Hybrid Electron Microscopy Normal Mode Analysis graphical interface and protocol

CrossMark

Carlos Oscar S. Sorzano [a], José Miguel de la Rosa-Trevín [a], Florence Tama [b], Slavica Jonić [c,*]

[a] Biocomputing Unit, Centro Nacional de Biotecnología – CSIC, 28049 Madrid, Spain
[b] RIKEN, Advanced Institute for Computational Sciences, Kobe, Hyogo 650-0047, Japan
[c] IMPMC, Sorbonne Universités – CNRS UMR 7590, UPMC Univ Paris 6, MNHN, IRD UMR 206, 75005 Paris, France

ARTICLE INFO

ABSTRACT

This article presents an integral graphical interface to the Hybrid Electron Microscopy Normal Mode Analysis (HEMNMA) approach that was developed for capturing continuous motions of large macromolecular complexes from single-particle EM images. HEMNMA was shown to be a good approach to analyze multiple conformations of a macromolecular complex but it could not be widely used in the EM field due to a lack of an integral interface. In particular, its use required switching among different software sources as well as selecting modes for image analysis was difficult without the graphical interface. The graphical interface was thus developed to simplify the practical use of HEMNMA. It is implemented in the open-source software package Xmipp 3.1 (http://xmipp.cnb.csic.es) and only a small part of it relies on MATLAB that is accessible through the main interface. Such integration provides the user with an easy way to perform the analysis of macromolecular dynamics and forms a direct connection to the single-particle reconstruction process. A step-by-step HEMNMA protocol with the graphical interface is given in full details in Supplementary material. The graphical interface will be useful to experimentalists who are interested in studies of continuous conformational changes of macromolecular complexes beyond the modeling of continuous heterogeneity in single particle reconstruction.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

To perform biological functions, macromolecular assemblies change their shapes driven by different factors such as hydrogen bonding, ionic interactions, Van der Waals forces, hydrophobic packing and interactions with different partners. The alternative structures are referred to as different conformations, and the transitions between them conformational changes. Structural dynamics elucidation by determining the alternative conformations can be essential for understanding functions of large macromolecular complexes. Although single-particle electron microscopy (EM) has shown to be promising in capturing the alternative conformations (Elad et al., 2008; Fu et al., 2007, 2011; Grob et al., 2006; Penczek et al., 2006; Scheres et al., 2007; Simonetti et al., 2008), describing gradual (continuous) conformational changes is still a major challenge. This is the case when the complex cannot be stabilized in one or a few discrete conformational states or the

microscope is used to study full dynamics of the complex that is allowed to freely change its conformation.

In our previous work, we have shown that Normal Mode Analysis (NMA) of a reference structure can be used to analyze EM images of large macromolecular complexes and obtain an overall view of the conformational distribution for elucidating continuous conformational changes (Jin et al., 2014). In the approach referred to as Hybrid Electron Microscopy Normal Mode Analysis (HEMNMA), normal modes are used within elastic alignment of EM images with a reference structure (atomic-resolution structure or EM structure), without image classification or 3D reconstruction, as each image is allowed to contain a unique particle conformation. Structures can be reconstructed from images after discretizing the output, full conformational distribution. The full conformational distribution allows modeling of transition pathways and analyzing the conformational changes more extensively than with the classic, discrete methods (*e.g.*, Scheres et al., 2007), as the classic methods give only a few conformations. The discrete methods can be combined with HEMNMA to extend the analyses, as was shown in Jin et al. (2014).

Although HEMNMA is a powerful methodology, its usage has until now been limited to researchers who are at ease with

* Corresponding author at: IMPMC-UMR 7590, Université Pierre & Marie Curie, Case courrier 115, 4 Place Jussieu, 75005 Paris, France. Fax: +33 1 44 27 37 85.
E-mail address: Slavica.Jonic@impmc.upmc.fr (S. Jonić).

installation of different software on Linux operating system and learning how to use it without a user-friendly graphical interface or guidance. Indeed, HEMNMA required the usage of a collection of methods that were not coming from the same software package and that were run on the console without a graphical interface or, in some cases, graphical interface was available on web servers. For instance, the elastic image alignment method is a part of the open-source software package Xmipp since Xmipp version 3.0.1 (http://xmipp.cnb.csic.es; for general references to the package, please see de la Rosa-Trevin et al. (2013), Scheres et al. (2008), Sorzano et al. (2004b)). It used to be run without graphical interface, through a Linux command-line program obtained after the Xmipp compilation. The values of a number of parameters had to be set by the user according to the documentation provided on the Xmipp web site. Also, normal modes of atomic structures used to be computed either using *ElNemo* web server (Suhre and Sanejouand, 2004) or by locally compiling and running several programs involved in the *ElNemo* methodology. Normal modes of EM structures used to be computed either using *3DEM Loupe* web server (Nogales-Cadenas et al., 2013) or by locally compiling and running several programs involved in the *3DEM Loupe* methodology. The advantage of using the web servers was the available user-friendly graphical interface and user guidance. However, the web servers could provide help for only one step of the HEMNMA methodology. Also, the results obtained on the web server had to be integrated in the rest of the workflow of the HEMNMA methodology manually by the user.

The HEMNMA methodology involves other steps such as statistical analysis of the results of the elastic alignment, clustering, 3D reconstruction, and identification and animation of characteristic trajectories of the conformational changes. These steps used to be done in GNU Octave (Eaton et al., 2009), with scripts written by the user. Thus, the entire workflow used to be managed by the user fully and required his/her full attention.

