



HAL
open science

MDTOMO method for continuous conformational variability analysis in cryo electron subtomograms based on molecular dynamics simulations

Rémi Vuillemot, Isabelle Rouiller, Slavica Jonić

► To cite this version:

Rémi Vuillemot, Isabelle Rouiller, Slavica Jonić. MDTOMO method for continuous conformational variability analysis in cryo electron subtomograms based on molecular dynamics simulations. *Scientific Reports*, 2023, 13, pp.10596. 10.1038/s41598-023-37037-9 . hal-04147646

HAL Id: hal-04147646

<https://hal.science/hal-04147646>

Submitted on 30 Jun 2023

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.



OPEN

MDTOMO method for continuous conformational variability analysis in cryo electron subtomograms based on molecular dynamics simulations

Rémi Vuillemot^{1,2}, Isabelle Rouiller^{2,3} & Slavica Jonić¹✉

Cryo electron tomography (cryo-ET) allows observing macromolecular complexes in their native environment. The common routine of subtomogram averaging (STA) allows obtaining the three-dimensional (3D) structure of abundant macromolecular complexes, and can be coupled with discrete classification to reveal conformational heterogeneity of the sample. However, the number of complexes extracted from cryo-ET data is usually small, which restricts the discrete-classification results to a small number of enough populated states and, thus, results in a largely incomplete conformational landscape. Alternative approaches are currently being investigated to explore the continuity of the conformational landscapes that in situ cryo-ET studies could provide. In this article, we present MDTOMO, a method for analyzing continuous conformational variability in cryo-ET subtomograms based on Molecular Dynamics (MD) simulations. MDTOMO allows obtaining an atomic-scale model of conformational variability and the corresponding free-energy landscape, from a given set of cryo-ET subtomograms. The article presents the performance of MDTOMO on a synthetic ABC exporter dataset and an in situ SARS-CoV-2 spike dataset. MDTOMO allows analyzing dynamic properties of molecular complexes to understand their biological functions, which could also be useful for structure-based drug discovery.

Cryo electron tomography (cryo-ET) is a structural biology technique that allows determining the three-dimensional (3D) structure of abundant biological macromolecular complexes. This technique is particularly important as it allows studying the complexes and their interactions in situ, i.e., in their native cellular environment¹.

In cryo-ET, a series of images are recorded by tilting the sample over a range of angles. During the image collection over multiple tilts, the sample accumulates radiation damage. The sample damage is reduced by (1) lowering the electron dose per image (the electron dose allowed to be accumulated by the sample is spread over the whole tilt range) and (2) collecting images over a smaller tilt range (usually between -60° and $+60^\circ$ in steps of 1° or 2°). The low electron dose to which the sample is exposed during the image collection results in images with a low signal-to-noise ratio (SNR), which in turn results in a noisy 3D volume (the so-called tomogram) that is obtained by 3D reconstruction from the tilt-image series. The limited tilt range used for the data collection results in an anisotropic resolution of the tomogram (the so-called missing-wedge problem)². A macromolecular complex extracted from a tomogram into an individual volume (the so-called subtomogram) is difficult to analyze due to the low SNR and the missing-wedge-induced deformations. Therefore, many instances of the same complex extracted from a tomogram into subtomograms are usually iteratively aligned and averaged (the procedure referred to as subtomogram averaging, STA), which increases the SNR and reduces the missing-wedge-induced deformations.

STA can be combined with discrete 3D classification to tackle the conformational heterogeneity of the complexes³⁻⁶. The goal of the classification is to make the groups (classes) of subtomograms that are as

¹IMPIC-UMR 7590 CNRS, Sorbonne Université, Muséum National d'Histoire Naturelle, CC 115, 4 Place Jussieu, 75005 Paris, France. ²Department of Biochemistry and Pharmacology and Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Melbourne, VIC 3010, Australia. ³Australian Research Council Centre for Cryo-Electron Microscopy of Membrane Proteins, Parkville, VIC 3052, Australia. ✉email: slavica.jonic@upmc.fr

homogeneous as possible regarding the conformation of the complex. To obtain a subtomogram average of good quality from a class, the class should be highly populated. However, considering a relatively small number of available subtomograms, discrete classification methods are usually used with a small number of classes, which makes them ill-suited in the context of highly heterogeneous datasets composed of a continuum of conformational states. This heterogeneity issue, known as continuous conformational variability, is well-known in single particle (SPA) cryo electron microscopy (cryo-EM) and is currently an active field of research^{7–17}. The number of methods for analyzing continuous conformational variability in SPA datasets has been growing over the last few years, but only a few such methods have been proposed for cryo-ET so far^{18,19}. This is likely due to the challenges of analyzing cryo-ET data that were mentioned earlier, namely low SNR, missing-wedge-induced deformations, and small number of subtomograms. Also, volumetric data processing induces additional computational complexity.

Recently, a method named TomoFlow has been proposed for analyzing continuous conformational variability in cryo-ET subtomograms, which uses a computer vision technique of dense 3D optical flow to estimate the displacement of voxels in each subtomogram with respect to the voxels in a global subtomogram average¹⁸. This subtomogram average (reference volume) is calculated at each iteration from all subtomograms that were elastically and rigid-body aligned with respect to the subtomogram average obtained in the previous iteration¹⁸. The reference volume used for the TomoFlow initialization can either be a subtomogram average obtained by classical STA from the given set of subtomograms or it can be a cryo-EM map. The estimated displacements are projected onto a linear subspace through principal component analysis (PCA) to place the subtomograms on a low-dimensional conformational landscape.

HEMNMA-3D is another method for analyzing continuous conformational variability in cryo-ET subtomograms, which is able to use either a reference-volume voxel displacement (as TomoFlow) or a reference-structure atomic displacement to match the conformational states in subtomograms¹⁹. The development of HEMNMA-3D was inspired by the SPA methods that describe the continuous conformational variability by displacing atoms of an available molecular structure (reference structure) to match the conformational states in the data^{7,10,13,16}. Such methods restrain the conformational exploration to the energy region that preserves the structural elements of the complex, which reduces the risk of overfitting low-SNR data. They displace atoms of a reference structure using different methods to model the deformation field, such as normal mode analysis^{7,13}, MD simulations^{10,15}, and Zernike polynomials¹⁶. HEMNMA⁷ and its deep-learning-based extension DeepHEMNMA¹³ perform a flexible alignment of the reference structure to single particle images along the direction in the energy landscape that is defined by a set of vectors of collective motions, called normal modes. The mentioned HEMNMA-3D method¹⁹ is an extension of HEMNMA⁷ to normal-mode-based analysis of subtomograms. Another SPA method follows a direction in the energy landscape that is defined by Zernike polynomials¹⁶. However, the methods based on normal modes and Zernike polynomials restrict the estimation of the dynamics to a predefined direction in the energy landscape (the direction defined by normal modes or Zernike polynomials), which may not be sufficient to fully describe the conformational heterogeneity in the data. Contrary to these methods, a recently proposed SPA method named MDSPACE uses molecular dynamics (MD) simulations, in combination with normal modes, which allows a better exploration of the energy landscape¹⁰. For atomic displacements using a combination of normal modes and MD simulations, MDSPACE employs the algorithm named Normal Mode Molecular Dynamics (NMMD)²⁰, which was initially proposed for flexible fitting of atomic structures into cryo-EM maps in the context of structural modeling.

