MDSPACE: Extracting continuous conformational landscapes from cryo-EM single particle datasets using 3D-to-2D flexible fitting based on Molecular Dynamics simulation

Rémi Vuillemot, Alex Mirzaei, Mohamad Harastani, Ilyes Hamitouche, Léo Fréchin, Bruno P Klaholz, Osamu Miyashita, Florence Tama, Isabelle Rouiller, Slavica Jonic

This is the author's version of an article accepted for publication in Journal of Molecular Biology (article reference: YJMBI_167951)

MDSPACE: Extracting continuous conformational landscapes from cryo-EM single particle datasets using 3D-to-2D flexible fitting based on Molecular Dynamics simulation

Rémi Vuillemot[1,6], Alex Mirzaei[1], Mohamad Harastani[1], Ilyes Hamitouche[1], Léo Fréchin[2], Bruno P. Klaholz[2], Osamu Miyashita[3], Florence Tama[3,4,5], Isabelle Rouiller[6], Slavica Jonic[1,*]

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

[2]Centre for Integrative Biology, Department of Integrated Structural Biology, IGBMC-UMR 7104 CNRS, U964 Inserm, Université de Strasbourg, Strasbourg, France

[3]RIKEN Center for Computational Science, Kobe, Japan

[4]Institute of Transformative Biomolecules, Graduate School of Science, Nagoya University, Nagoya, Japan

[5]Department of Physics, Graduate School of Science, Nagoya University, Nagoya, Japan

[6]Department of Biochemistry & Pharmacology and Bio21 Molecular Science and Biotechnology Institute, University of Melbourne, Victoria, Australia

Contact details of the corresponding author:

Dr. Slavica Jonić

Sorbonne Université

IMPMC - CNRS UMR 7590, CC 115

4 Place Jussieu, 75005 Paris, France

Phone: +33 1 44 27 72 05

Fax: +33 1 44 27 37 85

E-mail: slavica.jonic@upmc.fr

Running title: MDSPACE

Abstract: This article presents an original approach for extracting atomic-resolution landscapes of continuous conformational variability of biomolecular complexes from cryo electron microscopy (cryo-EM) single particle images. This approach is based on a new 3D-to-2D flexible fitting method, which uses molecular dynamics (MD) simulation and is embedded in an iterative conformational-landscape refinement scheme. This new approach is referred to as MDSPACE, which stands for Molecular Dynamics simulation for Single Particle Analysis of Continuous Conformational hEterogeneity. The article describes the MDSPACE approach and shows its performance using synthetic and experimental datasets.

Keywords: Cryo electron microscopy (cryo-EM); molecular dynamics simulation (MD simulation); continuous conformational variability; 3D-to-2D flexible fitting; conformational landscapes; principal component analysis; single particle analysis

1. Introduction

Cryo electron microscopy (cryo-EM) single particle analysis allows obtaining information about conformational variability of purified biomolecular complexes embedded in random orientations and positions in a thin layer of vitreous ice. Sophisticated image processing algorithms and powerful computational hardware are required to disentangle multiple co-existing conformations of the complexes (particles) from parallel electron-beam projection images of the sample. The particles first need to be extracted in individual images from the imaged zones of the sample and, then, a complex problem of combined orientational, positional, and conformational heterogeneity needs to be solved. Biological samples are imaged at a low electron dose to limit irradiation damage, which results in a low signal-to-noise ratio (SNR) of the collected images that makes image processing tasks difficult.

Continuous conformational variability is a term introduced about ten years ago and currently largely used in cryo-EM to refer to the concept according to which biomolecular complexes can be considered to follow continuous trajectories of conformational changes and the conformational states contained in the images are samples on these trajectories [1-5]. This concept has been the basis for the development of methods to obtain the conformational landscapes of the complexes *in vitro*, by single particle analysis [1, 2, 6-15]. Similar methods are currently also being developed for analyzing continuous conformational variability of the complexes *in situ*, by cryo electron tomography [16, 17].

Methods for deciphering continuous conformational variability from cryo-EM single particle images do not make any assumption on the number of the different conformational states present in the sample, but rather consider that each particle image may come from a different particle conformation [1, 2, 6-15]. On the contrary, the currently still most used methods, based on maximum likelihood classification [18-20], limit the number of the output conformations as they in advance fix the number of classes (much smaller than the number of particle images) into which the particle images should be separated (separation by assigning each particle image to its most likely class), which is followed by further iterative refinement of the highest-resolution classes (with most similar conformations) and discarding other classes. This allows fast high-resolution 3D reconstructions of a few conformational averages, provided that the optimized biochemical procedures were used to maximize the homogeneity of the sample [21-28]. Such methods are often referred to as discrete classification methods [29-34].