In this article, we describe the graphical interface of HEMNMA that was recently developed to facilitate the practical use of the methodology. The interface simplifies the use of HEMNMA and helps understanding the workflows behind different HEMNMA steps. Also, it allows easier setting of values of different parameters of a run, and easier start and comparison of different runs. The main part of the graphical interface was implemented in Xmipp version 3.1. The source files for NMA of atomic and EM structures (Nogales-Cadenas et al., 2013; Suhre and Sanejouand, 2004) were incorporated into Xmipp (version 3.1) so that the NMA programs can now be accessed and run as any other Xmipp program. Only a part of the interface was implemented in MATLAB (MATLAB and Statistics Toolbox Release 2011b, The MathWorks, Inc., Natick, Massachusetts, United States), and it allows an interactive analysis of results of an advanced statistical analysis step. The graphical interface uses Chimera (Goddard et al., 2007; Pettersen et al., 2004; Yang et al., 2012) and VMD (Hsin et al., 2008; Humphrey et al., 1996) to visualize the input and output PDB-format files and EM density volumes, and output animations.

This integral graphical interface provides the experimentalists in single-particle EM with an easy way to explore continuous heterogeneity of samples and it forms a direct connection to the single-particle reconstruction process. As an integral tool for capturing continuous motions of macromolecular complexes from single-particle EM images, it is an important addition to the single-particle EM field.

In the article, the main workflow of the interface is described together with the logic implemented in three main modules, which is necessary to know for a correct interpretation and usage of the output of each module. A detailed step-by-step protocol for using HEMNMA with the graphical interface is given in Supplementary material. The usage is shown with a simple and intuitive synthetic data set of adenalyte kinase that was specifically designed so that the readers can easily catch the whole workflow and see how to interpret the results of many available graphics outputs.

2. Results and discussion

The developed graphical interface to HEMNMA comprises the following three modules: (1) *Normal Mode Analysis*; (2) *Flexible Alignment*; and (3) *Advanced Results Analysis*. The main workflow of this integral interface is shown in Fig. 1. The workflow requires an intermediate-resolution EM structure or atomic-resolution structure (reference structure) and a set of single particle images. The ML3D, Multireference Projection Matching or Relion methods (Scheres, 2012; Scheres et al., 2008, 2007) can be used to obtain a reference EM structure if the structure at atomic resolution is unavailable. ML3D and Multireference Projection Matching are a part of Xmipp and Relion can be accessed through it.

After computing and analyzing normal modes, in the *Normal Mode Analysis* module, the user can either stop the processing or continue it by selecting the normal modes to be used for image analysis. Finally, the user can perform the image analysis and visualize the image analysis results in the *Flexible Alignment* module. The image analysis results can further be analyzed using the *Advanced Results Analysis* module. One can run in parallel multiple copies of the same workflow using different data sets or parameters. The image analysis part of the *Flexible Alignment* module is MPI-parallelized, as it is the most time-consuming part of the workflow. The other parts of the workflow perform a single-core processing. We start this section by describing each of the three modules of the workflow and we finish it by showing an example of using the graphical interface and some of the most important graphics output.

2.1. Normal Mode Analysis module

The *Normal Mode Analysis* module allows computing normal modes of an input atomic or EM structure as well as an interactive visualization of the computed modes by animating displacements of the structure along the modes. Finally, it allows selecting the modes to be used in the *Flexible Alignment* module.

2.1.1. EM volume-to-pseudoatoms conversion

If an EM structure is used to compute normal modes, the EM volume is converted into a pseudoatomic structure with the method that also runs on *3DEM Loupe* web server (Nogales-Cadenas et al., 2013). The input volume is converted into a collection of Gaussian functions (pseudoatoms) with a given standard deviation (pseudoatom size) so that the pseudoatomic structure converted back to a volume approximates the input volume with a given error. In the context of using normal modes for image alignment, it has been shown that smaller radius and approximation errors (higher numbers of pseudoatoms) result in higher resolutions of computed projections, and thus, more precise image alignments (Jin et al., 2014). In the case of heavy background noise, a masked EM structure can be used for the volume-to-pseudoatoms conversion. The pseudoatomic structure and the computed modes are written in the format that is readable by the *Flexible Alignment* module.

