Elastic image registration to fully explore macromolecular dynamics by electron microscopy

Qiyu Jin, Carlos Oscar Sanchez Sorzano, Isabelle Callebaut, Florence Tama,
Slavica Jonic

► To cite this version:

Qiyu Jin, Carlos Oscar Sanchez Sorzano, Isabelle Callebaut, Florence Tama, Slavica Jonic. Elastic image registration to fully explore macromolecular dynamics by electron microscopy. 2014 IEEE International Conference on Image Processing (ICIP), Oct 2014, Paris, France. pp.2075 - 2079, 10.1109/ICIP.2014.7025416 . hal-01431471

HAL Id: hal-01431471

<https://hal.science/hal-01431471>

Submitted on 10 Jan 2017

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.

ELASTIC IMAGE REGISTRATION TO FULLY EXPLORE MACROMOLECULAR DYNAMICS BY ELECTRON MICROSCOPY

Qiyu Jin ¹, Carlos Oscar Sanchez Sorzano ², Isabelle Callebaut ¹, Florence Tama ³, and Slavica Jonic ^{1,*}

¹ IMPMC, Sorbonne Universités - CNRS UMR 7590, UPMC, MNHN, IRD, 75005 Paris, France.

² Biocomputing Unit, Centro Nacional de Biotecnología – CSIC, 28049 Madrid, Spain.

³ RIKEN, Advanced Institute for Computational Sciences, Kobe, Hyogo, 650-0047, Japan.

* Corresponding author (Slavica.Jonic@impmc.upmc.fr)

ABSTRACT

Structural changes are critical for biological functions of proteins and describing conformational changes in large macromolecular complexes is a major challenge. We have recently developed a hybrid method (HEMNMA) combining transmission electron microscopy (EM), normal mode analysis (NMA), and image analysis to study macromolecular dynamics. NMA is traditionally used to study macromolecular motions while HEMNMA provides insight into actual conformational changes seen by EM. HEMNMA uses normal modes to elastically align EM images with a reference structure in order to determine the conformations present in images and evaluate their pertinence. In this paper, we show how HEMNMA can be used with an atomic-resolution reference structure, using as an example the study of the conformational dynamics of Tomato Bushy Stunt Virus.

Index Terms— electron microscopy, elastic registration, structure, dynamics, macromolecules

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. To understand the functions of macromolecular assemblies, it is often necessary to elucidate their structural dynamics by determining the alternative conformations.

Three-dimensional (3D) structure of a wide range of biological macromolecular assemblies can be computed from two-dimensional (2D) images collected by transmission electron microscopy (EM). This information integrated with other structural data (e.g., from X-ray crystallography) helps structural biologists understand the function of macromolecular complexes. Single-particle analysis (SPA) is a method used for studies of macromolecular assemblies whose structure and dynamics can be analyzed in isolation (e.g., proteins, ribosomes, viruses) [2]. It is complementary to nuclear

magnetic resonance (NMR) since it allows computing the structure of large assemblies (diameter of 10-30 nm). It is also complementary to X-ray crystallography since it allows studying non-crystalline matter. Typically SPA uses a large number of images of randomly oriented individual macromolecular complexes to reconstruct the 3D structure. In practice, the analysis requires EM images of thousands of instances of the same complex captured in a unique conformation taken in random orientation. Complex image processing algorithms are required to compute the structure from the images. When the angular distribution of single-particle orientations samples Fourier space completely and the population is homogeneous, standard image processing strategies allow computing an average structure at resolution of 0.4-1 nm [3, 4].

In addition, SPA has shown to be promising in capturing alternative conformations of the same macromolecular complex [5-9]. To study macromolecular dynamics, currently standard methods separate the mixed particles into classes with similar conformations but random particles orientations and then analyze the 3D structures computed from the image classes. However, in the framework in which the conformational change is considered as a continuous movement, any kind of classification (in 2D or in 3D) samples the continuous movement into discrete, average conformations that are limited in resolution, which means that some different but close conformations may not be detected.

We recently developed Hybrid Electron Microscopy Normal Mode Analysis (HEMNMA) method that was specifically designed for studying gradual conformational changes where a discrete number of states cannot fully describe the continuum of conformations [1]. HEMNMA provides an overall view of the distribution of conformations without classification of images or 3D reconstruction. The method requires a structure of the complex in one conformation (e.g., structure obtained by X-ray crystallography or by EM) and EM images of the complex in multiple conformations. It uses Normal Mode Analysis (NMA) to predict possible motions from the reference structure that are then confronted with EM images to identify the motions taking actually place in the sample. The confrontation is done within an elastic 3D-to-2D alignment procedure that aligns EM images with the reference structure deformed using possible directions of the conformational change predicted by NMA. In this approach,

each image is analyzed as possibly containing a unique particle conformation.