In 2014, we introduced HEMNMA (Hybrid Electron Microscopy Normal Mode Analysis), one of the first methods for analyzing continuous conformational variability from cryo-EM images, which is based on a combined 3D-to-2D flexible and rigid-body fitting of an initial conformation of the particle (an atomic model or a cryo-EM map) against particle images [2]. The 3D-to-2D fitting in HEMNMA determines the conformations based on a linear combination of normal modes of the initial conformation and the amplitudes of the normal modes. Usually, a few low-frequency high-collectivity normal modes are enough to describe

global motion directions of the particle [35-39]. HEMNMA elastically deforms the initial conformation using tentative normal-mode amplitudes and rigid-body aligns the obtained conformations with the images to determine the best matching conformation for each particle image simultaneously with the particle orientation and position in the image. Very recently, we published a deep learning extension of HEMNMA, named DeepHEMNMA [7], which involves a deep neural network that can be trained to emulate the HEMNMA operations, achieving an impressive speed-up in determining the conformations, orientations, and positions of particles for large datasets, with respect to HEMNMA alone (e.g., at least 40 times faster in the experiments shown in [7]). Both HEMNMA and DeepHEMNMA are available in open-source ContinuousFlex plugin [40] for Scipion [41].

The methods with publicly available and user-friendly software that are additionally able to quasi-simultaneously determine or refine the particle poses (orientations and positions) and the particle conformations in the particle images have advantages over other methods. HEMNMA was the first such method [2]. The majority of other continuous conformational variability methods [1, 6, 9-14] perform the conformational analysis using fixed orientations and positions of the particles, obtained with discrete classification methods [18-20]. CryoDRGN2 is a publicly available deep learning software that allows a refinement of the particle poses by alternating between a sequence of iterations of volume learning and a pose refinement iteration (by an exhaustive search of orientations and positions over a discretized 5D parameter space), but this scheme has been reported to suffer from vanishing gradients due to the change of the training objective during the course of the training [42].

HEMNMA/DeepHEMNMA can directly produce an atomic model for any given particle image (when using an atomic model as the initial conformation), whereas other continuous conformational variability methods produce EM maps that are then used to derive atomic models. More precisely, the initial atomic conformation is obtained by X-ray crystallography or derived from an available EM map before using HEMNMA, but it is directly used to obtain a refined atomic model in terms of conformation for each particle image. Indeed, HEMNMA and DeepHEMNMA allow obtaining an atomic-scale conformational landscape when the initial conformation is given by an atomic structure and a “pseudoatomic”-scale conformational landscape when the initial conformation is given by an EM map. The pseudoatoms are 3D Gaussian functions that represent the EM map and replace atoms in the 3D-to-2D fitting based on normal modes [43]. Additionally, HEMNMA and DeepHEMNMA can produce EM map reconstructions in the conformational landscape, by assembling images with similar conformations in 3D reconstructions.

The methods based on normal modes perform fast changes of the given conformation using a few variables (the amplitudes of a few normal modes) [35-39]. However, large amplitudes of motion along normal modes may induce distortions of the conformational model. Model distortions are usually avoided using Molecular Dynamics (MD) simulation approaches (deterministic) or Monte Carlo simulations approaches (stochastic), as in the case of 3D-to-3D flexible fitting of an atomic initial conformation against a cryo-EM map [44-50].

These methods produce good quality structures using a large number of variables (degrees of freedom), determined by the number of atomic coordinates. However, they have a very high computational cost compared to the methods based on normal modes and, until very recently, they remained computationally prohibitive for characterizing continuous conformational variability in large and highly heterogeneous sets of single particle images. For instance, a stochastic approach called Natural Moves Monte Carlo, which refines an initial conformation using a 2D particle class average, was used with image datasets containing mixtures of only two conformational states or conformations very close to these two states (e.g., open and closed) [51]. The method refines the conformation by employing a coarse-grained model composed of rigid segments representing domains connected by flexible loops and by sampling the conformations using rigid-body degrees of freedom of the domains while analyzing a high-SNR 2D class average. The Natural Moves Monte Carlo method has a low robustness to noise and cannot be used with individual, low-SNR single particle images [51]. In a more recent study, a 1D conformational landscape was obtained by generating samples from a posterior distribution using a Monte Carlo approach to determine a 1D free energy profile from single particle images, based on a pre-determined path collective variable [5]. However, the method cannot be easily extended to map images onto conformational landscapes of higher dimensions (e.g., 2D or 3D) and its success depends on the choice of the path collective variable. The determination of the path collective variable requires the availability of at least two relevant atomic conformational models, which might be difficult to obtain for some systems.

Recently, we combined the displacement based on normal modes and the displacement based on MD simulation within a 3D-to-3D flexible fitting method named NMMD (Normal Mode and Molecular Dynamics), which proved to speed up fitting of atomic initial conformations against cryo-EM maps while avoiding structural distortions induced by normal modes [52]. NMMD simultaneously integrates global atomic displacements (along normal modes) and local atomic displacements (based on MD simulation) [52]. The gain in speed obtained by NMMD opens doors for MD-based analysis of continuous conformational variability in large and highly heterogeneous sets of single particle images.

To allow a fast MD-based analysis of individual single particle images, we developed a method for 3D-to-2D flexible fitting of an initial atomic conformation against a given particle image that is able to use a combination of the displacement based on normal modes and the displacement based on MD simulation. The normal-mode empowered MD-based 3D-to-2D flexible fitting method uses pre-determined particle orientations and positions (e.g., by projection matching of the image with the initial conformation), which are refined during the fitting. However, this 3D-to-2D flexible fitting method alone revealed to be suboptimal for accurate extraction of conformational landscapes due to a difficulty to individually fit some of the particle views. To tackle this issue, we developed an iterative approach that extracts the information about the principal motion directions from the ensemble of the fitted conformations (obtained for a set of particle images through individual 3D-to-2D flexible fitting of each image) and uses this information for the next

round of the individual 3D-to-2D flexible fitting by replacing normal modes by the principal motion directions obtained in the previous round. The iterative refinement of the principal motion directions extracted from the ensemble of the fitted conformations helps to iteratively refine the individual 3D-to-2D flexible fitting of the particle images, which in turn results in the refinement of the entire conformational landscape.