2.1.2. Computation of normal modes, collectivity, and score

We use the elastic network model to compute normal modes, which does not require energy minimization (the reference conformation is considered to be at the energy minimum). 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
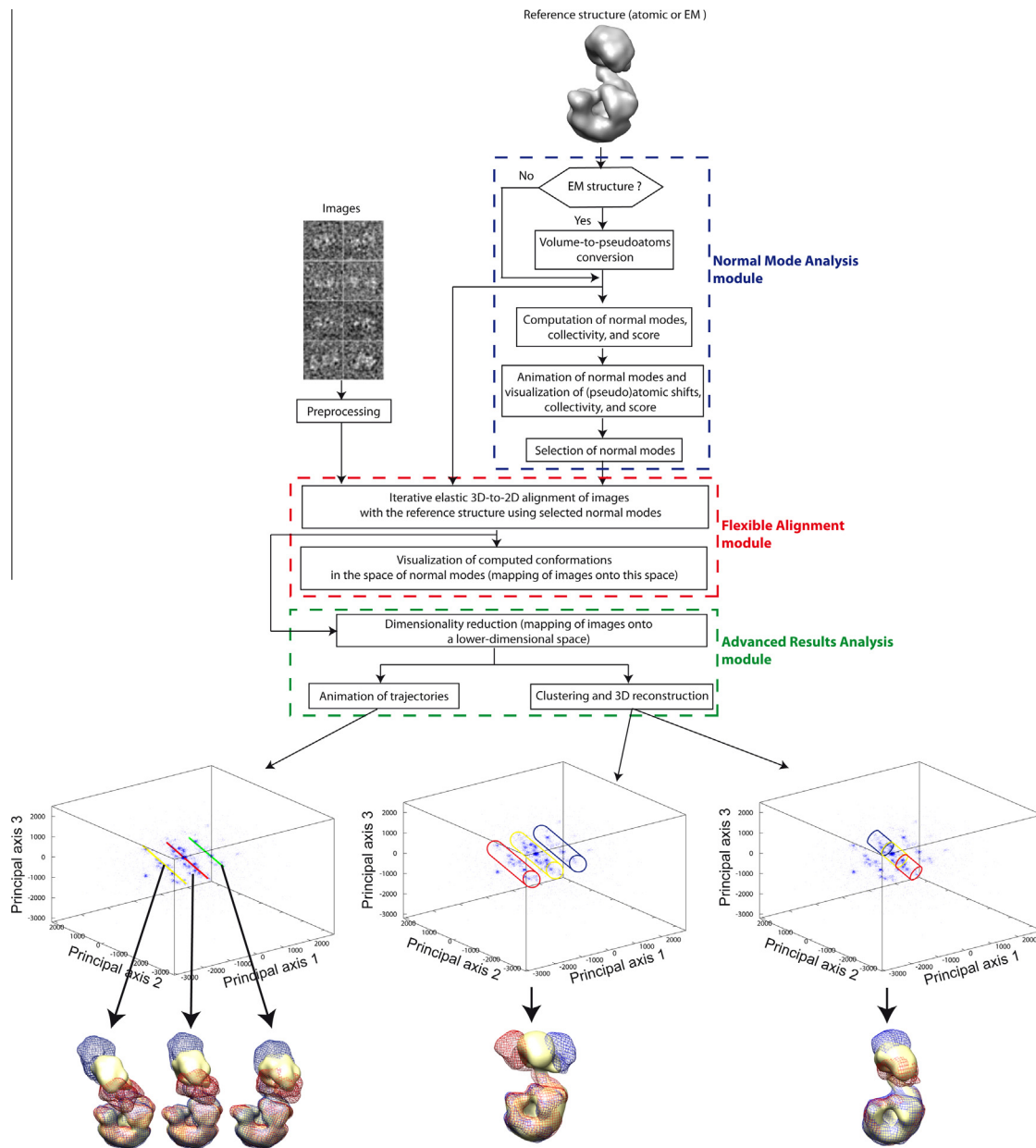
Fig.1. Flowchart of the graphical interface of HEMNMA comprising *Normal Mode Analysis*, *Flexible Alignme*nt, and *Advanced Results Analysis* modules. For illustration, the figure shows some of HEMNMA results obtained for DNA polymerase Pol α – B subunit complex of the eukaryotic primosome, together with the examples of images and the reference structure that were used in this case (reproduced with permission from Jin et al. (2014)).

number of nodes in the elastic network model. The nodes are atoms or pseudoatoms of the structure, and the maximum number of nodes is determined by the total number of atoms or pseudoatoms in the input PDB file. In the case of atomic structures, we use the rotation–translation block (RTB) method that divides the structure into blocks (one or a few consecutives residues per block) whose rotations and translations are considered only, rather than all degrees of freedom for all atoms (Durand et al., 1994; Tama et al., 2000). As the basis for the Hessian diagonalization is reduced in this case, the RTB method allows fast computing of normal modes. It has been shown that the RTB method provides a good approximation of low-frequency normal modes (Durand et al., 1994; Tama et al., 2000). The RTB method is also currently running on *ElNemo* web server (Suhre and Sanejouand, 2004). In the case of EM structures, normal modes are obtained by a direct diagonalization of the $3N \times 3N$ Hessian. Such calculation is performed with

ARPACK libraries (Lehoucq et al., 1998) and we refer to this method as the Cartesian method. This method was already used on EM structures (Chacon et al., 2003; Tama et al., 2002) and it is currently also running on *3DEM Loupe* (Nogales-Cadenas et al., 2013).

The accuracy of the computed normal modes will depend on the atomic or pseudoatomic interaction cutoff distance (the distance below which atoms or pseudoatoms do not interact). The cutoff distance can be given by the user or it can be computed automatically. The automatic computation is based on a given percentage of atomic or pseudoatomic distances that should be below the cutoff distance. The modes are computed along with their respective collectivity degrees. The collectivity degree counts the number of atoms or pseudoatoms that are affected by the mode, and it is computed as proposed in (Bruschweiler, 1995) (it is normalized between $1/N$ and 1 where N is the number of atoms or pseudoatoms; it approaches 1 for maximally collective motions and

approaches 0 for localized motions). Highly collective low-frequency modes have been shown to be relevant to functional conformational changes (Delarue and Dumas, 2004; Suhre et al., 2006; Tama and Sanejouand, 2001; Tama and Brooks, 2006; Tama et al., 2004; Wang et al., 2004). Two integer scores are attributed to each mode, to score the frequency and the collectivity of the mode (one score increases, from 1 to the total number of modes, as the mode frequency increases while the other increases, from 1 to the total number of modes, as the mode collectivity decreases). An overall normalized score, referred to as nmaScore, is then computed for each mode by summing the frequency and collectivity scores of the mode and dividing the sum by the total number of modes. Such score penalizes modes with higher frequencies and lower collectivities (such modes will be assigned higher scores) and translates the finding that the functional conformational changes can usually be described by highly collective low-frequency modes (Delarue and Dumas, 2004; Suhre et al., 2006; Tama and Sanejouand, 2001; Tama and Brooks, 2006; Tama et al., 2004; Wang et al., 2004).

2.1.3. Animation of normal modes and visualization of (pseudo) atomic shifts, collectivity, and score

A visualization module was developed as a part of the *Normal Mode Analysis* module to allow checking the computed normal modes, the normalized atomic or pseudoatomic shifts that the modes produce on the structure, the collectivity and the score for each mode. If an EM structure was used to compute normal modes, this visualization module also allows checking the volume-to-pseudoatoms conversion results, which is important for evaluating quality of the computed normal modes. The module thus allows Chimera visualization of the overlap between the input EM volume and the computed pseudoatomic structure. In the same context, the slices of the input EM volume and those of the volume approximated by the pseudoatomic structure are allowed to be visualized along all three axes, using the standard visualization program in Xmipp.