NMA is known to be a powerful tool to predict functional motions and it is usually used within flexible 3D-to-3D alignment procedures [10]. A consistency of the conformations present in images with these predicted by NMA was shown earlier although only visually and using a small number of images [11]. On the contrary, HEMNMA provides an automated iterative framework allowing processing a high number of images to evaluate the pertinence of different conformations. The statistical analysis of these identified conformations allows modeling the deformation pathways compatible with the experimental data and analyzing the conformational changes more extensively than using the classic discrete methods.

In the previous paper, we showed the performance of HEMNMA with experimental data when using a reference structure obtained by EM [1]. In this paper, we show how HEMNMA can be used with an atomic-resolution reference structure, using as an example the study of the conformational dynamics of Tomato Bushy Stunt Virus (TBSV).

2. METHODS

We start this section by briefly describing the basic steps of HEMNMA method. In **Subsection 2.1**, we give more technical details on the used image analysis method. The method is schematically shown in **Fig. 1**. Each particle image is analyzed independently of the remaining images. The images can be preprocessed (e.g., phase correction of electron microscope contrast transfer function, image size reduction, etc). Normal modes are computed by NMA of a reference structure. Displacements of the reference structure with a combination of normal modes determine possible conformations. To determine the conformation of the complex in the given image, the possible conformations are confronted with the conformation actually contained in the image using an iterative elastic 3D-to-2D alignment. The results of the alignment procedure are the values of the elastic (normal-mode displacement amplitudes) and rigid-body (orientation and translation) parameters of the geometrical transformation for which the transformed reference structure projects onto the image plane producing an image that matches best the given particle image.

To minimize the image analysis time, the elastic alignment should be done using only a subset of the computed normal modes. If *a priori* knowledge about the conformational change is unavailable, the collectivity degree is computed to count the number of atoms that are significantly affected by the mode [12] (the collectivity degree approaches 1 for maximally collective movements, whereas it approaches 0 for localized motions). Then, the lowest-frequency modes with the collectivity degree above a threshold (e.g., 0.5-0.75) are selected as highly collective low-frequency modes have been shown to be relevant to functional conformational changes [13-16] (a few examples of this type of modes selection are given in [1]). Also, the modes can be selected based on some *a priori* knowledge about the conformational change (a few examples of this type of modes selection are given in [1] and an example is also given here).

Fig. 1: Flowchart of HEMNMA method.

The computed displacement amplitudes are the conformational parameters and can be plotted to show the distribution of images in the space determined by the analyzed normal mode directions or in the space determined by the principal component axes for the computed deformation amplitudes. A principal axis (principal component mode) is determined as a linear combination of normal modes using principal component analysis (PCA) of the computed deformation amplitudes.

To visualize the conformational changes on the structural level, the computed deformation amplitudes can be analyzed using one of the following two approaches. In the first approach, the deformation amplitudes along a trajectory in the NMA space or the PCA space are applied onto the reference structure directly, resulting in its deformation (displacement) along the trajectory. The computed deformations can also be visualized by 3D reconstruction from images separated by classifying the deformation amplitudes assigned to them (standard 3D reconstruction methods based on rigid-body orientation and translation parameters are then used to compute structures from the classes). Since different but close conformations may be lost through classification and 3D averaging, the first approach should be preferred over the second when using the method to fully explore the dynamics (no stabilization of the complex in one or a few conformational states).

2.1. Iterative elastic 3D-to-2D registration

The strategy uses Powell's UOBYQA method, to estimate the elastic parameters, and a gradient-based method referred to as Wavelets-and-Splines, to estimate rigid-body parameters. The UOBYQA method uses quadratic approximation for building a local model of the objective function [17]. The structure coordinates, modified for the last UOBYQA-estimated deformation amplitudes, are converted into a volume that is then aligned with the particle image using rigid-body projection matching by Wavelets-and-Splines (**Fig. 2**).

Fig. 2: Flowchart of the iterative 3D-to-2D alignment step of HEMNMA method shown in **Fig. 1**.