In this article, this iterative approach is introduced and referred to as MDSPACE, which stands for Molecular Dynamics simulation for Single Particle Analysis of Continuous Conformational hEterogeneity. To the best of our knowledge, MDSPACE is the first method for analyzing conformational variability from cryo-EM single particle images that is based on classical mechanics MD simulation, meaning deterministic trajectories describing the evolution of the atomic model over time. MDSPACE can obtain the conformational landscapes from highly heterogeneous image datasets containing continuous conformational heterogeneity, by analyzing individual low-SNR single particle images. MDSPACE requires a single input atomic structure, which is used as the initial conformation for fitting each particle image. As shown in our experiments, MDSPACE can recover the conformational space even when the initial conformation is distant from the target conformations in the particle images. Furthermore, MDSPACE allows obtaining conformational landscapes of any dimension (1, 2, 3, etc.), which can be visualized using 1D histograms, 2D plots, or 3D plots.

The MDSPACE method is described in **Section 2**. **Section 3** shows the performance of the new method with synthetic data of the heterodimeric ABC exporter TmrAB [26] and with experimental cryo-EM data of yeast 80S ribosome-tRNA complexes from EMPIAR-10016 dataset [21]. A general discussion and conclusion are provided in **Section 4**.

2. Methods

MDSPACE

MDSPACE is an iterative approach (**Figure 1**) for analyzing continuous conformational variability in single particle images based on MD simulation. The MD simulation is guided by a 2D biasing potential. MDSPACE iteratively refines an initial conformation of the particle in each image. The initial conformation is the same for all particle images and determined by a given atomic structure. Also, MDSPACE iteratively refines an initial particle orientation and position in each image (the initial rigid-body alignment obtained by classical approaches based on classification and projection matching or by other continuous conformational variability approaches prior to using MDSPACE). In the first iteration, a provided atomic structure is flexibly fitted to each particle image, independently of other images, using an original 3D-to-2D flexible fitting approach based on normal-mode empowered MD simulation (indicated as “NMMD step” in **Figure 1**) and the given initial particle orientation and position in the image. The obtained ensemble of fitted atomic models is then rigid-body aligned to the initial atomic structure and principal component analysis (PCA) is performed on the rigid-body aligned models. The principal components obtained by PCA represent the dominant conformational changes extracted from the ensemble of fitted atomic models in one iteration of MDSPACE. In the next

iteration of MDSPACE, the previously extracted principal component vectors replace the normal mode vectors used in the "NMMD step", to incorporate the ensemble information in the fitting of individual particles (principal-component empowered MD simulation indicated as "PCMD step" in **Figure 1**). Each new iteration of MDSPACE (involving a new round of 3D-to-2D flexible fitting) is based on principal-component empowered MD simulations that encourage the conformation to move along the principal component vectors representing the dominant motions, with the effect of refining the fitting. Rigid-body movements that occurred during the MD-based flexible fitting are measured at the end of each MDSPACE iteration and are then used to refine the parameters of the initial rigid-body alignment of the particle images for the next MDSPACE iteration. In each MDSPACE iteration, it is possible to visualize the PCA space (determined by the first few principal axes), individual atomic models, movies of atomic-model displacements along different directions in this space, or 3D reconstructions from the groups of images with similar particle conformations (close points in the PCA space) along these directions. These directions can be the principal axes or can be determined by a path traversing the densest regions in this space. The information obtained from the PCA space can be compared between different iterations.

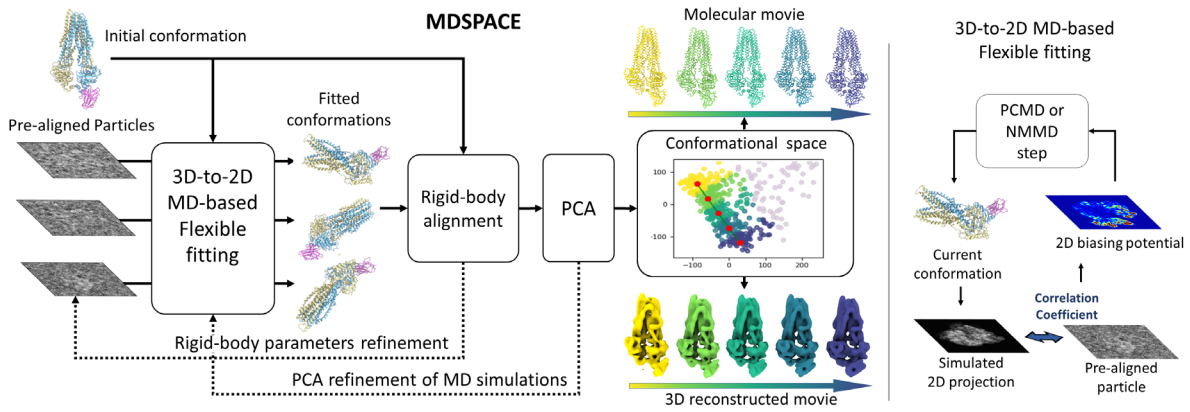

Figure 1 : Flowchart of the MDSPACE method (left) proposed for iterative continuous conformational analysis of single particle images, which is based on a 3D-to-2D flexible fitting approach (right) that can use normal-mode empowered MD simulation (indicated as "NMMD step" in this figure) or principal-component empowered MD simulation (indicated as "PCMD step" in this figure). The MD simulation is guided by a 2D biasing potential (right). The dotted lines represent the iterative process, which may be repeated several times to refine the conformational space.