The computed modes can be evaluated visually, by animating the displacement of the reference structure along the specified mode for a given displacement amplitude. The animated displacements are written in the form of multiblock PDB files, known also as animated PDB files, using a modified version of the script "animate.py" from the Python-based program package ModeHunter (Stember and Wriggers, 2009). The animated PDB files are played in VMD. The structure can be rotated, shifted, or zoomed during the animation in VMD. By analogy with the visualization of pseudoatomic structures, the atomic-resolution structure is represented by a set of spheres for a faster interactive display and animation. The atom/pseudoatom representation can be changed inside VMD, using the VMD tools.

The animated modes represent the predicted movements for the given displacement amplitude. The displacement amplitudes along normal modes have no physical units and the displacement amplitude for the modes animation should be selected according to the size of the studied complex so that a good visibility of the movements is obtained while avoiding degenerated movements of the complex. The *Normal Mode Analysis* module can be relaunched with increased or decreased values of the displacement amplitude if one wants to see the same movement with larger or smaller amplitudes. When rerunning the module, the only steps that are repeated are those that involve the modified parameters. For instance, when entering a new value of the displacement amplitude for the animation of modes, only the animated PDB files are recomputed while the other steps are skipped (volume-to-pseudoatoms conversion in the case of EM structure and the computation of normal modes).

The computed modes can also be explored by analyzing the atomic/pseudoatomic shifts along the specified mode as well as the maximum atomic/pseudoatomic shifts over all computed modes, which are given in the form of 1D plots. These plots can help detect the atoms/pseudoatoms with localized movements, which can be interpreted as a normal behavior (*e.g.*, binding of a factor to a ribosomal complex can be responsible for localized conformational changes such as manly movement of the binding factor) or as an abnormal behavior (*e.g.*, when a localized movement can only be explained by a low-quality of the elastic network caused by specifying inappropriate values of the atomic/pseudoatomic interaction cutoff).

The collectivity and the score can be visualized by displaying the output modes metadata file. This file contains the file names of different modes (the column with the label "nmaModefile"), their current status as "selected" or "deselected" (the column "enabled"), the collectivity (the column "nmaCollectivity"), and the score (the column "nmaScore"). All modes with the collectivity above a given threshold value will have the status "selected" except the first six modes that are related to rigid-body movements. This "selection" status can later be modified.

2.1.4. Selection of normal modes for image analysis

To reduce the image analysis time in the next module (*Flexible Alignment*), the elastic alignment can be done using a subset of the computed normal modes. This can be done by modifying the modes "selection" status in the output modes metadata file. The interface allows saving the modified modes metadata file. To decide which modes to select, one can use a previous knowledge about possible movements or the highest-collectivity criterion, as proposed in (Jin et al., 2014). Additionally, the interface allows using a least-score criterion (as a general rule, modes with low nmaScore are preferred because they are highly collective and have low frequency) or any combination of these three criterions to select modes.

2.2. Flexible Alignment module

The *Flexible Alignment* module allows an iterative elastic 3D-to-2D alignment of images with the reference structure, using the selected normal modes. Also, it allows visualizing the computed conformations in the space of normal modes by mapping images onto this space.

2.2.1. Iterative elastic 3D-to-2D alignment

To determine the conformation in the given single-particle image and the image orientation and translation, the image is compared with projections of possible conformations. The possible conformations are obtained by displacing atomic or pseudoatomic reference coordinates using a combination of selected normal modes. The data required for the alignment are thus the images, the reference structure, and the selected modes of the reference structure. The results of the alignment are the parameters of the best matching conformation and projection for each single-particle image. The parameters are the displacement amplitude along each used normal mode (elastic parameters) and the image orientation and translation (rigid-body parameters). The alignment parameters are obtained with a combination of a local-search trust-region-based Powell's UOBYQA optimization (unconstrained optimization in the absence of derivatives of the objective function (Powell, 2002)), to determine elastic parameters, and a projection-matching-like strategy to determine rigid-body parameters.

The graphical interface allows using the following two projection-matching-like strategies: (1) *Wavelets-and-Splines*, as proposed in Jin et al. (2014), and (2) *Projection Matching-and-Splines* (optional strategy). In both strategies, a global search for

the rigid-body parameters is done using a library of reference projections, and so obtained initial alignment is then refined locally in a continuous, spline framework using the method based on the Fourier central-slice theorem, as proposed in Jonic et al. (2005). In the *Wavelets-and-Splines* strategy, the initial alignment is done in wavelet space using the method proposed in Sorzano et al. (2004a). In the *Projection Matching-and-Splines* strategy, the initial alignment is done in real space using the method proposed in Scheres et al. (2008). The library of reference projections is computed by sampling 3D space quasi-uniformly, with a given angular sampling step. In the *Wavelets* method, the search for the in-plane angle is done with the step of 5°. In the *Projection Matching* method, there is a continuous search of the in-plane rotation by exploiting the correlation property of the Fourier transform in polar coordinates. In both methods, the search is done in 5D space, the maximum absolute value of the expected shift is one tenth of the image size at the full-resolution level, and the shift search is done with the step of 1 pixel. The advantage of *Wavelets-and-Splines* is a better robustness to noise, as the alignment is done in the wavelet space (Sorzano et al., 2004a). The advantage of *Projection Matching-and-Splines* is a faster processing, as the library of reference projections is computed in Fourier space (by slicing the volume) and the in-plane shifts and rotations are also estimated in Fourier space (by exploiting the Fourier correlation property) (Scheres et al., 2008). A limitation of *Wavelets-and-Splines* is that the wavelet transform requires a power of 2 pixels for the image size, which may require some image rescaling before entering the *Flexible Alignment* module.