The processing is done in a coarse-to-fine multi-resolution fashion using a two-level multi-resolution image and volume pyramids for speed and robustness-to-noise reasons [18]. 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 final vector of deformation amplitudes from Level 1 is used as the initial point for the first iteration at the fine-resolution level (Level 0). The main algorithmic difference between the processing at the two levels concerns the rigid-body projection matching at Step 2. At Level 1, a global search for the rigid-body alignment parameters is followed by a local refinement (combined Wavelets and Splines strategies). At Level 0, the global search is omitted and the parameters found at Level 1 are refined locally (Splines strategy), after scaling up the translations by 2.

The Wavelets strategy computes roughly five rigid-body parameters, based on the maximum cross-correlation between the experimental image and a discrete library of projections of the reference volume [19]. This is done in 2D wavelet-transform domain that provides fast and robust-to-noise processing [19]. The directions of the reference projections sample the 3D space quasi-uniformly, with a given angular sampling step (typically, 10°). Typical values for the in-plane angle and shift steps are 5° and 1 pixel, respectively. The Splines strategy refines the parameters locally, in a continuous framework, thanks to the fast gradient-based iterative minimization of a least-squares measure of dissimilarity between the image and the volume projections, using central slice theorem and optimized B-spline interpolation in the Fourier domain [20]. The value of the dissimilarity measure is taken as the new evaluation of the objective function for UOBYQA.

3. EXPERIMENT

Swelling of many icosahedral plant viruses has been observed upon changes of pH and EDTA chelation of divalent cations at the interface between subunits (Ca^{2+} or Mg^{2+} that appear to play a critical role in virion stability) but the swelling mechanism is still not well understood [21]. In our previous work, we used HEMNMA to analyze TBSV cryo-EM images

with an 1.3 nm-resolution reference structure obtained by EM and with 4 modes selected using the collectivity threshold of 0.75 [1]. We demonstrated that the capsid globally remains symmetrical during the conformational change and the conformational change can be mainly described as a combination of an increase in the particle radius, a rotation of the subunits around the 5-fold symmetry axis, and an increase in the size of the intersubunit space at the subunits quasi-trimer [1], which is consistent with the models of the TBSV conformational change proposed in the literature [22-24]. Most importantly, the visualization of the full conformational distribution revealed many intermediate states of the change [1], which described more extensively the conformational heterogeneity that was detected previously [22].

Here, these findings were used as an *a priori* knowledge to select the modes to analyze TBSV images with an atomic-resolution reference structure (a coarse-grain model of a 0.29 nm-resolution X-ray crystal structure of the compact TBSV conformation (PDB entry 2TBV [25]) that contains 53700 Ca atoms). Normal modes for this structure were computed as described in [26].

Cryo-electron micrographs were acquired on a JEOL JEM 2100F microscope with an ultra high-resolution pole piece ($\text{Cs}=0.5$ mm). Samples with heterogeneous populations of TBSV particles (at different phases of swelling) were prepared for cryo-EM as described previously [22]. Individual particle images were extracted from digitized micrographs, scaled down to the size of 128^2 pixels (pixel size of 0.32 nm), and 4000 particles were randomly selected from different micrographs (different defocus values). A continuous particle size variation can be seen on a digitized micrograph (**Fig. 3a**). The images were analyzed using the combined modes 28 (related to the capsid expansion), 80 (related to the subunits movement away and towards the 5-fold symmetry axis), and 107 (related to the subunits rotation around the 5-fold symmetry axis), as these individual modes were found to be the only ones describing symmetrical capsid movements. The computed deformation amplitudes along the modes were analyzed by PCA. The images were classified into five classes with respect to the computed deformation amplitude along mode 28 (**Fig. 3b**).

Fig. 3: (a) A zone of a cryo-electron micrograph. (b) PCA of the computed deformation amplitudes. Each particle image is represented by a point and the classes of points are colored with different colors, from the smallest deformation amplitude along mode 28 (red) to the largest one (cyan). The number of images per class: 24% (red), 22.5% (magenta), 23.3% (yellow), 19% (green), and 11.2% (cyan). Blue asterisks: principal trajectory.

Fig. 4: Reference structure displaced along the principal trajectory shown in **Fig. 3b** ((a) and (b) are two frames of a recorded movie for the same view of the structure).

Fig. 5: Dynamics at the level of pentamers and quasi-trimers. (a) A-subunits pentamers extracted from the two structures shown in **Fig. 4** and overlapped, (b) A-B-C subunits quasi-trimers extracted from the two structures shown in **Fig. 4** and overlapped. Blue: from the structure shown in **Fig. 4a**. Yellow: from the structure shown in **Fig. 4b**. Only S domains of the subunits are shown for a better visibility.