In this article, we present MDTOMO, a novel approach for analyzing continuous conformational variability in subtomograms, which is based on extracting atomic models from subtomograms by flexible fitting of a reference atomic structure using a combination of normal modes and MD simulation. MDTOMO analyzes each individual subtomogram by performing NMMD-based flexible fitting to extract the atomic model that matches the conformational state in the subtomogram. MDTOMO allows interpreting datasets in the form of atomic conformational landscapes and their further analysis in terms of free-energy landscapes. Moreover, MDTOMO allows obtaining subtomogram averages from localized regions in the energy landscape. While other MD-based flexible fitting methods are used for fitting cryo-EM maps or SPA images, this is the first method that uses MD-based flexible fitting in the context of analyzing subtomograms.

We tested MDTOMO on a synthetic dataset of an ABC exporter²¹ and in situ dataset of a SARS-CoV-2 spike protein (S)²². The synthetic dataset provides a controlled environment where the results of the method can be compared with the ground-truth data, which allows assessing the method performance and robustness to noise and missing-wedge-induced deformations. Although MDTOMO does not currently correct for the missing wedge, it was able to produce meaningful results even for high amounts of noise. MDTOMO applied to in situ data indicates its ability to decipher the conformational landscape as a continuum of intermediate states.

Methods

MDTOMO is a method for extracting information about continuous conformational variability of macromolecular complexes from subtomograms. More precisely, by flexibly fitting an initial atomic model (reference structure) to each subtomogram, MDTOMO can estimate an atomic-scale model of the conformational variability landscape from which it is possible to obtain the energy landscape. Moreover, MDTOMO can visualize principal conformational changes in terms of the atomic displacements or the subtomogram averages of localized regions in the energy landscape.

First, the subtomograms are pre-aligned using a rigid-body alignment with respect to the initial model. Then, the conformational state in each subtomogram is estimated by performing “biased” NMMD simulations. NMMD flexibly displaces the initial atomic coordinates to match the underlying conformational state in the subtomogram. It extracts the conformational state and refines the rigid-body pre-alignment by estimating the rigid-body rearrangement that occurred during the simulation. Finally, the obtained set of atomic models is

projected onto a low-dimensional space (representing an essential conformational space), which allows visualizing conformational distribution and deciphering potential trajectories of conformational changes.

To speed up the NMMD fitting of multiple subtomograms and reduce the risk of overfitting noise in the data, MDTOMO can be used with a coarse-grained initial atomic model, as shown in this article.

Normal mode molecular dynamics. To reduce the computational cost, MDTOMO uses the recently proposed Normal Mode Molecular Dynamics (NMMD)²⁰ to handle the MD simulation. NMMD is a method that incorporates Normal Mode Analysis (NMA) into the simulation to increase the speed of the global-dynamics atomic displacements allowing a better efficiency of the fitting. In NMMD, a set of low-frequency normal modes, representing the most collective motions (the motions that move the largest number of atoms synergistically), is incorporated in the MD integration scheme to encourage the MD simulation to move along the normal mode directions. The normal modes allow a fast displacement along the global dynamics and the MD simulation ensures the structural stability of the system and corrects for the distortions that can be induced by normal modes. The incorporation of normal modes typically reduces the length of the simulation required to fit large-amplitude conformational changes and therefore allows reducing the overall simulation length and the computational cost of the simulation. NMMD has been used to fit atomic models into cryo-EM maps and proved to be, on average, 40% faster than the flexible fitting solely based on MD simulation²⁰. In MDTOMO, NMMD is used to fit an initial atomic model into each subtomogram from the given set of subtomograms.

As in standard MD-based fitting approaches (solely based on MD simulation), NMMD incorporates an additive biasing potential to the total energy that is minimized during the fitting. This additional potential biases the simulation by attracting the atoms into the regions of highest densities in the given EM map while the MD simulation maintains the integrity of the structure. As in the majority of the fitting approaches^{20,23–25}, the biasing potential in MDTOMO is based on the correlation coefficient (CC) between the experimental map and a synthetic map calculated from the atomic model that is being fitted.

Coarse-grained simulations. To further reduce the computational cost of the analysis, we used a coarse-grained model based on off-lattice Ca Gō model²⁶, which drastically reduces the computational time compared to all-atom simulations. The Gō-like models are a class of models that share similar concepts as the model proposed by Gō (e.g., defining the long-range interactions based on the native contacts of the initial model, which tends to bias the simulation towards the reference). Ca Gō model represents the protein with Ca atoms of the amino acid chain and captures better the native dynamics than the original Gō model²⁷. Here, the coarse-grained models were obtained by the SMOG2 software²⁸.

Low-dimensional conformational space and free-energy calculation. As MDTOMO results in one atomic model per subtomogram, a further analysis must be performed to visualize the conformational landscape. In this context, the set of atomic models obtained by MDTOMO can be projected onto a low-dimensional (e.g., 2D or 3D) space. In this study, two methods were used to perform the dimension reduction of the atomic coordinates, namely Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP)²⁹. PCA is the well-established technique for dimension reduction, offering a fast, deterministic and linear reduction method, which is useful to describe the overall variability in most datasets. UMAP is a recent method that allows capturing the non-linear motions, which can, in some cases, achieve a better separation of the different populations compared to PCA. The density of points obtained in the low-dimensional space can then be converted into free energy differences through the Boltzmann factor $\Delta G/k_B T = -\ln(n/n_0)$ by counting the number of particles n in each region of the space and the number of particles in the most populated region n_0 , k_B being the Boltzmann constant and T the temperature of the system. Here, T was set to 300 K. Both PCA and UMAP have been used as dimensionality reduction methods to construct free-energy landscape from MD simulations^{30,31}.

Rigid-body pre-alignment of subtomograms. MDTOMO refines the rigid-body pre-alignment of the subtomograms with the initial model. The pre-alignment prevents the simulation to get trapped into local minima. For the synthetic ABC exporter experiments, the pre-alignment was obtained by STA based on the fast rotational matching algorithm (FRM)³² implemented in Scipion³³. Concerning the in situ dataset of SARS-CoV-2 spikes, we used the STA alignment that had been obtained in the original work²².

The rigid-body pre-alignment is refined in MDTOMO by estimating the rigid-body rearrangements that occurred during the MD simulation, and using them to correct the initial alignment. The estimation of the rigid-body alignment parameters is obtained by optimizing the rotation and the translations that minimize the root mean square deviation (RMSD) between the initial conformation and the fitted conformation, using an optimization algorithm based on singular value decomposition that is available in BioPython³⁴.

Software implementation. MDTOMO is implemented as part of ContinuousFlex³⁵, a software package for continuous conformational heterogeneity analysis in SPA and cryo-ET data, which is also available as a plugin of Scipion³³. The Scipion back-end allows a user-friendly graphical interface for MDTOMO and integrating the data analysis results from most of the recent cryo-ET methods through the new ScipionTomo suite³⁶. The MD and NMMD simulation back-ends are handled by GENESIS³⁷.

Results

In this section, we show the performance of MDTOMO using a set of synthetic subtomograms of an ABC exporter²¹ and using a set of experimental in situ subtomograms of a SARS-CoV-2 spike protein (S)²².