The current implementation of the UOBYQA method is based on CONDOR libraries (Berghen and Bersini, 2005). The method is initialized by specifying an initial vector of deformation amplitudes (initial point) and the initial and final values, ρ_{start} and ρ_{end} of the trust region radius ρ ($\rho_{start} \geqslant \rho \geqslant \rho_{end}$). The radius ρ is a typical distance between successive points at which the objective function is evaluated and its value is reduced automatically as soon as the objective function stops decreasing (Powell, 2002). For noisy objective functions, noise contribution to the estimates of first and second derivatives of the objective function can thus be controlled by controlling the distances between the points (Powell, 2002). The method uses an additional trust region radius Δ, satisfying $\Delta \geqslant \rho$, that is adjusted as it is traditionally done in trust region methods. The advantage of introducing Δ is to allow the changes of the deformation amplitudes to exceed ρ, which helps to avoid some loss of efficiency that may occur otherwise (ρ is never increased by the algorithm because the increases would necessitate more decreases later, which is usually onerous) (Powell, 2002). The initial value for Δ is equal to ρ_{start}.

The processing is done in a coarse-to-fine multi-resolution fashion using two-level multi-resolution image and volume pyramids for speed and robustness-to-noise reasons (Thevenaz et al., 1998). To initialise the first iteration at the coarse-resolution data level (Level 1), the current deformation amplitude for each mode is set to zero (initial point), which corresponds to the non-deformed reference structure. The initial and final values of the trust region radius at Level 1 are set to $\rho_{start} = 250$ and $\rho_{end} = 50$. The final vector of deformation amplitudes found at Level 1 is used as the initial point for the first iteration at the fine-resolution level (Level 0). Also, the final value of ρ from Level 1 is used to initialize the first iteration at Level 0. The initial value of ρ at Level 0 is reduced until $\rho_{end} = 0.5$. At Level 1, a global search for the rigid-body alignment parameters is followed by a local refinement. At Level 0, the global search is omitted and the rigid-body alignment parameters found at Level 1 are refined locally, after scaling up the translations by 2. Our experiments have shown that noisy, EM images should have at least the size of 32×32 pixels at the coarsest pyramid level, in order to provide resolution for a good

alignment. The input images should thus not be smaller than 64×64 pixels. The default values of ρ_{start} and ρ_{end} at Level 1 can be changed by modifying their scaling factor (here, referred to as trust region scale) whose default value is 1.

2.2.2. Visualization of computed conformations in the space of normal modes

The conformational parameters computed with the elastic 3D-to-2D alignment can be visualized in the space determined by the normal modes used for the alignment. The graphical interface allows displaying histograms, 2D scatter plots, or 3D scatter plots of the computed displacement amplitudes along the modes. The visualization is interactive in the case of 3D plots (3D plots can be rotated and zoomed).

2.3. Advanced Results Analysis module

We have created a data analysis tool based on MATLAB that is especially tailored to the exploration of deformation amplitudes (displacements along modes) computed with the *Flexible Alignment* module and to modeling of pathways of conformational changes from such data. The MATLAB-based tool has a graphical interface designed with the MATLAB tool GUIDE. To be able to use the MATLAB analysis tool, Xmipp has to be compiled so that it is aware that MATLAB is in the system. In this case, a binding function will be compiled and its path has to be added to the MATLAB path, so that MATLAB is able to read Xmipp metadata files. More details on the Xmipp compilation with MATLAB can be found in the installation instructions on the Xmipp web site (http://xmipp.cnb.csic.es).

The MATLAB-based tool allows visualizing the raw displacements in 1D, 2D or 3D (*i.e.*, histograms, 2D scatter plots or 3D scatter plots, respectively), as the visualization tool of the *Flexible Alignment* module. However, the MATLAB-based tool additionally allows coloring of points (images) in the scatter plots with the color corresponding to the alignment quality (the measure of similarity between the input image and the corresponding projection of the best-matching model) so that the user can easily identify which images do fit the deformed model and which images do not. The MATLAB-based tool also allows the definition of data filters for removing outlier points (images whose assigned deformation amplitudes are far from the main population), which is not possible with the visualization tool of the *Flexible Alignment* module. Most importantly, the MATLAB-based tool allows reducing the dimensionality of the raw data, mapping of images onto the lower-dimensional space, animating characteristic trajectories through the data, clustering, and 3D reconstruction from clusters.

2.3.1. Dimensionality reduction and visualization of computed conformations in the lower-dimensional space

The dimension of the raw data is determined by the number of selected normal modes. The raw data can be projected onto a lower-dimensional space using one of the following available dimensionality reduction techniques: Principal Component Analysis (PCA), Kernel PCA, Probabilistic PCA, Local Tangent Space Alignment (LTSA), Linear LTSA, Diffusion Map, Linearity Preserving Projection, Laplacian Eigenmap, Hessian Locally Linear Embedding, Stochastic Proximity Embedding, and Neighborhood Preserving Embedding (Van der Maaten et al., 2009). Some of these techniques are linear (with the advantage of being relatively simple) while others are non-linear (with the advantage of being well suited to represent non-linear manifolds). They are executed by calling an Xmipp program so that no additional MATLAB toolbox is required to be installed. The projection results can be exported into the MATLAB workspace for further processing such as clustering and 3D reconstruction. Once the raw data is projected onto the lower-dimensionality space, it is treated as in the case of raw data

(1D, 2D or 3D representations and data filters are available). The dimensionality reduction to two or three dimensions allows an easier global data display and analysis as well as an easier modeling of pathways of conformational changes as shown next.

2.3.2. Animation of trajectories through data

In 2D and 3D representations, characteristic trajectories through the data can be determined and, then, animated to see how these displacements act on the structure. To determine a trajectory, one selects two points from the data, which automatically creates 9 points on the line between the two given points. The points can be dragged to better fit the data on the 2D or 3D representation. To animate the trajectory, the reference atomic or pseudoatomic structure is displaced along the trajectory using VMD. The VMD software provides an interactive analysis of the structure while it is displacing along a trajectory (e.g., rotation, shift, and zoom). Also, the atomic/pseudoatomic visual representation can be modified in VMD and different VMD plugins can be used to better analyze the conformational change. The animation is saved as an animated PDB file that can directly be played with VMD. If the trajectory is drawn on the lower-dimensional space, a mechanism is required to convert it back to the original displacement space, for generating animation. This can be done with linear methods (PCA, Linear LTSA, Linearity Preserving Projection, Probabilistic PCA and Neighborhood Preserving Embedding) because there exists a matrix mapping from one space to the other.