3D-to-2D flexible fitting using MD simulations

In 3D-to-3D flexible fitting approaches based on MD simulation [44-48, 52], MD simulation drives an initial atomic conformation towards the target conformation defined by a cryo-EM map, thanks to the so-called biasing potential that is added to the classical MD force field potential and depends on the cryo-EM map. For instance, the biasing potential can be defined as the correlation coefficient (CC) between the cryo-EM map and a map simulated from the atomic model to guide the motion of the molecule following MD algorithm [45] or normal-mode empowered MD algorithm NMMD [52] so that the lowest potential value is achieved

for the highest CC, i.e., when the atoms fit well the cryo-EM map. Then, the effective potential used for the fitting is as follows [45, 52]:

$$V = V_{ff} + k(1 - CC) \quad (1)$$

where V_{ff} in the potential from the classical MD-based force field, CC is the correlation coefficient between the cryo-EM map and the map simulated from the atomic model, and k is the force constant that determines the weight given to the biasing potential with respect to the classical MD potential. The force constant is an important parameter as it allows the MD simulation to go through energy barriers and explore the conformational space.

In the method proposed here for 3D-to-2D flexible fitting of an atomic model to a single particle image, we use the total potential (**Eq.** 1) where the biasing potential term is defined as the CC between the given particle image (p_{exp}) and a 2D projection simulated using the atomic model (p_{sim}). This CC is defined as follows:

$$CC = \frac{\sum_{i,j} p_{sim}(i,j) p_{exp}(i,j)}{\sqrt{\sum_{i,j} p_{exp}^2(i,j) \sum_{i,j} p_{sim}^2(i,j)}}, \quad (2)$$

where $\sum_{i,j}(\cdot)$ is the sum over all pixel indexes (i,j) in the image, and the simulated projection and the particle image have the same number of pixels in each dimension and the same pixel size.

The force applied to move the atoms is obtained by deriving the potential energy with respect to the atomic coordinates, $\boldsymbol{X}_n = (x^n, y^n, z^n), n = 1, N$, as:

$$\boldsymbol{F} = -\frac{\partial}{\partial \boldsymbol{X}_n(l)} V, l = 1,3, \quad (3)$$

where N is the number of atoms, and $\boldsymbol{X}_n(1) = x^n$, $\boldsymbol{X}_n(2) = y^n$, $\boldsymbol{X}_n(3) = z^n$, $n = 1, N$.

By combining **Eq.** 1 and **Eq.** 3, the total force $\boldsymbol{F}$ can be written as follows:

$$\boldsymbol{F} = \boldsymbol{F}_{ff} - k \frac{\partial}{\partial \boldsymbol{X}_n(l)} CC, \quad (4)$$

where $\boldsymbol{F}_{ff}$ is the force resulting from the classical MD-based force field.

The derivatives of the CC are calculated using:

$$\frac{\partial CC}{\partial \boldsymbol{X}_n(l)} = \frac{\sum_{i,j} p_{exp}(i,j) \frac{\partial p_{sim}(i,j)}{\partial \boldsymbol{X}_n(l)}}{\sqrt{\sum_{i,j} p_{exp}^2(i,j) \sum_{i,j} p_{sim}^2(i,j)}} - \frac{\left(\sum_{i,j} p_{sim}(i,j) \frac{\partial p_{sim}(i,j)}{\partial \boldsymbol{X}_n(l)}\right)\left(\sum_{i,j} p_{sim}(i,j) p_{exp}(i,j)\right)}{\sqrt{\sum_{i,j} p_{exp}^2(i,j)} \left(\sum_{i,j} p_{sim}^2(i,j)\right)^{\frac{3}{2}}}. \quad (5)$$

In the following, we show how the simulated 2D projection (p_{sim}) and its derivatives ($\frac{\partial p_{sim}}{\partial \boldsymbol{X}_n(l)}$) are obtained.