Experiment with a synthetic dataset of ABC exporter. *Synthetic dataset of ABC exporter.* To test the performance of MDTOMO, we synthesized a dataset as close as possible to experimental cryo-ET conditions. The dataset simulates a highly heterogeneous cryo-ET dataset of a heterodimeric ABC exporter, TmrAB, along the substrate translocation cycle. During the translocation cycle, the ABC exporter changes its conformation from outward-facing to inward-facing, allowing the substrate to enter the intracellular cavity. ATP-binding first induces closing of the intracellular gate (an occluded conformation of the exporter, where both intracellular and extracellular gates are closed) and, then, opening of the extracellular gate (an outward-facing conformation, where the extracellular gate opens while the intracellular gate remains closed), which allows a release of the substrate in the extracellular environment. The structure of TmrAB in inward-facing, outward-facing, and occluded conformations were derived from cryo-EM maps²¹ and are available in the Protein Data Bank (PDB) under the accession codes 6RAF, 6RAH, and 6RAK, respectively. Figure 1a shows the PDB: 6RAK structure of TmrAB in the occluded conformation while Fig. 1b shows a sketch of the translocation cycle, with all three mentioned conformational states of TmrAB during the cycle.

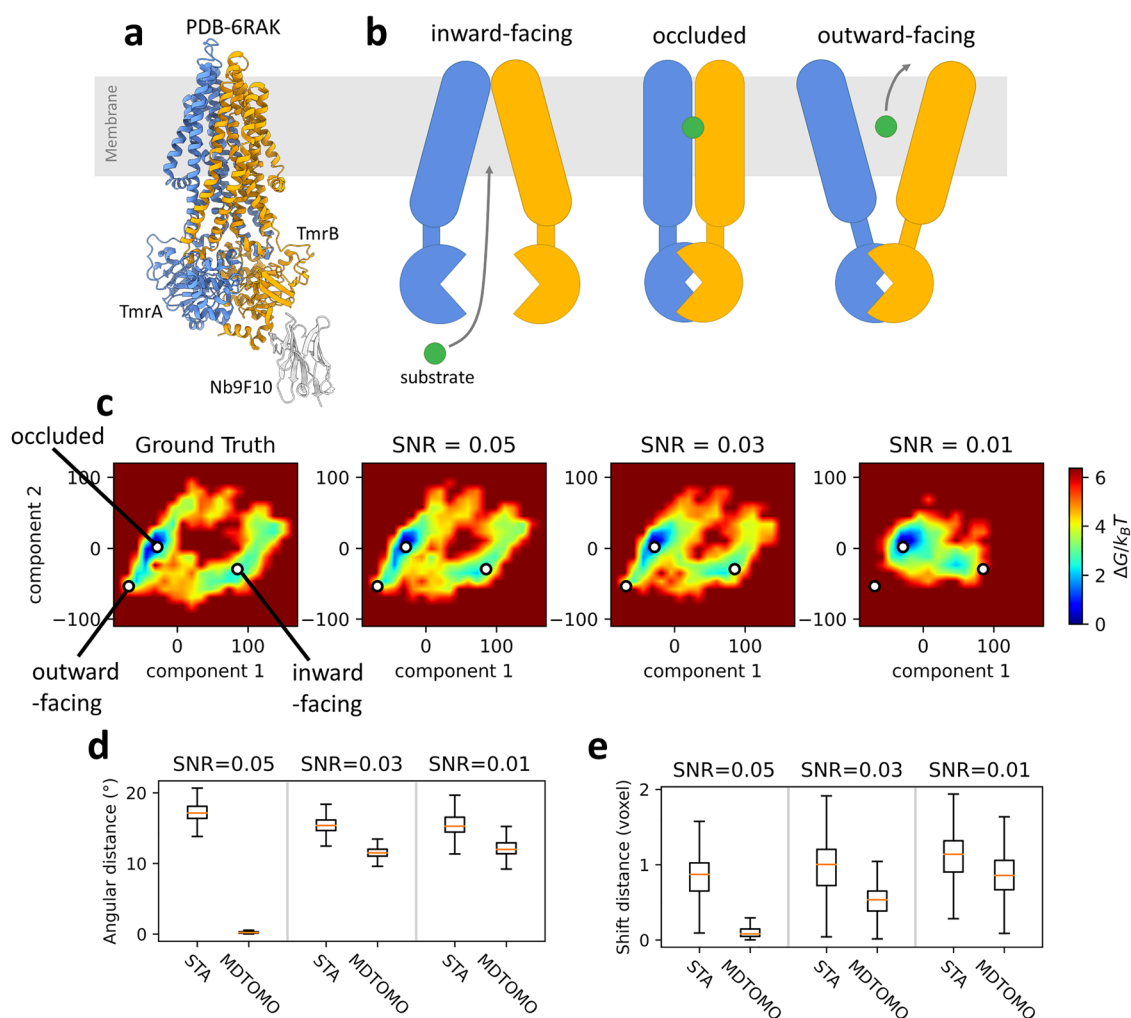


Figure 1. MDTOMO analysis of the synthetic dataset of TmrAB. **(a)** Structure of TmrAB in the occluded conformation. **(b)** Sketch of the substrate translocation process, from left to right: inward-facing, occluded, and outward-facing conformations of TmrAB. **(c)** Free-energy landscapes along the first two PCA components, from left to right: the ground-truth data used to synthesize the subtomograms and the MDTOMO results for the datasets with the SNR of 0.05, 0.03, and 0.01. The location of the three given PDB structures (6RAF, 6RAH, and 6RAK), used to target the MD simulations during the data synthesis, are also shown (white discs) in the free-energy landscapes. **(d,e)** Comparison of the initial and final rotational **(d)** and translational **(e)** alignment errors, for the three datasets (three SNR values).

The conformational heterogeneity was generated by standard MD-based flexible fitting (without normal mode acceleration) to obtain a dataset with realistic conformational variability, composed of a number of more populated conformational states (representing more stable states), close to the three mentioned states of TmrAB during the translocation cycle, and a number of less populated, intermediate states between these more stable states. The MD simulation, starting from the occluded conformation, was directed towards one of the two other conformational states at a time (inward-facing or outward-facing), alternating between these two states, by choosing the biasing potential that forces the simulation to go to one of the two conformational states, with a force constant of 700 kcal/mol. The biasing potential was defined by a simulated EM map of the corresponding conformational state which was changed every 10 ps, resulting in a simulated MD trajectory of the exporter alternating between the outward-facing and inward-facing conformations while passing through a large number of intermediate states, including the occluded conformation. The MD simulation was performed for 1 ns using C α -G \ddot{o} model. One can notice that a 10-ps period may not be long enough for reaching the target conformation for the fitting, meaning that the fitted model at the end of each 10-ps MD-based flexible fitting run may be slightly different from the target conformation. Nevertheless, the dataset consisting of a continuum of conformational states visited during this MD simulation is rich and representative of the translocation cycle enough for performing the tests of the method that is proposed here. It should be noted that we make the problem of recovery of the ground-truth motions more challenging by using NMMD without adding normal modes to MD simulation in the case of data synthesis and by adding normal modes to MD simulation in the case of fitting (next subsection). Additionally, the C α -G \ddot{o} model allows a large enough number of structural degrees of freedom to avoid a bias of synthetic data and the use of two different methods to simulate density maps from the C α -G \ddot{o} model in the data synthesis task and in the fitting task makes the ground-truth motions retrieval even more difficult. For the synthesis of subtomograms, a method based on scattering factors³⁸ is used to convert the structure into a density map, which is then followed by a 2D projection method (to generate the image tilt series) and a 3D reconstruction method (Fourier back-projection). For the fitting in MDTOMO, the simulated density map is obtained directly from the structure by placing 3D Gaussian kernels on C-alpha atomic coordinates.