2.3.3. Clustering and 3D reconstruction

The data can be clustered semi-automatically by specifying rules for data splitting, and the clusters can be saved for further 3D reconstruction purposes. This can be done in the lower- or higher-dimensional space. Then, the 3D reconstructions can be computed using the output metadata files with the rigid-body and elastic alignment parameters corresponding to each cluster. The reconstruction can be performed using any of the reconstruction methods available in Xmipp. To perform clustering or reconstruction using the methods implemented in other software packages, the output alignment files can be exported to be readable by these software packages. The reconstructed structures can be overlapped in Chimera to visualize the conformational change.

2.4. Example

Here, we show the most important graphical outputs returned by the software (Supplementary Fig. 1), when analyzing images with continuous conformational changes using an atomic reference structure. In this context, we highlight the output of the tools for the results visualization and analysis in the space of normal
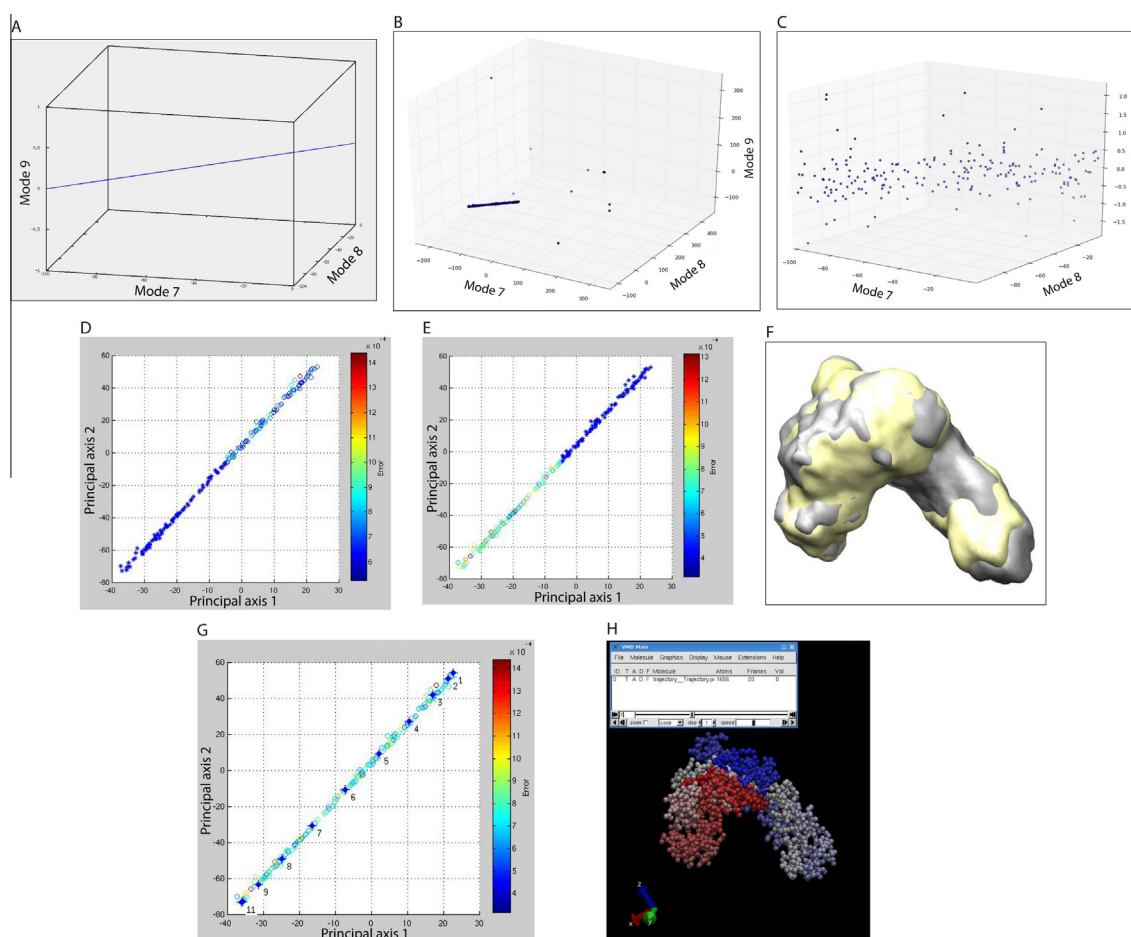

Fig.2. Examples of the most important outputs of the HEMNMA graphical interface shown with a synthetic data set. (A–C) Ground-truth deformation amplitudes (A), deformation amplitudes computed using *Projection Matching-and-Splines* strategy (B), and deformation amplitudes computed using *Wavelets-and-Splines* strategy (C), for 200 synthetic images represented by points (the plot in (A) was produced with GNU Octave and the plots in (B and C) were produced with the HEMNMA graphical interface). (D and E) Clustering of deformation amplitudes shown in (B), after removing the outlier points and projecting onto the first two principal axes (a cluster of images is shown by blue asterisks in each of the panels (D and E)). (F) Chimera visualization of the overlap of two structures reconstructed from the respective clusters (yellow and gray structures from the clusters shown in (D and E), respectively). (G) Trajectory drawing with nine points placed automatically between two manually drawn points (the position of these eleven points can be readjusted with the user's help, until the curve fits the data in all planes). (H) Animation of the reference structure along the trajectory shown in (G), using the VMD software. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

modes, in a lower-dimensional space, in the form of animated trajectories, and in the form of 3D reconstructions from clusters (Fig. 2), which are the outputs of the *Flexible Alignment* and *Advanced Results Analysis* modules. Other available graphical outputs are given in Supplementary material (*e.g.*, the output of the *Normal Modes Analysis* module, EM volume as the reference structure, etc.).