Fig. 6: Reconstructed structures from the classes shown in **Fig. 3b**, with the indicated intersubunit space and 5-fold symmetry axis. The same color code as in **Fig. 3b**. Only three overlapped volumes are shown for a better visualization of the structural differences.

The deformation amplitudes were fitted with a linear regression line in the space determined by the first two principal axes, and the reference structure was displaced along this line referred to as principal trajectory (**Fig. 3b** and **Fig. 4**). The conformational change can be mainly described as a combination of a change in the radius of the capsid and a rotation of the subunits around the 5-fold symmetry axis (**Figs. 4-5**). Inherent to the increase in the radius and the pentamer rotation is an increase in the size of the space between the three types of subunits in the subunits quasi-trimer (**Fig. 5b**). The RMSD between the atomic structures at the two ends of the animated line is 2.1 nm.

The overlapped volumes reconstructed from the image classes also show an increase in the size of the virus and in the size of the intersubunit space (**Fig. 6**). However, the pentamers rotation is less visible on the overlapped volumes, which may be explained by a lower resolution due to mixing of differently rotated pentamers in the same volume reconstruction.

These findings, consistent with the results of our previous study [1], complement the previous results by providing information about the conformational change at the atomic level.

4. CONCLUSION AND FUTURE WORK

This paper shows how our new method, HEMNMA, can be used to study gradual conformational transitions at atomic-resolution level. We showed the study of conformational dynamics of a virus but the method can be applied to any complex that is suitable for electron microscopy, including asymmetric structures [1].

HEMNMA accuracy was shown to be good in the case of synthesized very noisy (SNR=0.1) and much defocused (defocus=2 μ m) images (RMSD between the computed and ground-truth conformations is below 0.1 nm) [1]. In real-data case, the ground truth is unknown and accuracy cannot be evaluated, but it may be lower than in the synthetic case (imperfect noise and image formation models). To improve accuracy, we plan to test other optimization methods, such as simultaneous gradient-based estimation of all (elastic and rigid-body) alignment parameters in combination with global optimization for the initial alignment.

Elastic registration is the most time consuming step of HEMNA. It currently takes around 2 minutes per mode for one image (image size 128x128 pixels, Xeon 3.00GHz processor, one image per computing core), and its speed up is also planned for the future work.

5. ACKNOWLEDGMENTS

The work was partially funded by the CNRS (France) and the CSIC (Spain) [PICS 2011]; the French National Research Agency ANR [ANR-11-BSV8-010-04]; the European Social Fund and the Ministerio de Educación y Ciencia; the Spanish Ministry of Economy and Competitiveness [AIC-A-2011-0638 and BIO2010-16566]; the Comunidad de Madrid [CAM S2010/BMD-2305]; and FOCUS Establishing Supercomputing Center of Excellence (Japan).

6. REFERENCES

- [1] Q. Jin, C. O. S. Sorzano, J. M. de la Rosa-Trevin, J. R. Bilbao-Castro, R. Nunez-Ramirez, O. Llorca, F. Tama, and S. Jonic, "Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes," *Structure*, vol. 22, pp. 496-506, 2014.
- [2] J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in Their Native State*. New York: Oxford University Press, 2006.

