F., 2017. Flexible fitting to cryo-EM density map using ensemble molecular dynamics simulations. *Journal of computational chemistry* 38, 1447-1461.
- Moebel, E., Kervrann, C., 2020. A Monte Carlo framework for missing wedge restoration and noise removal in cryo-electron tomography. *Journal of Structural Biology: X* 4, 100013.
- Mori, T., Terashi, G., Matsuoka, D., Kihara, D., Sugita, Y., 2021. Efficient Flexible Fitting Refinement with Automatic Error Fixing for De Novo Structure Modeling from Cryo-EM Density Maps. *Journal of Chemical Information and Modeling* 61, 3516-3528.
- Mori, T., Kulik, M., Miyashita, O., Jung, J., Tama, F., Sugita, Y., 2019. Acceleration of cryo-EM Flexible Fitting for Large Biomolecular Systems by Efficient Space Partitioning. *Structure* 27, 161-174.e163.

- Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P.M.G.E., Grigoras, I.T., Malinauskaite, L., Malinauskas, T., Miehl, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E.V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S.W., Chirgadze, D.Y., Murshudov, G., Aricescu, A.R., Scheres, S.H.W., 2020. Single-particle cryo-EM at atomic resolution. *Nature* 587, 152-156.
- Navarro, P.P., Stahlberg, H., Castaño-Díez, D., 2018. Protocols for Subtomogram Averaging of Membrane Proteins in the Dynamo Software Package. *Frontiers in molecular biosciences* 5, 82-82.
- Orzechowski, M., Tama, F., 2008. Flexible Fitting of High-Resolution X-Ray Structures into Cryoelectron Microscopy Maps Using Biased Molecular Dynamics Simulations. *Biophysical Journal* 95, 5692-5705.
- Punjani, A., Fleet, D.J., 2021. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of Structural Biology* 213, 107702.
- Sanchez Sorzano, Carlos O., Alvarez-Cabrera, Ana L., Kazemi, M., Carazo, Jose M., Jonić, S., 2016. StructMap: Elastic Distance Analysis of Electron Microscopy Maps for Studying Conformational Changes. *Biophysical Journal* 110, 1753-1765.
- Schur, F.K., Obr, M., Hagen, W.J., Wan, W., Jakobi, A.J., Kirkpatrick, J.M., Sachse, C., Kräusslich, H.G., Briggs, J.A., 2016. An atomic model of HIV-1 capsid-SP1 reveals structures regulating assembly and maturation. *Science* 353, 506-508.
- Shah, S.T.H., Xuezi, X., 2021. Traditional and modern strategies for optical flow: an investigation. *SN Applied Sciences* 3, 289.
- Sorzano, C.O., Jonic, S., Nunez-Ramirez, R., Boisset, N., Carazo, J.M., 2007. Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. *J Struct Biol* 160, 249-262.
- Sorzano, C.O.S., Jiménez, A., Mota, J., Vilas, J.L., Maluenda, D., Martínez, M., Ramírez-Aportela, E., Majtner, T., Segura, J., Sánchez-García, R., Rancel, Y., Del Caño, L., Conesa, P., Melero, R., Jonic, S., Vargas, J., Cazals, F., Freyberg, Z., Krieger, J., Bahar, I., Marabini, R., Carazo, J.M., 2019. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. *Acta Crystallogr F Struct Biol Commun* 75, 19-32.
- Strelak, D., Jiménez-Moreno, A., Vilas, J.L., Ramírez-Aportela, E., Sánchez-García, R., Maluenda, D., Vargas, J., Herreros, D., Fernández-Giménez, E., de Isidro-Gómez, F.P., Horacek, J., Myska, D., Horacek, M., Conesa, P., Fonseca-Reyna, Y.C., Jiménez, J., Martínez, M., Harastani, M., Jonić, S., Filipovic, J., Marabini, R., Carazo, J.M., Sorzano, C.O.S., 2021. Advances in Xmipp for Cryo–Electron Microscopy: From Xmipp to Scipion. *Molecules* 26, 6224.
- Suhre, K., Sanejouand, Y.-H., 2004. ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Research* 32, W610-W614.
- Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., Ludtke, S.J., 2007. EMAN2: An extensible image processing suite for electron microscopy. *J Struct Biol* 157, 38-46.