A continuous representation of an atomic model is commonly simulated by placing a 3D Gaussian function at each atomic position [45]. A 2D projection of such a Gaussian-based representation of an atomic model,

at an orientation determined by a rotation matrix $\boldsymbol{R}$ and a position determined by a translation vector $\boldsymbol{T}$, is defined as follows (projection along the *z*-axis):

$$p_{sim}(x,y) = \frac{1}{2\pi\sigma^2}\sum_{n=1}^{N} e^{-\frac{1}{2\sigma^2}\left((x-r_x^n)^2+(y-r_y^n)^2\right)}, \quad (6)$$

$$\begin{pmatrix} r_x^n \\ r_y^n \\ r_z^n \end{pmatrix} = \boldsymbol{R}^{-1}\begin{pmatrix} x^n \\ y^n \\ z^n \end{pmatrix} - \boldsymbol{T} = \boldsymbol{R}^{-1}\ \boldsymbol{X}_n - \boldsymbol{T}, \quad n = 1, N,$$

where σ is the standard deviation of the Gaussian functions, N is the number of atoms in the model with the coordinates $\boldsymbol{X}_n = (x^n, y^n, z^n), n = 1, N$, and r_x^n and r_y^n are the *x* and *y* coordinates of the n-th atom rotated and translated with a 3 × 3 rotation matrix $\boldsymbol{R}$ and a 3 × 1 translation vector $\boldsymbol{T}$. Note that the projection depends on the pose ($\boldsymbol{R}$, $\boldsymbol{T}$) and on the provided atomic coordinates $\boldsymbol{X}_n$.

Let us write **Eq. 6** as:

$$p_{sim}(x,y) = \sum_{n=1}^{N} p_{sim}^n(x,y), \quad (7)$$

where $p_{sim}^n(x,y)$ is the contribution of the n-th atom to the projection, given by:

$$p_{sim}^n(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left((x-r_x^n)^2+(y-r_y^n)^2\right)}. \quad (8)$$

Using the expression in **Eq. 7**, the derivative of the simulated projection with respect to the atomic coordinates $\boldsymbol{X}_n(l), l = 1{,}3$ ($\boldsymbol{X}_n(1) = x^n$, $\boldsymbol{X}_n(2) = y^n$, and $\boldsymbol{X}_n(3) = z^n$) can be written, after the discretization on a regular image pixel grid (pixel coordinates(i,j)), in the following form:

$$\frac{\partial p_{sim}(i,j)}{\partial \boldsymbol{X}_n(l)} = \frac{1}{\sigma^2} p_{sim}^n(i,j)\boldsymbol{Q}_{i,j}(l), \quad l = 1{,}3, \quad (9)$$

where $\boldsymbol{Q}_{i,j}(l)$ is the l-th element ($l = 1{,}3$) of the vector $\boldsymbol{Q}_{i,j}$:

$$\boldsymbol{Q}_{i,j} = \boldsymbol{R}\begin{pmatrix} i - r_x^n \\ j - r_y^n \\ 0 \end{pmatrix}. \quad (10)$$

3D-to-2D flexible fitting using normal-mode empowered MD simulations

In the NMMD method [52], which performs 3D-to-3D flexible fitting, a set of normal modes, describing the most collective motions of the structure, is incorporated in MD simulation through a simultaneous integration of the conformational parameters related to normal modes (normal-mode amplitudes) and the conformational parameters related to MD simulation (Cartesian atomic coordinates) [52]. The NMMD method uses the force vector on MD-induced atomic coordinate displacement ($\boldsymbol{F}_{ff}$) to implement the force vector on normal-mode amplitudes ($\boldsymbol{F}_q$), according to the following linear relationship between them [52]:

$$\boldsymbol{F}_q = \boldsymbol{A}\,\boldsymbol{F}_{ff}, \quad (11)$$

where $\boldsymbol{A}$ is the matrix of normal mode vectors.

The NMMD algorithm encourages the simulation to move along the normal modes to faster perform global dynamics while increasing the computational efficiency of MD-based local dynamics.

In the first iteration of MDSPACE, 3D-to-2D flexible fitting is performed using the NMMD algorithm (simultaneous integration of the global displacement along normal modes and the MD-based local displacement), which accelerates flexible fitting of a given initial atomic conformation to target conformations in cryo-EM single particle images (the refinement of the atomic conformation against each particle image). In the next iterations of MDSPACE, the normal modes in NMMD are replaced by the principal component vectors that are extracted from the conformational ensemble obtained at the previous MDSPACE iteration, which iteratively refines the conformation and rigid-body alignment parameters of the particle in each image, as described in the subsection "**PCA-based refinement of MD simulations**".

Initial orientation and position of the particles by rigid-body pre-alignment

The MDSPACE method presented in this article assumes that a set of rotations $\boldsymbol{R}_1$ and translations $\boldsymbol{T}_1$, corresponding to $\boldsymbol{R}$ and $\boldsymbol{T}$ in **Eq. 6**, have been obtained by pre-alignment of the initial conformation with the particle images (regardless of the method used for this pre-alignment). Indeed, although rigid-body motions are allowed in MD simulations, it is preferable for the initial conformation to be pre-aligned with the particle images in order to prevent the simulation to get trapped into local minima.

The accuracy of the pre-alignment depends on a multitude of factors. For instance, as in the case considered in this article, where large continuous conformational heterogeneity is present in the particle images and rigid-body alignment methods are used for the pre-alignment (e.g., based on projection matching between the initial conformation and the particle images), the pre-alignment errors will be larger. In this article, we show that MDSPACE can refine the initial rigid-body alignment. We show that a coarse rigid-body pre-alignment is sufficient to guide the MD simulation in the right direction, as rigid-body alignments are further performed during the MD-based 3D-to-2D flexible fitting (fine rigid-body movements of the atomic model during MD simulation to finely match the particle in the image), as explained in the next subsection.