To simulate the sub-tomograms, we followed a previously published method^{18,19}. We started by collecting 3000 atomic models from the simulated trajectory and we converted each model to a volume of size 100^3 voxels and a voxel size of $(2 \text{ \AA})^3$ using a method based on scattering factors³⁸. The obtained volumes were low-pass filtered at a cutoff frequency of $1/(6 \text{ \AA})$ to take into account the impact of various imperfections (e.g., average radiation damage per image and tilt-series alignment errors). Then, the volumes were rotated and shifted randomly, and projected with a tilt angle from -60° to $+60^\circ$ using an angular step of 2° . The tilt-series images were modulated by a contrast transfer function (CTF) that corresponds to a 200 kV microscope, a defocus of -0.5 \mu m , and a spherical aberration of 2 mm. Following a method that adds Gaussian noise before and after the CTF³⁹, three sets of images were obtained corresponding to a SNR of 0.05, 0.03, and 0.01. The synthetic subtomograms were obtained from the simulated images using weighted back-projection in Fourier space⁴⁰. The subtomograms were aligned using STA based on the FRM algorithm³² in Scipion. It should be noted that more accurate tilt-series simulations would incorporate modeling of variations of different parameters and their effects, such as variations of the particle position, radiation damage, and defocus over the same tilt series and between different tilt series.

Recovery of the ground-truth conformational space. We applied MDTOMO to each of the synthetic datasets (one for each SNR value) composed of 3000 subtomograms using 100-ps NMMD simulations incorporating the 10 lowest frequency normal modes, with a force constant of 5000 kcal/mol starting from the PDB:6RAK model (occluded conformation). Figure 1c shows the free-energy landscape along the first two PCA components for the synthetic MD trajectory (ground truth) and the free-energy landscapes obtained with MDTOMO for each of the three datasets (subtomograms obtained from the images with the SNR of 0.05, 0.03, and 0.01). We observe that MDTOMO was able to recover the conformational landscape very accurately for the datasets with the SNR of 0.05 and 0.03. Indeed, Fig. 1c shows that the most dense regions of the conformational space (i.e., the regions of the lowest free energy) are those that are the closest to the three given PDB structures (6RAF, 6RAH, and 6RAK) that were used to target the MD simulations during the data synthesis. In the case of the dataset with the highest level of noise (SNR of 0.01), the outward-facing conformation was detected less accurately (some densities are missing for the outward-facing state) than the two other states (occluded and inward-facing).

Recovery of the rigid-body alignments. The final rigid-body alignment, obtained by MDTOMO, was compared with the ground truth rigid-body displacement (the rotations and shifts used to synthesize the data). The initial rigid-body alignment, done by STA, was also compared with the ground truth rigid-body displacement. The initial and final alignment errors are compared in Fig. 1d,e, which show respectively the rotational and translational errors for the three datasets (three SNR values). We observe that, for both rotational and translational parameters and each value of the SNR, MDTOMO allows a refinement of the rigid-body alignment (the final errors are smaller than the initial errors).

Experiment with a dataset of SARS-CoV-2 spike protein in situ. *In situ* dataset of SARS-CoV-2 spike protein. The SARS-CoV-2 spike (S) protein is a trimer of three identical glycoproteins (Fig. 2a). It is involved in the cell infection as it mediates the viral entry by binding to the angiotensin-converting enzyme 2 (ACE2) receptor on the cell surface^{41,42}. The S-protein head includes three receptor binding domains (RBDs) that are located at the top of the spike and exists in different conformations on the viral capsid. We used MDTOMO to analyze a set of experimental subtomograms of the S protein in situ that had previously been obtained and analyzed using STA and 3D classification (the corresponding tilt-series data are available in the EMPIAR database; accession code: EMPIAR-10453)²². In that previous publication²², the tilt series were reconstructed into

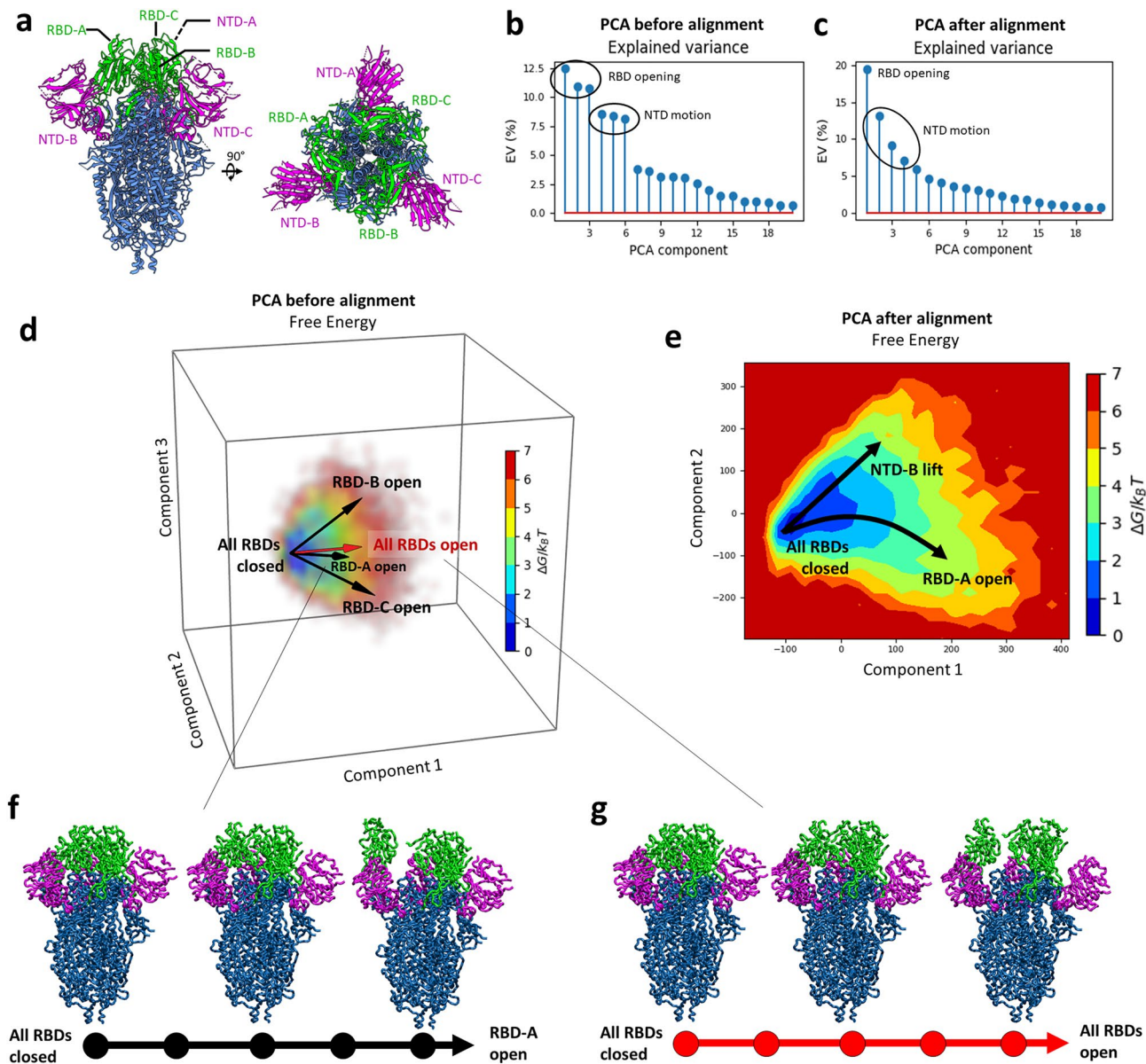


Figure 2. PCA analysis of MDTOMO results with a cryo-ET dataset of SARS-CoV-2 S protein. **(a)** Structure of the S protein (PDB 6VXX). **(b,c)** Explained variance of the PCA space obtained before **(b)** and after **(c)** reducing the number of principal components to describe opening of RBDs. **(d)** Free-energy landscape determined by the first three principal components, obtained before reducing the number of components to describe opening of RBDs. **(e)** Free-energy landscape determined by the first and second principal components obtained after reducing the number of components to describe opening of RBDs. The arrows show the directions associated with the RBD and NTD motions discussed here. **(f)** RBD-A opening trajectory following the direction “RBD-A open” shown in **(d)**. **(g)** Trajectory of opening of all three RBDs together following the direction “All RBDs open” shown in **(d)**.

tomograms with 3D-CTF correction⁴³, 20,830 subtomograms were picked and classified several classes, out of which two classes were further described. They correspond to two distinct conformational states, out of which one has three closed RBDs (fully closed state) and the other has one RBD opened (the classes containing 8273 and 4321 subtomograms, respectively).