- Tirion, M.M., 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical review letters* 77, 1905.
- Trabuco, L.G., Villa, E., Mitra, K., Frank, J., Schulten, K., 2008. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. *Structure* 16, 673-683.
- Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J.H., Welsch, S., Blanc, F.E.C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., van Zandbergen, G., Landry, J., de Azevedo, N.T.D., Mosalaganti, S., Schwarz, A., Covino, R., Mühlebach, M.D., Hummer, G., Krijnse Locker, J., Beck, M., 2020. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* 370, 203-208.
- Vanden Berghen, F., Bersini, H., 2005. CONDOR, a new parallel, constrained extension of Powell's UOBYQA algorithm: Experimental results and comparison with the DFO algorithm. *Journal of Computational and Applied Mathematics* 181, 157-175.
- von Kügelgen, A., Tang, H., Hardy, G.G., Kureisaite-Ciziene, D., Brun, Y.V., Stansfeld, P.J., Robinson, C.V., Bharat, T.A.M., 2020. In Situ Structure of an Intact Lipopolysaccharide-Bound Bacterial Surface Layer. *Cell* 180, 348-358.e315.
- Vuillemot, R., Miyashita, O., Tama, F., Rouiller, I., Jonic, S., 2022. NMMD: Efficient Cryo-EM Flexible Fitting Based on Simultaneous Normal Mode and Molecular Dynamics atomic displacements. *Journal of Molecular Biology* 434, 167483.
- Wan, W., Kolesnikova, L., Clarke, M., Koehler, A., Noda, T., Becker, S., Briggs, J.A.G., 2017. Structure and assembly of the Ebola virus nucleocapsid. *Nature* 551, 394-397.
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems* 2, 37-52.
- Zhong, E.D., Bepler, T., Berger, B., Davis, J.H., 2021. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature Methods* 18, 176-185.

This is the Supplementary Material of an article accepted for publication in Journal of Structural Biology, <https://doi.org/10.1016/j.jsb.2022.107906>

## ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data

Mohamad Harastani, Rémi Vuillemot, Ilyes Hamitouche, Nima Barati Moghadam, Slavica Jonic\*

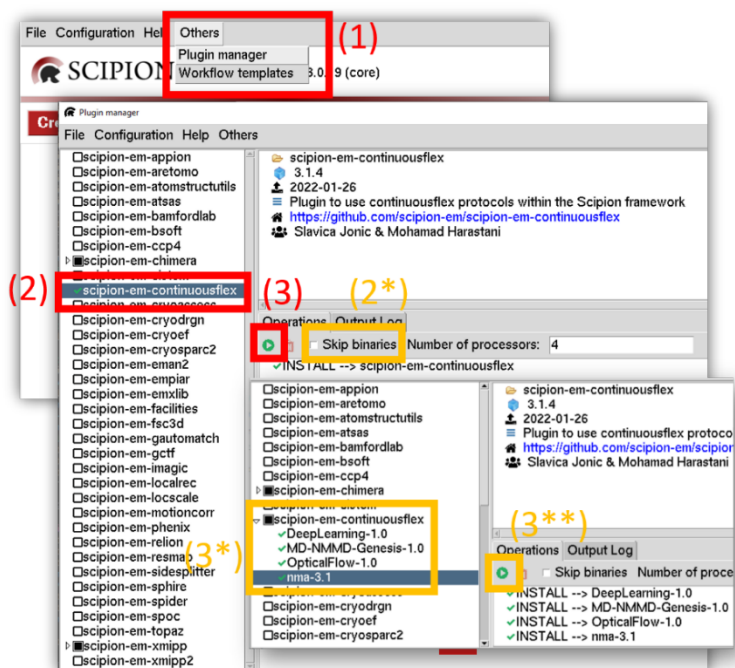
IMPMC-UMR 7590 CNRS, Sorbonne Université, Muséum National d'Histoire Naturelle, Paris, France

\*corresponding author: Slavica Jonic, IMPMC-UMR 7590 CNRS, Sorbonne Université, 4 Place Jussieu, 75005, Paris, France ([Slavica.Jonic@upmc.fr](mailto:Slavica.Jonic@upmc.fr))

### 1. ContinuousFlex installation, testing, and bug reports

#### 1.1. Installation

For information on installing Scipion and its plugins, the reader is referred to the Scipion documentation (<https://scipion-em.github.io/docs>). The ContinuousFlex installation steps using Scipion plugin manager are shown in **Supplementary Figure 1**.



**Supplementary Figure 1** Steps needed to install ContinuousFlex plugin in Scipion: (1) open the plugin manager, (2) tick the plugin name, (3) click the play button to install the plugin. In case of difficulties in installing the plugin or no intention to use all the functionalities of the plugin, the order (1, 2, 2\*, 3, 3\*, 3\*\*) may be tried by ticking only a subset of the toolboxes (binaries) that the install depends on for different methods.

## 1.2. Testing the methods contained in the plugin

Scipion allows testing the different methods it integrates through test modules, or simply "tests". Tests provide the following advantages: 1) ensuring a proper installation of the plugin and its configuration, 2) providing simple showcases for users, i.e., providing a sense for the choice of parameters of different protocols and the steps needed to follow to process data, and 3) allowing easier development and bug fixes (test-driven development).

In order to test all the functions of the ContinuousFlex plugin, it is possible to run the command "scipion3 tests continuousflex.tests". However, this test might take unnecessarily long if the user is interested in testing and using a single method, e.g., HEMNMA. In this case, the user can search for the specific test by first running the "scipion3 tests" command and finding the test corresponding to the method of interest, e.g., "scipion3 tests continuousflex.tests.test\_workflow\_HEMNMA".

The test will automatically create a Scipion project and run protocols for importing and processing data that can be visualized as in the case of projects that the user creates.

## 1.3. Bug reports and questions

Bug reports can be directly reported on the ContinuousFlex plugin page on GitHub (<https://github.com/scipion-em/scipion-em-continuousflex>). Other questions and inquiries can be sent to the corresponding author.