Refinement of the initial orientation and position of the particles

At each MDSPACE iteration and for each particle image, two sets of values of the rotation and translation parameters can be distinguished: initial and final. The initial values at an MDSPACE iteration are used to determine the pose of the projection image (**Eq. 6**), which remains unchanged until the end of that MDSPACE iteration. For the first MDSPACE iteration, the initial rotation and translation parameter values ($\boldsymbol{R}_1$ and $\boldsymbol{T}_1$, respectively) are obtained by rigid-body pre-alignment (as explained in the previous subsection). During each MDSPACE iteration, the atomic structure undergoes rigid body transformations, driven by the MD-biasing potential towards the target particle conformation and position in the image. The final rotation and translation parameter values at an MDSPACE iteration are obtained by combining the initial values ($\boldsymbol{R}_1$ and

$\boldsymbol{T}_1$) and the values ($\boldsymbol{R}_2$ and $\boldsymbol{T}_2$) estimated through a rigid-body alignment between the initial conformation and the finally fitted conformation (by minimizing the RMSD between the two conformations). The final rotation and translation parameter values obtained at one MDSPACE iteration ($\boldsymbol{R}_3$ and $\boldsymbol{T}_3$, respectively) are used as the initial values (new $\boldsymbol{R}_1$ and $\boldsymbol{T}_1$) for the next MDSPACE iteration.

At each MDSPACE iteration, the rotation $\boldsymbol{R}_2$ and the translation $\boldsymbol{T}_2$ are obtained by estimating rigid-body motion between the initial conformation (atomic coordinates $\boldsymbol{r}_i^n, n = 1, N$) and the finally fitted conformation (atomic coordinates $\boldsymbol{r}_f^n, n = 1, N$), by the following minimization of the RMSD between these two conformations:

$$\min_{\boldsymbol{R}_2, \boldsymbol{T}_2} \sqrt{\frac{1}{N}\sum_{n=1}^{N} \| \boldsymbol{r}_f^n - (\boldsymbol{R}_2 \cdot \boldsymbol{r}_i^n + \boldsymbol{T}_2) \|^2}, \qquad (12)$$

where N is the number of atoms. The RMSD minimization is performed using an optimization algorithm based on singular value decomposition that is available in BioPython [53].

At each MDSPACE iteration, the final rotation and translation parameter values ($\boldsymbol{R}_3$ and $\boldsymbol{T}_3$, respectively) are obtained by combining the initial parameter values ($\boldsymbol{R}_1$ and $\boldsymbol{T}_1$) with the parameter values extracted from the MD simulation ($\boldsymbol{R}_2$ and $\boldsymbol{T}_2$), as follows:

$$\boldsymbol{R}_3 = \boldsymbol{R}_2 \cdot \boldsymbol{R}_1, \qquad \boldsymbol{T}_3 = \boldsymbol{R}_2 \cdot \boldsymbol{T}_1 + \boldsymbol{T}_2, \qquad (13)$$

The $\boldsymbol{R}_3$ and $\boldsymbol{T}_3$ parameter values are then used as the refined values of the initial rotation and translation parameters for the next MDSPACE iteration (in place of $\boldsymbol{R}_1$ and $\boldsymbol{T}_1$, respectively).

PCA-based refinement of MD simulations

The 3D-to-2D flexible fitting accurately fits most particle images. However, some images are more difficult to fit, especially the images associated with specific particle views for which the conformational change is less detectable or ambiguous in the projection plane (e.g., a motion mainly along the projection axis will affect the 2D projection but the conformational change will be less detectable in the projection). In such cases, the fitting will most likely induce no displacement from the initial atomic positions (the initial conformation) or perform small-scale, local rearrangements. Most likely, it will not find the correct conformation in images with such views. This is a direct consequence of fitting each particle image individually (independently of other images) in the presence of high level of noise in the images.

To make the approach more robust to particle views and noise, MDSPACE involves an iterative approach that refines the MD-based 3D-to-2D flexible fitting of individual particle images by incorporating the ensemble conformational information into MD simulation, which encourages MD simulations of the particle images with “bad” views to follow the principal motion directions learned from the images with “good” views at the previous iteration. In this approach, the information about the principal motions learned at one iteration is

used as a prior to reanalyze the dataset at the next iteration, which boosts the MD-based fitting of individual particle images making it robust to difficult views and noise. For this reason, at the end of each MDSPACE iteration, the conformations obtained by individually fitting different particle images are rigid-body aligned with respect to the initial conformation and then analyzed using PCA. The PCA results in principal component vectors that represent global motions seen in data at that MDSPACE iteration, which are then incorporated in a new round of the fitting at the next MDSPACE iteration. More precisely, the PCA-space conformational information is incorporated into normal-mode empowered MD-based flexible fitting by replacing normal-mode vectors by principal component vectors of the PCA space, which we refer to as PCA-empowered MD-based flexible fitting. This encourages the simulation to move along the principal component vectors found in the previous MDSPACE iteration. The PCA-empowered MD-based flexible fitting yields a new set of atomic structures, whose principal components can be obtained by PCA and then incorporated in a new round of the fitting. MDSPACE, which alternates the fitting with the PCA analysis, refines the PCA space over the iterations, making it closer to the target conformational landscape, as illustrated in **Figure 1**.