- [3] M. Cotteville, E. Larquet, S. Jonic, M. V. Petoukhov, G. Caprini, S. Paravisi, D. I. Svergun, M. A. Vanoni, and N. Boisset, "The subnanometer resolution structure of the glutamate synthase 1.2-MDa hexamer by cryoelectron microscopy and its oligomerization behavior in solution: functional implications," *J Biol Chem*, vol. 283, pp. 8237-49, 2008.
- [4] S. J. Ludtke, M. L. Baker, D. H. Chen, J. L. Song, D. T. Chuang, and W. Chiu, "De novo backbone trace of GroEL from single particle electron cryomicroscopy," *Structure*, vol. 16, pp. 441-8, 2008.
- [5] N. Elad, D. K. Clare, H. R. Saibil, and E. V. Orlova, "Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections," *J Struct Biol*, vol. 162, pp. 108-20, 2008.
- [6] P. Grob, M. J. Cruse, C. Inouye, M. Peris, P. A. Penczek, R. Tjian, and E. Nogales, "Cryo-Electron Microscopy Studies of Human TFIIID: Conformational Breathing in the Integration of Gene Regulatory Cues," *Structure*, vol. 14, pp. 511-20, 2006.
- [7] P. A. Penczek, J. Frank, and C. M. Spahn, "A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation," *J Struct Biol*, vol. 154, pp. 184-94, 2006.
- [8] S. H. Scheres, H. Gao, M. Valle, G. T. Herman, P. P. Eggermont, J. Frank, and J. M. Carazo, "Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization," *Nat Methods*, vol. 4, pp. 27-9, 2007.
- [9] A. Simonetti, S. Marzi, A. G. Myasnikov, A. Fabbretti, M. Yusupov, C. O. Gualerzi, and B. P. Klaholz, "Structure of the 30S translation initiation complex," *Nature*, vol. 455, pp. 416-20, 2008.
- [10] F. Tama, O. Miyashita, and C. L. Brooks, 3rd, "Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM," *J Struct Biol*, vol. 147, pp. 315-26, 2004.
- [11] J. Brink, S. J. Ludtke, Y. Kong, S. J. Wakil, J. Ma, and W. Chiu, "Experimental verification of conformational variation of human fatty acid synthase as predicted by normal mode analysis," *Structure*, vol. 12, pp. 185-91, 2004.
- [12] R. Bruschweiler, "Collective protein dynamics and nuclear spin relaxation," *J Chem Phys*, vol. 102, pp. 3396-3403, 1995.
- [13] M. Delarue and P. Dumas, "On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models," *Proc Natl Acad Sci U S A*, vol. 101, pp. 6957-62, 2004.
- [14] F. Tama, O. Miyashita, and C. L. Brooks, 3rd, "Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis," *J Mol Biol*, vol. 337, pp. 985-99, 2004.
- [15] F. Tama and Y. H. Sanejouand, "Conformational change of proteins arising from normal mode calculations," *Protein Eng*, vol. 14, pp. 1-6, 2001.
- [16] Y. Wang, A. J. Rader, I. Bahar, and R. L. Jernigan, "Global ribosome motions revealed with elastic network model," *J Struct Biol*, vol. 147, pp. 302-14, 2004.
- [17] M. J. D. Powell, "UOBYQA: unconstrained optimization by quadratic approximation," *Mathematical Programming*, vol. 92, pp. 555-582, 2002.
- [18] P. Thevenaz, U. E. Ruttimann, and M. Unser, "A pyramid approach to subpixel registration based on intensity," *IEEE Trans Image Process*, vol. 7, pp. 27-41, 1998.
- [19] C. O. S. Sorzano, S. Jonic, C. El-Bez, J. M. Carazo, S. De Carlo, P. Thevenaz, and M. Unser, "A multiresolution approach to orientation assignment in 3D electron microscopy of single particles," *J Struct Biol*, vol. 146, pp. 381-92, 2004.
- [20] S. Jonic, C. O. S. Sorzano, P. Thevenaz, C. El-Bez, S. De Carlo, and M. Unser, "Spline-based image-to-volume registration for three-dimensional electron microscopy," *Ultramicroscopy*, vol. 103, pp. 303-17, 2005.
- [21] J. Witz and F. Brown, "Structural dynamics, an intrinsic property of viral capsids," *Arch Virol*, vol. 146, pp. 2263-74, 2001.
- [22] R. Aramayo, C. Merigoux, E. Larquet, P. Bron, J. Perez, C. Dumas, P. Vachette, and N. Boisset, "Divalent ion-dependent swelling of tomato bushy stunt virus: a multi-approach study," *Biochim Biophys Acta*, vol. 1724, pp. 345-54, 2005.
- [23] J. Kruse, K. M. Kruse, J. Witz, C. Chauvin, B. Jacrot, and A. Tardieu, "Divalent ion-dependent reversible swelling of tomato bushy stunt virus and organization of the expanded virion," *J Mol Biol*, vol. 162, pp. 393-414, 1982.
- [24] I. K. Robinson and S. C. Harrison, "Structure of the expanded state of tomato bushy stunt virus," *Nature*, vol. 297, pp. 563-568, 1982.
- [25] S. C. Harrison, A. Olson, C. E. Schutt, F. K. Winkler, and G. Bricogne, "Tomato bushy stunt virus at 2.9 Å resolution," *Nature*, vol. 276, pp. 368-373, 1978.
- [26] F. Tama and C. L. Brooks, 3rd, "Diversity and identity of mechanical properties of icosahedral viral capsids studied with elastic network normal mode analysis," *J Mol Biol*, vol. 345, pp. 299-314, 2005.