MDTOMO was used to analyze the entire set of 20,830 subtomograms, which were first downsampled from their original size (256^3 voxels) to the size of 128^3 voxels and then pre-aligned using the rigid-body alignment parameters obtained by STA in the previous publication²². The initial conformation for the NMMD fitting with MDTOMO was PDB 6VXX, which corresponds to the fully closed state. The simulations included an initial energy minimization, followed by 100-ps NMMD simulations incorporating the 10 lowest frequency normal modes, with a force constant of 7000 kcal/mol.

Continuous conformational landscape of the S protein. The conformational variability was extracted by PCA from the set of atomic models obtained by MDTOMO. This PCA revealed multiple asymmetric conformations of the S protein. The plot in Fig. 2b shows the variance described by each PCA component, and indicates that the first six principal components are responsible for more than 60% of the total variability. By exploring the conformational changes in different directions in the PCA space for the first six components, we could observe that the first three principal components (the most dominant motions) describe asymmetric opening and closing motions of the RBDs, whereas the next three principal components (less dominant motions) describe asymmetric up and down motions of the N-terminal domains (NTDs). Here, “asymmetric motion” means that each domain (RBD or NTD) can move independently from other domains.

Figure 2d shows a space determined by the first three principal components. Each of the three black arrows in this space, denoted by “RBD-A open”, “RBD-B open”, and “RBD-C open” (Fig. 2d), shows a direction of motion that mainly describes a single RBD opening (one RBD is opening while the other two RBDs remain closed), for each of the three RBDs denoted in Fig. 2a by “RBD-A”, “RBD-B” and “RBD-C”, respectively. Figure 2f shows the conformational change along the “RBD-A open” direction (opening of the RBD-A domain). In other directions of this space (Fig. 2d), we could observe other conformational changes. In particular, when going from low-energy regions to high-energy regions in other directions than those indicated by black arrows (Fig. 2d), we could observe opening of more than one RBD. For instance, the red arrow denoted by “All RBDs open” in Fig. 2d indicates a direction along which all three RBDs are opening together, and Fig. 2g shows the conformational changes along this direction (from the conformation with all three RBDs closed to a conformation with all three RBDs open).

Although the original study (which used STA and discrete classification) revealed only a fully closed state and a state with a single RBD open²², simultaneous opening of more than one RBD has been discussed in the literature⁴⁴. However, considering the symmetry (C3) of the conformation with all three RBDs fully closed and our results showing that each of the three RBDs can undergo opening motion independently of the other two RBDs, we explored the possibility that a large number of the conformational models obtained by MDTOMO could be identical up to a rotation by 120° or 240° around the C3 pseudo-symmetry axis, which could have caused a difficulty in well separating the principal components by the PCA. Therefore, we designed a method to reduce the variability in order to simplify the PCA. This method is based on aligning the conformational models obtained by MDTOMO with respect to 5 conformations along the “RBD-A open” direction (Fig. 2d), which describe opening of the RBD-A domain. Three out of the five conformational models used for this alignment are shown in Fig. 2f. This method, which simplifies conformational variability analysis of the S protein, is explained next.

Simplifying PCA of conformational variability based on pseudo-symmetries of conformations. As introduced in the previous subsection, we reduced the number of principal components describing the variability due to RBD motions by aligning the conformational states obtained by MDTOMO with the states of a selected trajectory of a single RBD opening extracted from the initial PCA space (Fig. 2d), which we here refer to as reference trajectory. The reference trajectory was selected as a five-state trajectory that corresponds to the most dominant single-RBD motion in the space determined by the first three PCA components, which is the opening of RBD-A (Fig. 2d,f). The alignment was done by rotating the MDTOMO-derived conformations around the C3 pseudo-symmetry axis (here the z-axis) by multiples of 120° and matching the rotated conformations with the states on the reference trajectory. Note that we chose to split the trajectory in five states as this is the minimal number of states for obtaining a comprehensive description of the different intermediate states in this dataset.

More precisely, the set of 20,830 MDTOMO-derived conformational models was extended three times, by applying $n \times 120^\circ$ rotation around the z-axis with $n = \{0, 1, 2\}$, resulting in 62,490 models (20,830 triplets, each triplet composed of the model rotated by $n \times 120^\circ$). Each of the three models in a triplet was compared by RMSD with the five models from the reference trajectory resulting in 15 RMSD values (five RMSD values per model of a triplet). The lowest value of the 15 RMSD values was identified and the associated rotation factor n was collected to determine the angle for the realignment of the MDTOMO-derived conformational model.

A new PCA was performed on the set of models obtained after the alignment, resulting in a new conformational variability space, described mostly by two principal components corresponding to a motion of RBD-A opening and a motion of NTD-B lifting, as shown in Fig. 2e. Based on Fig. 2e, it can be noted that the number of principal components necessary to describe RBD motions was reduced from 3 to 1. Also, in the new PCA space, the NTD motions are described by lower principal components (Fig. 2c) than in the initial PCA space (components 2–4 in the new space vs. components 4–6 in the initial space).

The set of MDTOMO-derived conformational models aligned with the selected trajectory of single RBD opening (the trajectory shown in Fig. 2f) was further analyzed using UMAP, which allowed a better separation of the different conformational states and identification of more populated distinct conformations. UMAP was obtained with a reduction dimension of 5, an Euclidean metric for computing the distance in the ambient space, a neighborhood size of 15, and a minimum distance between points of 0.1.

UMAP representation of the conformational landscape simplified using pseudo-symmetries of conformations. The conformational landscape obtained after reducing the number of principal components to describe opening of RBDs (Fig. 2e) shows a motion of RBD-A opening and a motion of NTD-B lifting. The set of aligned MDTOMO-derived conformational models was analyzed with UMAP to better separate different conformational states. The first two UMAP components are presented in terms of free-energy in Fig. 3a. The conformational models and the density maps shown in Fig. 3 are the conformational-model and subtomogram averages calculated from the four most populated regions (regions of lowest energy), manually selected and depicted in Fig. 3a.