It is worth noting that the PCA-empowered MD-based flexible fitting in MDSPACE not only refines the flexible fitting (the finally fitted conformations) but also refines the initial rigid-body alignment of the particle images. More precisely, at the end of each iteration of MDSPACE, the rigid-body parameters that initiated this iteration are combined with the alignment parameters extracted from the MD simulation performed at the same iteration (following **Eq. 13**) and the obtained refined rigid-body parameters are used to start the new MDSPACE iteration.

General recommendation for running MDSPACE iteratively

To increase speed and accuracy of analyzing large datasets of single particle images, we generally recommend using normal-mode empowered MD simulations for the first MDSPACE iteration and using principal-component empowered MD simulations (PCA-based refinement) for all other iterations. Adding normal modes to MD simulation accelerates the fitting. Adding principal component vectors from the previous MDSPACE iteration to MD simulation in the next MDSPACE iteration improves robustness to difficult views and noise. Regarding the number of principal component vectors to use for the PCA-based refinement, 3 principal component vectors may generally be enough, but possibly more will be required for some systems. The number of principal component vectors can be selected based on the observed decrease in the singular values of the PCA components. The computational cost of adding more principal component vectors is negligible.

A coarse-to-fine data processing scheme can be used to additionally speed up MDSPACE processing (e.g., in the case of large data sets of large complexes such as ribosomes). More precisely, at the first MDSPACE iteration (normal-mode empowered MD simulations), the principal components of the conformational

variability can be learned by processing a small data subset, which will result in a coarsely estimated low-dimensional conformational landscape. At the next MDSPACE iteration (principal-component empowered MD simulation), the conformational landscape can be refined using a larger number of images and the principal components obtained at the previous MDSPACE iteration.

Software implementation

MD simulation, Normal Mode Analysis, and PCA methods and software: The iterative MDSPACE method requires a large number of computationally costly MD simulations, as MD-based fitting is applied several times for each particle (*M* times per particle, for *M* iterations of MDSPACE). To reduce the computational cost, we choose a coarse-grained approach for MD simulations using off-lattice Cα Gō model as described by Clementi and collaborators [54]. The Cα Gō model can be extended to Phosphorus atoms, as done in this article for the experiments with 80S ribosome-tRNA cryo-EM dataset [21]. Off-lattice Gō models can successfully capture native dynamics and conformational transitions of diverse systems [55] using much smaller computational resources than all-atom simulations. However, the non-local interaction in Gō-like models are determined from a "native state" corresponding to the experimental structure, which tends to bias the dynamics towards the experimental structure [56]. In the case of smaller systems (e.g., much smaller than the ribosome studied in this article), one could replace the Cα Gō model by all-atom simulation as it would avoid such bias. Gō models were obtained by SMOG2 software [57]. As shown in this article, short 30-picosecond MD simulations (used with both synthetic and experimental datasets) are sufficient for 3D-to-2D flexible fitting with MDSPACE. In all experiments in this article, MD simulations were conducted using the MD simulation software GENESIS [58], which is also available as part of ContinuousFlex [40] (a plugin for the image processing software package Scipion [41]) that now also includes MDSPACE software. The software for calculating normal modes here is ElNemo [59], which is also available in ContinuousFlex. ElNemo accelerates normal mode analysis using the rotation-translation block approach [60]. Both GENESIS and ElNemo were used in our previous work on NMMD [52]. PCA is obtained by probabilistic principal component analysis implementation in Scikit-Learn [61, 62].

PCA space clustering and 3D reconstructions: To interpret continuous conformational variability from a low-dimensional space such as a PCA space, we performed clustering of close points in the PCA space and grouping of the corresponding particle images into 3D reconstructions to visualize conformational variability in terms of EM maps. A new clustering tool was developed to allow a manual or automated drawing of a trajectory of points in the PCA space and performing an automatic clustering of each particle image to the closest point on the trajectory. An automatically generated trajectory consists of a set of points regularly spaced on a line that can be drawn along a PCA axis or any other direction in the PCA space. The trajectory points can be manually dragged to adjust to data distribution. The resulting clusters of particle images in the

PCA space can be manually refined (by adding or removing data, including removal of outliers). The 3D reconstruction of the EM maps from the clusters is obtained by direct Fourier interpolation using Xmipp [63, 64].

***MDSPACE software*:** MDSPACE is implemented as part of the ContinuousFlex plugin [40] for Scipion [41], which also offers NMMD and HEMNMA/DeepHEMNMA software. As NMMD, MDSPACE uses GENESIS for MD simulations and ElNemo for calculating normal modes. The 3D-to2D flexible fitting is implemented in GENESIS by modifying the 3D-to-3D flexible fitting algorithm [45] to use 3D-to-2D projection and 2D biasing potential. The Scipion frontend of MDSPACE allows highly parallelized calculations (MPI protocol, CPU processing) and running the software on high performance computers. Additionally, it provides an intuitive and user-friendly graphical interface for running MDSPACE, which also simplifies the reproducibility of the analyses.

3. Results

3.1 Experiment with synthetic cryo-EM data

In this subsection, we show the performance of MDSPACE using synthetic data of the heterodimeric ABC exporter TmrAB [26].