The data set used here consists of an atomic structure of adenalyte kinase (PDB entry 4AKE), its normal modes computed and visualized using the *Normal Mode Analysis* module (Supplementary Figs. 2–4), and synthetic images computed by randomly deforming 4AKE with a combination of normal modes and then projecting the deformed structure with random orientation and position onto the image plane. A total number of 200 synthetic images of size 128×128 pixels (pixel size: $(1 \text{ Å})^2$) were computed using the atomic structure 4AKE being displaced with a combination of modes 7–9. The displacement along modes 7 and 8 was done with the same random displacement (deformation) amplitude that was uniformly distributed between -100 and 0 while the deformation amplitude along mode 9 was 0 for all images. These ground-truth deformation amplitudes are shown in Fig. 2A and the corresponding ground-truth conformational change is shown in Supplementary Movies 1 and 2. The particles orientations and shifts were also randomly distributed (two out-of-plane rotations were distributed uniformly in the ranges [0°, 360°] and [0°, 180°], the in-plane rotation was 0°, and each of the two in-plane translations was distributed uniformly in the range [0,5] pixels).

To show how the ground-truth (known) deformation amplitudes (Fig. 2A) can be retrieved by analyzing images using HEMNMA graphical interface, we executed the *Flexible Alignment* module using, as the input data, the synthetic images, the 4AKE structure, and the modes file with selected modes 7–9 (Supplementary Fig. 5). The results of the iterative elastic 3D-to-2D alignment were then visualized for both *Projection Matching-and-Splines* and *Wavelets-and-Splines* strategies (Fig. 2B and C) using the visualization interface of the *Flexible Alignment* module (Supplementary Fig. 6). Then, the results were further analyzed with the *Advanced Results Analysis* module (Supplementary Figs. 7–11). As the *Projection Matching-and-Splines* strategy produced some outlier points with this synthetic data set (Fig. 2B), we give in Fig. 2D–H the graphics output when using this strategy, which shows the usefulness of the tool for rejecting outlier points. Note, however, that the results of the *Wavelets-and-Splines* strategy were significantly better (no outlier points, Fig. 2C). After rejecting the outlier points (Supplementary Figs. 7 and 8) and projecting the remaining points onto a 2D space by PCA (Supplementary Fig. 9), we separated images into two groups with a similar number of images (Fig. 2D and E and Supplementary Fig. 10) and computed one structure for each group, by 3D reconstruction from images of the group (Fig. 2F and Supplementary Figs. 12 and 13). One can note that the two overlapped structures (Fig. 2F) obtained by 3D reconstruction from such discrete classes (the classes that do not correspond to isolated clusters (Fig. 2D and E)) clearly reflect the main characteristics of the ground-truth continuous conformational change (Supplementary Movies 1 and 2). The animated reference structure along the trajectory identified through the data (Fig. 2G and H and Supplementary Fig. 11) showed much better the characteristics of a continuous conformational change. The movies animating the reference structure along the trajectory are not shown here as the movies are not significantly different from Supplementary Movies 1 and 2 which is not surprising as the computed amplitudes of displacement along modes coincide well with the ground-truth amplitudes (computed displacement amplitudes are very similar for modes 7 and 8, they are in the range [−100,0], and the displacements along mode 9 are very close to 0 (Supplementary Fig. 8B and D)).

3. Conclusions

In this paper, we described a user-friendly graphical interface for Hybrid Electron Microscopy Normal Mode Analysis (HEMNMA) that has been recently developed. The graphical interface makes easier the use of HEMNMA that relies on several different methods (Jin et al., 2014; Nogales-Cadenas et al., 2013; Suhre and Sanejouand, 2004) for two different types of input structure, atomic-resolution structure and EM structure. Also, it gives access to additional methods such as real-space projection matching for rigid-body alignment, scoring of modes for modes selection, methods for visualization of input and output data, and a MATLAB-based analysis tool for statistical analysis of obtained results and interactive modeling of pathways of conformational changes. Finally, it allows a direct start of different runs on MPI-parallelized PC machines (local computers, clusters, or supercomputers) and a direct use of results obtained with other methods accessible in Xmipp (*e.g.*, particles picking, CTF phase inversion, image resizing, 3D reconstruction, preliminary model computing by ML3D, Multireference Projection Matching or Relion methods in case of unavailable atomic-resolution structure etc.). Also, using a simple and intuitive synthetic data set of adenalyte kinase, we showed here some of the most important outputs that can be obtained with the interface while other available outputs can be found in Supplementary material, along with a detailed step-by-step protocol for using the interface.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2014.09.005.

References

Berghen, F.V., Bersini, H., 2005. CONDOR, a new parallel, constrained extension of Powell's UOBYQA algorithm: experimental results and comparison with the DFO algorithm. J. Comput. Appl. Math. 181, 157–175.

Bruschweiler, R., 1995. Collective protein dynamics and nuclear spin relaxation. J. Chem. Phys. 102, 3396–3403.

Chacon, P., Tama, F., Wriggers, W., 2003. Mega-Dalton biomolecular motion captured from electron microscopy reconstructions. J. Mol. Biol. 326, 485–492.

de la Rosa-Trevin, J.M., Oton, J., Marabini, R., Zaldivar, A., Vargas, J.M., Carazo, J.M., Sorzano, C.O., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J. Struct. Biol. 184, 321–328.

Delarue, M., Dumas, P., 2004. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. Proc. Natl. Acad. Sci. U.S.A. 101, 6957–6962.

Durand, P., Trinquier, G., Sanejouand, Y.H., 1994. A new approach for determining low-frequency normal modes in macromolecules. Biopolymers 34, 759–771.