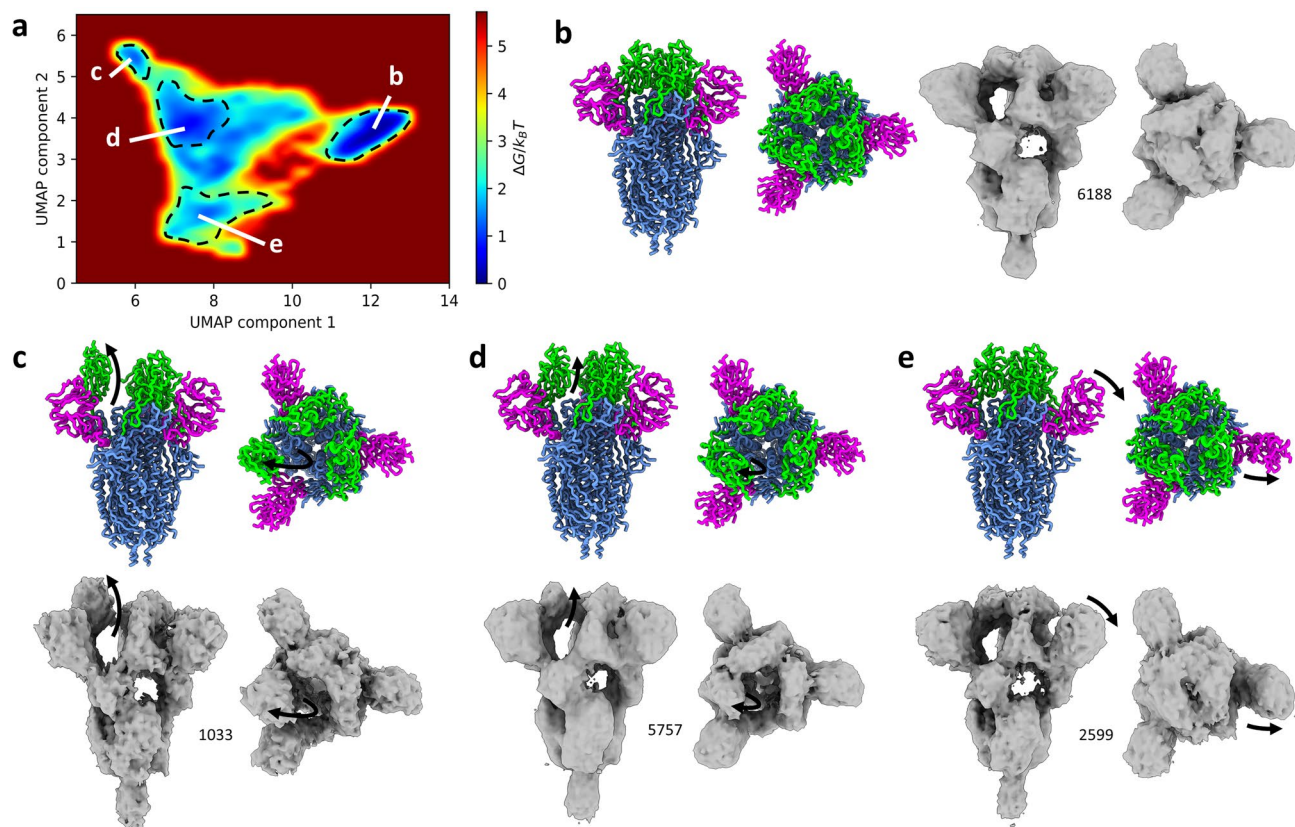


Figure 3. UMAP analysis of MDTOMO results with a cryo-ET dataset of SARS-CoV-2 S protein, after simplifying description of the variability due to opening of RBDs. **(a)** Free-energy landscape along the UMAP components 1 and 2. **(b–e)** Conformational-model and sub-tomogram averages from four manually selected regions in the free-energy landscape (denoted by **b–e**). The colored model corresponds to the average of the conformational models in the selected region. The density maps are the averages of the subtomograms from the selected regions in the energy landscape. The number of the averaged particles is indicated near each map. Black arrows represent the conformational change from the initial structure used for the fitting within MDTOMO (PDB-6VXX, which is a conformation with all RBD closed).

The most populated region, designated as **(b)** in Fig. 3a, corresponds to a conformer with all 3 RBDs closed. The corresponding conformational-model average and subtomogram average are shown in Fig. 3b. The most distant low-density region from the region **(b)** along the first UMAP component, designated as **(c)** in Fig. 3a, corresponds to a conformer with a single open RBD, i.e., RBD-A (Fig. 3c). The central region, designated as **(d)** in Fig. 3a, corresponds to intermediate states of a single RBD opening (Fig. 3d). The bottom region along the second UMAP component, designated as **(e)** in Fig. 3a, corresponds to a conformer with the NTB-B lifted (Fig. 3e).

Discussion

In this article, we presented MDTOMO, a novel approach to explore continuous conformational landscapes of biomolecular complexes by analyzing subtomogram datasets. MDTOMO uses efficient, coarse-grained and normal-mode-empowered MD simulations to extract the conformational states from subtomograms, which are typically averaged out during classical STA, and allows reconstructing low-dimensional (usually 2D or 3D) energy profiles of the structures from the experimental subtomogram data. MDTOMO is the first method for analyzing subtomograms using MD-based flexible fitting.

MDTOMO starts by extracting an atomic model from each individual subtomogram, by NMMD flexible fitting of a given reference atomic model into the subtomogram. The extracted atomic models are then projected onto a low-dimensional space, which can be considered as an essential conformational space. In this space, closer points represent more similar conformations, and denser regions of points represent the regions containing more likely or more stable (low-energy) conformations in the given dataset. The density of points in this space can be interpreted in terms of the free energy of the complex or by computing and analyzing the averages of the subtomograms and their associated atomic models in the regions of the space characterized by the lowest free energy of the complex. In this article, we showed that MDTOMO produces expected results using synthetic data (ABC exporter). Also, we showed that the results produced using experimental SARS-CoV-2 spike protein data are coherent with previously published results regarding the existence of conformations with all three RBDs closed and those with a single RBD open²². Additionally, they revealed the existence of gradual conformational

transitions related to the opening of single RBDs, a lifting motion of a single NTB, and additional conformations with more than one fully or partially open RBD.

In this article, we analyzed two datasets of two different proteins (ABC exporter and SARS-CoV-2 spike protein) of two different sizes (3000 particles and 20,000 particles, respectively). Although this study shows only two examples of application of MDTOMO, the method can be applied on macromolecular complexes of various sizes. Its computing time scales with the number of atoms included in the model and the size of the subtomograms. For instance, a method using a similar approach as MDTOMO (NMMD-based fitting) for applications in single particle image analysis, MDSPACE, successfully analyzed the variability of the large complex of 80S ribosome¹⁰. MDTOMO can also be applied on cryo-ET datasets with smaller voxel sizes and larger volumes (the subtomograms of SARS-CoV-2 spikes had a size of 128^3 voxels and a voxel size of 2.65 \AA^3).

In the experiment with SARS-CoV2 spike data, we showed the results obtained using PCA and UMAP dimensionality reduction methods. PCA is a linear dimensionality reduction method that decomposes the data variability into a linear combination of principal component vectors. UMAP is a non-linear dimensionality reduction method that in some cases gives a better separation of different conformational populations than PCA, as was the case with the analysis of SARS-CoV2 spike data in this article. In particular, UMAP allowed revealing two clusters of the conformational states previously observed using the same dataset and the classical STA data analysis procedure²² (all three closed RBDs and one-open RBD, Fig. 3b,c) and two clusters of the states unobserved previously (Fig. 3d,e). In the case of synthetic dataset of ABC exporter, we only showed the results obtained with PCA because PCA and UMAP produced similar conformational landscapes in the case of this dataset.

However, as UMAP extracts non-linear features from the data, it produces a conformational landscape with non-linear parametrization in which the interpretation of energy densities must be done carefully. In this conformational embedding, equivalent distances are not necessarily equivalent conformational difference. For instance, in the UMAP embedding in Fig. 3, the distance between the centers of the c and d regions and between the centers of the d and b regions is 2 and 4.5, respectively, but these two distances correspond to approximately the same conformational difference (a rotation of the RBD of approximately 30°). This also implies that local minima observed in the energy landscape are not necessarily meaningful stable conformational states, but rather close intermediate states that were regrouped during the embedding (e.g., a number of local minima are visible in the central region of Fig. 3a). To prevent any misinterpretation of the landscape, the future MDTOMO improvements will include the improvement of the conformational landscape interpretation using methods that could preserve local point densities⁴⁵.

In the method presented in this article, the correlation coefficient was used as the metric for fitting the conformation in subtomograms. In standard methods, the alignment and classification of subtomograms use the correlation coefficient in combination with a compensation for the missing wedge (e.g., constrained correlation coefficient⁴⁶, wedge-masked differences⁴⁷) because of the sensitivity of the correlation coefficient to anisotropic deformations of subtomograms caused by the missing wedge. MDTOMO uses correlation-driven MD-based fitting without compensating for the missing wedge. The use of missing-wedge compensated metric such as constrained correlation coefficient (CCC) to drive the MD fitting was considered. However, a prototype implementation of the gradient of the CCC required for the energy calculation showed to be too computationally expensive for running MD fitting and requires further developments in order to be efficient. Nevertheless, MDTOMO was able to produce expected results with synthetic and experimental data without compensating for the missing wedge, which can be explained by the fact that the NMMD flexible alignment in MDTOMO is not only driven by the force that depends on the correlation coefficient (derivatives of the correlation coefficient with respect to the atomic positions), but also by the classical MD-based force field (Eq. 1 and Eq. 4 in Vuillemot et al.¹⁰). The weight given to the potential defined by the correlation coefficient, with respect to the classical MD potential, is a free parameter (the so-called force constant; k in Eq. 1 in Vuillemot et al.¹⁰) whose value is set so as to drive the fitting toward the conformation in the subtomogram by preserving structural constraints, which reduces the risk of overfitting the noise and the missing-wedge-induced deformations in the subtomograms.

Compared to the majority of other methods for analyzing conformational flexibility in single particle cryo-EM datasets (e.g., 3DVA¹¹ in cryoSPARC⁴⁸ or cryoDRGN¹²) or cryo-ET datasets (e.g., Tomoflow¹⁸), MDTOMO uses an available atomic structure, whose coordinates are displaced (by performing NMMD simulation) to match the conformations in the subtomograms. Such atomic structure is not available for all datasets, but those datasets for which an atomic structure is available can efficiently be analyzed using MDTOMO. Indeed, MDTOMO incorporates structural constraints into the analysis via MD simulation, which is particularly important for analyzing cryo-ET subtomograms that suffer from noise and missing-wedge-induced deformations, as these structural constraints limit the motions to those that are feasible (underlying biochemical processes of the complex) and, therefore, they reduce the risk of overfitting mentioned previously. Also, the use of an available atomic structure in MDTOMO allows obtaining a conformational landscape in which the conformational transitions can be visualized directly on the atomic structure, contrary to the majority of other methods that visualize the conformational transitions on density volumes.

Another method that can use an available atomic structure to analyze conformational flexibility in cryo-ET datasets is HEMNMA-3D¹⁹. More precisely, HEMNMA-3D can use either an available atomic structure or an available density volume (e.g., subtomogram average or cryo-EM map)¹⁹, which is flexibly fitted into subtomograms by solely using normal mode analysis. Normal modes are useful for their simplicity and computational efficiency. However, they tend to induce distortions of the structure in case of large amplitudes of conformational changes. Also, they restrict the conformational exploration to a set of collective motions that must be defined in advance and that may be insufficient to describe the full range of conformational changes. MDTOMO overcomes these limitations by employing coarse-grained MD simulations to allow a full exploration of the conformational landscape without inducing structural distortions, and by empowering these MD simulations using normal modes to make simulations fast.

Complex	Weight (kDa)	100 ps-simulation MDTOMO on Intel 6248 2.50 GHz	
		One CPU core—one particle (s)	320 CPU cores—full dataset
ABC exporter	150.85	372.3	0.9 h (3000 particles)
SARS-CoV2 spike	438.26	987.5	17.8 h (20,830 particles)

Table 1. Execution time of MDTOMO for the two datasets analyzed in this article.

The MDTOMO approach requires running one MD simulation per subtomogram, which can be computationally expensive. The MDTOMO software is parallelized so that multiple simulations can run on multiple CPU cores simultaneously, with one simulation running only on one CPU core. Furthermore, the use of coarse-grained simulation coupled with normal-mode-based acceleration allows performing MDTOMO in a reasonable amount of time. The total execution time on 8 nodes of 2×20 Intel 6248 2.50 GHz for the two presented datasets is shown in Table 1. However, it is important to note that the computational cost scales with the molecular weight of the complex (the higher the number of atoms, the higher the cost of the MD simulation).

Data availability

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

Received: 3 March 2023; Accepted: 14 June 2023

Published online: 30 June 2023

References

- Mahamid, J. *et al.* Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016).
- Turoňová, B., Marsalek, L. & Slusallek, P. On geometric artifacts in cryo electron tomography. *Ultramicroscopy* **163**, 48–61 (2016).
- Scheres, S. H. W., Melero, R., Valle, M. & Carazo, J.-M. Averaging of electron subtomograms and random conical tilt reconstructions through likelihood optimization. *Structure* **17**, 1563–1572. <https://doi.org/10.1016/j.str.2009.10.009> (2009).
- Bartesaghi, A. *et al.* Classification and 3D averaging with missing wedge correction in biological electron tomography. *J. Struct. Biol.* **162**, 436–450. <https://doi.org/10.1016/j.jsb.2008.02.008> (2008).
- Zhang, P. Advances in cryo-electron tomography and subtomogram averaging and classification. *Curr. Opin. Struct. Biol.* **58**, 249–258. <https://doi.org/10.1016/j.sbi.2019.05.021> (2019).
- Castaño-Díez, D. & Zanetti, G. In situ structure determination by subtomogram averaging. *Curr. Opin. Struct. Biol.* **58**, 68–75. <https://doi.org/10.1016/j.sbi.2019.05.011> (2019).
- Jin, Q. *et al.* Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure* **22**, 496–506 (2014).
- Sorzano, C. O. S. *et al.* Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. *Acta Crystallogr. Struct. Biol. Commun.* **75**, 19–32. <https://doi.org/10.1107/S2053230X18015108> (2019).
- Jonić, S. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. *Curr. Opin. Struct. Biol.* **43**, 114–121. <https://doi.org/10.1016/j.sbi.2016.12.011> (2017).
- Vuillemot, R. *et al.* MDSPACE: Extracting continuous conformational landscapes from Cryo-EM single particle datasets using 3D-to-2D flexible fitting based on molecular dynamics simulation. *J. Mol. Biol.* **435**, 167951. <https://doi.org/10.1016/j.jmb.2023.167951> (2023).
- Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J. Struct. Biol.* **213**, 107702 (2021).
- Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks. *Nat. Methods* **18**, 176–185 (2021).
- Hamitouche, I. & Jonic, S. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. *Front. Mol. Biosci.* **9**, 965645. <https://doi.org/10.3389/fmolb.2022.965645> (2022).
- Andén, J. & Singer, A. Structural variability from noisy tomographic projections. *SIAM J. Imaging Sci.* **11**, 1441–1492 (2018).
- Giraldo-Barreto, J. *et al.* A Bayesian approach to extracting free-energy profiles from cryo-electron microscopy experiments. *Sci. Rep.* **11**, 13657. <https://doi.org/10.1038/s41598-021-92621-1> (2021).
- Herreros, D. *et al.* Estimating conformational landscapes from Cryo-EM particles by 3D Zernike polynomials. *Nat. Commun.* **14**, 154. <https://doi.org/10.1038/s41467-023-35791-y> (2023).
- Tagare, H. D., Kucukelbir, A., Sigworth, F. J., Wang, H. & Rao, M. Directly reconstructing principal components of heterogeneous particles from cryo-EM images. *J. Struct. Biol.* **191**, 245–262 (2015).
- Harastani, M., Eltsov, M., Leforestier, A. & Jonic, S. TomoFlow: Analysis of continuous conformational variability of macromolecules in cryogenic subtomograms based on 3D dense optical flow. *J. Mol. Biol.* **434**, 167381 (2022).
- Harastani, M., Eltsov, M., Leforestier, A. & Jonic, S. HEMNMA-3D: Cryo electron tomography method based on normal mode analysis to study continuous conformational variability of macromolecular complexes. *Front. Mol. Biosci.* **8**, 663121. <https://doi.org/10.3389/fmolb.2021.663121> (2021).
- Vuillemot, R., Miyashita, O., Tama, F., Rouiller, I. & Jonic, S. NMMD: Efficient cryo-EM flexible fitting based on simultaneous normal mode and molecular dynamics atomic displacements. *J. Mol. Biol.* **434**, 167483 (2022).
- Hofmann, S. *et al.* Conformation space of a heterodimeric ABC exporter under turnover conditions. *Nature* **571**, 580–583. <https://doi.org/10.1038/s41586-019-1391-0> (2019).
- Turoňová, B. *et al.* In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203–208. <https://doi.org/10.1126/science.abd5223> (2020).
- Orzechowski, M. & Tama, F. Flexible fitting of high-resolution X-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys. J.* **95**, 5692–5705 (2008).
- Miyashita, O., Kobayashi, C., Mori, T., Sugita, Y. & Tama, F. Flexible fitting to cryo-EM density map using ensemble molecular dynamics simulations. *J. Comput. Chem.* **38**, 1447–1461 (2017).