Eaton, J.W., Bateman, D., Hauberg, S., 2009. GNU Octave Version 3.0.1 Manual: A High-level Interactive Language for Numerical Computations. CreateSpace Independent Publishing Platform.

Elad, N., Clare, D.K., Saibil, H.R., Orlova, E.V., 2008. Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. J. Struct. Biol. 162, 108–120.

Fu, J., Gao, H., Frank, J., 2007. Unsupervised classification of single particles by cluster tracking in multi-dimensional space. J. Struct. Biol. 157, 226–239.

Fu, J., Munro, J.B., Blanchard, S.C., Frank, J., 2011. Cryoelectron microscopy structures of the ribosome complex in intermediate states during tRNA translocation. Proc. Natl. Acad. Sci. U.S.A. 108, 4817–4821.

Goddard, T.D., Huang, C.C., Ferrin, T.E., 2007. Visualizing density maps with UCSF Chimera. J. Struct. Biol. 157, 281–287.

Grob, P., Cruse, M.J., Inouye, C., Peris, M., Penczek, P.A., Tjian, R., Nogales, E., 2006. Cryo-electron microscopy studies of human TFIID: conformational breathing in the integration of gene regulatory cues. Structure 14, 511–520.

Hsin, J., Arkhipov, A., Yin, Y., Stone, J.E., Schulten, K., 2008. Using VMD: an introductory tutorial. Curr. Protoc. Bioinformatics (Chapter 5, Unit 5.7).

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: visual molecular dynamics. J. Mol. Graph. 14 (33–38), 27–38.

Jin, Q., Sorzano, C.O., de la Rosa-Trevin, J.M., Bilbao-Castro, J.R., Nunez-Ramirez, 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.

Jonic, S., Sorzano, C.O., Thevenaz, P., El-Bez, C., De Carlo, S., Unser, M., 2005. Spline-based image-to-volume registration for three-dimensional electron microscopy. Ultramicroscopy 103, 303–317.

Lehoucq, R., Sorensen, D., Yang, C., 1998. ARPACK Users' Guide Society for Industrial and Applied Mathematics.

Nogales-Cadenas, R., Jonic, S., Tama, F., Arteni, A.A., Tabas-Madrid, D., Vazquez, M., Pascual-Montano, A., Sorzano, C.O., 2013. 3DEM Loupe: analysis of macromolecular dynamics using structures from electron microscopy. Nucleic Acids Res. 41, W363–367.

Penczek, P.A., Frank, J., Spahn, C.M., 2006. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. J. Struct. Biol. 154, 184–194.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera – a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.

Powell, M.J.D., 2002. UOBYQA: unconstrained optimization by quadratic approximation. Math. Program. 92, 555–582.

Scheres, S.H., 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 180, 519–530.

Scheres, S.H., Gao, H., Valle, M., Herman, G.T., Eggermont, P.P., Frank, J., Carazo, J.M., 2007. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods 4, 27–29.

Scheres, S.H., Nunez-Ramirez, R., Sorzano, C.O., Carazo, J.M., Marabini, R., 2008. Image processing for electron microscopy single-particle analysis using XMIPP. Nat. Protoc. 3, 977–990.

Simonetti, A., Marzi, S., Myasnikov, A.G., Fabbretti, A., Yusupov, M., Gualerzi, C.O., Klaholz, B.P., 2008. Structure of the 30S translation initiation complex. Nature 455, 416–420.

Sorzano, C.O., Jonic, S., El-Bez, C., Carazo, J.M., De Carlo, S., Thevenaz, P., Unser, M., 2004a. A multiresolution approach to orientation assignment in 3D electron microscopy of single particles. J. Struct. Biol. 146, 381–392.

Sorzano, C.O., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R., Scheres, S.H., Carazo, J.M., Pascual-Montano, A., 2004b. XMIPP: a new generation of an open-source image processing package for electron microscopy. J. Struct. Biol. 148, 194–204.

Stember, J.N., Wriggers, W., 2009. Bend-twist-stretch model for coarse elastic network simulation of biomolecular motion. J. Chem. Phys. 131, 074112.

Suhre, K., Navaza, J., Sanejouand, Y.H., 2006. NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. Acta Crystallogr. D Biol. Crystallogr. 62, 1098–1100.

Suhre, K., Sanejouand, Y.H., 2004. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. 32, W610–614.

Tama, F., Brooks, C.L., 2006. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu. Rev. Biophys. Biomol. Struct. 35, 115–133.

Tama, F., Gadea, F.X., Marques, O., Sanejouand, Y.H., 2000. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins 41, 1–7.

Tama, F., Miyashita, O., Brooks 3rd, C.L., 2004. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. J. Mol. Biol. 337, 985–999.

Tama, F., Sanejouand, Y.H., 2001. Conformational change of proteins arising from normal mode calculations. Protein Eng. 14, 1–6.

Tama, F., Wriggers, W., Brooks 3rd, C.L., 2002. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. J. Mol. Biol. 321, 297–305.

Thevenaz, P., Ruttimann, U.E., Unser, M., 1998. A pyramid approach to subpixel registration based on intensity. IEEE Trans. Image Process. 7, 27–41.

Van der Maaten, L., Postma, E., Van den Herik, J., 2009. Dimensionality Reduction: A Comparative Review. Tilburg Centre for Creative Computing, Tilburg University.

Wang, Y., Rader, A.J., Bahar, I., Jernigan, R.L., 2004. Global ribosome motions revealed with elastic network model. J. Struct. Biol. 147, 302–314.

Yang, Z., Lasker, K., Schneidman-Duhovny, D., Webb, B., Huang, C.C., Pettersen, E.F., Goddard, T.D., Meng, E.C., Sali, A., Ferrin, T.E., 2012. UCSF Chimera, MODELLER, and IMP: an integrated modeling system. J. Struct. Biol. 179, 269–278.