25. Igaev, M., Kutzner, C., Bock, L. V., Vaiana, A. C. & Grubmüller, H. Automated cryo-EM structure refinement using correlation-driven molecular dynamics. *Elife* **8**, e43542 (2019).
26. Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: What determines the structural details of the transition state ensemble and “en-route” intermediates for protein folding? An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953 (2000).
27. Takada, S. Gö model revisited. *Biophys. Physicobiol.* **16**, 248–255 (2019).
28. Noel, J. K. *et al.* SMOG 2: A versatile software package for generating structure-based models. *PLoS Comput. Biol.* **12**, e1004794. <https://doi.org/10.1371/journal.pcbi.1004794> (2016).
29. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. [arXiv:abs/1802.03426](https://arxiv.org/abs/1802.03426) (2018).
30. Trozzi, F., Wang, X. & Tao, P. UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: A comparison study. *J. Phys. Chem. B* **125**, 5022–5034 (2021).
31. Papaleo, E., Mereghetti, P., Fantucci, P., Grandori, R. & De Gioia, L. Free-energy landscape, principal component analysis, and structural clustering to identify representative conformations from molecular dynamics simulations: The myoglobin case. *J. Mol. Graph. Model.* **27**, 889–899 (2009).
32. Chen, Y., Pfeffer, S., Hrabe, T., Schuller, J. M. & Förster, F. Fast and accurate reference-free alignment of subtomograms. *J. Struct. Biol.* **182**, 235–245. <https://doi.org/10.1016/j.jsb.2013.03.002> (2013).
33. de la Rosa-Trevín, J. M. *et al.* Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J. Struct. Biol.* **195**, 93–99. <https://doi.org/10.1016/j.jsb.2016.04.010> (2016).
34. Cock, P. J. A. *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423. <https://doi.org/10.1093/bioinformatics/btp163> (2009).
35. Harastani, M., Vuillemot, R., Hamitouche, I., Moghadam, N. B. & Jonic, S. ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. *J. Struct. Biol.* **214**, 107906. <https://doi.org/10.1016/j.jsb.2022.107906> (2022).
36. Jiménez de la Morena, J. *et al.* ScipionTomo: Towards cryo-electron tomography software integration, reproducibility, and validation. *J. Struct. Biol.* **214**, 107872. <https://doi.org/10.1016/j.jsb.2022.107872> (2022).
37. Kobayashi, C. *et al.* Genesis 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. *J. Comput. Chem.* **38**, 2193–2206 (2017).
38. Peng, L.-M., Ren, G., Dudarev, S. & Whelan, M. Robust parameterization of elastic and absorptive electron atomic scattering factors. *Acta Crystallogr. A* **52**, 257–276 (1996).
39. Sorzano, C. O., Jonic, S., Nunez-Ramirez, R., Boisset, N. & Carazo, J. M. Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. *J. Struct. Biol.* **160**, 249–262. <https://doi.org/10.1016/j.jsb.2007.08.013> (2007).
40. Strelak, D. *et al.* Advances in Xmipp for cryo-electron microscopy: From Xmipp to Scipion. *Molecules* **26**, 6224 (2021).
41. Hamming, I. *et al.* Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J. Pathol.* **203**, 631–637 (2004).
42. Hoffmann, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280. [e278](https://doi.org/10.1016/j.cell.2020.05.027) (2020).
43. Turoňová, B., Schur, F. K., Wan, W. & Briggs, J. A. Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195 (2017).
44. Benton, D. J. *et al.* Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, 327–330 (2020).
45. Narayan, A., Berger, B. & Cho, H. Density-preserving data visualization unveils dynamic patterns of single-cell transcriptomic variability. *bioRxiv* **10**, 12.077776 (2020).
46. Förster, F., Pruggnaller, S., Seybert, A. & Frangakis, A. S. Classification of cryo-electron sub-tomograms using constrained correlation. *J. Struct. Biol.* **161**, 276–286 (2008).
47. Heumann, J. M., Hoenger, A. & Mastrorade, D. N. Clustering and variance maps for cryo-electron tomography using wedge-masked differences. *J. Struct. Biol.* **175**, 288–299 (2011).
48. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).

Acknowledgements

We acknowledge the support of the French National Research Agency—ANR (ANR-19-CE11-0008 to SJ and IR and ANR-20-CE11-0020-03 to SJ); the cooperation between the CNRS and the University of Melbourne (The Melbourne-CNRS Network, PRC 2889 to SJ and IR, CNRS 80 Prime to SJ, joint PhD scholarship to RV); The University of Melbourne start-up fund (to IR); and access to HPC resources of CINES and IDRIS granted by GENCI (A0100710998R, A0100710998, A0070710998, AP010712190, AD011012188 to SJ). We thank Beata Turoňová (Max Planck Institute of Biophysics, Frankfurt am Main, Germany) for providing the subtomograms and the initial rigid-body alignment parameters that were used in the SARS-CoV-2 S protein experiment described in this article (from the dataset available in the EMPIAR database under the code EMPIAR-10453) as well as for critical reading of our manuscript and discussing the results that are presented here.

Author contributions

R.V.: Conceptualization, Methodology, Software, Investigation, Validation, Writing—Original draft preparation. I.R.: Validation, Writing—Review and Editing, Funding acquisition. S.J.: Conceptualization, Methodology, Investigation, Validation, Writing—Review and Editing, Project administration, Funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023