Synthetic dataset of TmrAB

To assess the performances of MDSPACE in a controlled environment (with a known ground-truth solution), we synthesized a dataset by simulating experimental cryo-EM conditions as much as possible. The system studied is a heterodimeric ABC exporter, TmrAB, composed of two ABC proteins, TmrA and TmrB, in complex with a nanobody, Nb9F10 (**Figure 2a**). Multiple atomic models of the complex in different conformations, derived from cryo-EM maps, are available in the Protein Data Bank (PDB) [26] including an outward-facing conformation (PDB-6RAH, designated here as $TmrAB_{OF}$) and an inward facing conformation (PDB-6RAF, designated here as $TmrAB_{IF}$). $TmrAB_{IF}$ has a closed extracellular gate and an open intracellular gate (**Figure 2a**), whereas $TmrAB_{OF}$ has an open extracellular gate and a closed intracellular gate (**Figure 2b**, left).

In this experiment, $TmrAB_{OF}$ was used as the initial conformation for 3D-to-2D flexible fitting against synthetic images, whereas the $TmrAB_{IF}$ was used with its normal modes 7 and 8 to generate multiple synthetic conformations from which images were generated, by randomly sampling a continuous trajectory of the conformational transition defined by these two normal modes. We note here that normal modes are ordered according to their frequency and that the first 6 lowest-frequency normal modes (modes 1-6) are never used, as representing rigid-body displacements. The $TmrAB_{IF}$'s mode 7 describes opening and closing of the

intracellular gate and mode 8 describes a rotation of the intracellular part of TmrB together with the nanobody, as shown in **Figure 2b**. To synthesize multiple conformations of the complex, we performed random atomic displacements of $TmrAB_{IF}$ along modes 7 and 8, using the same normal-mode amplitude for both normal modes, where this normal-mode amplitude was sampled from a random uniform distribution in the range [-100,100]. This procedure was used to generate 500 different conformations along the trajectory. **Figure 2c** shows the ground-truth synthetic continuous conformational variability that is projected onto the first two principal component axes (ground-truth two-dimensional PCA space), together with the projections of $TmrAB_{IF}$ and $TmrAB_{OF}$ on the same PCA space. By observing the relative distances between the points in the ground-truth PCA space (**Figure 2c**), one can notice that the initial conformation for 3D-to-2D fitting against the synthetic images ($TmrAB_{OF}$) is significantly far from the target conformations (ground-truth trajectory). More precisely, the average root mean square deviation (RMSD) between $TmrAB_{OF}$ and the synthetic conformations generated from $TmrAB_{IF}$ is 8.1 Å.

The synthetic conformations were converted into high-resolution density maps using a method based on scattering factors [65] and projected (1 projection per conformation) using ray casting in real space to obtain a set of 2D particle images of 128 × 128 pixels and a pixel size of 2 Å × 2 Å. The projection orientations followed a uniform random angular distribution around a 3D sphere and the particles in the images were shifted following a uniform random distribution in the [-5,5] pixel range (i.e., [-10,10] Å). We simulated the effect of the electron microscope by applying a contrast transfer function (CTF) with noise added before and after the CTF, using the image formation model described elsewhere [66]. We synthesized realistically looking images using a simulated 200 kV microscope with a spherical aberration of 2 mm, a defocus of -0.5 µm, and Gaussian noise distributed during the image formation so as to simulate a SNR of 0.1 in the images.

Both the low SNR of the synthetic images and the large RMSD between the initial and target conformations (average RMSD of 8.1 Å) are challenges for MDSPACE. For each synthetic image, the 3D-to-2D flexible fitting of $TmrAB_{OF}$ should close the extracellular gate and precisely fit the opening of the intracellular gate and the rotation of the nanobody, in the presence of high noise in the images and using 2D biasing potential.

Rigid-body pre-alignment and MDSPACE iterations

To rigid-body pre-align the initial conformation with the particle images, we performed 4 iterations of projection matching between the images and the density map from the initial conformation, in Xmipp [63, 64], using the angular sampling rate of 7°, 5°, 3°, and 2° over the 4 iterations as well as no limit on the angular search in the first iteration and a limit on the angular search range of 10°, 6°, and 4° in the other three iterations. As the initial conformation is very different from the target conformations in the images, the projection matching gives only a rough alignment. In this article, we demonstrate that MDSPACE is able to

refine such rough initial rigid-body alignment while also iteratively refining deciphering of the conformational variability.

We run 4 iterations of MDSPACE using 30-picosecond MD simulations and the force constant of 2000 kcal/mol on the dataset of 500 synthetic particle images. Although our general recommendation is to run the first iteration of MDSPACE using normal-mode empowered MD-based fitting and all other iterations using principal-component empowered MD-based fitting (as described in the **Methods** section), the fitting here was performed without normal modes to avoid bias as the images were synthesized using normal modes. More precisely, flexible fitting using only MD simulations was used in the first iteration and PCA-based refinement in the other 3 iterations of MDSPACE. Starting from the second iteration, the PCA-based refinement at each iteration was performed using the first 3 principal components (3 most dominant motions) obtained at the previous iteration. As shown below, this protocol refines both the initial rigid-body alignment (pre-alignment) and the conformational variability retrieval (fitted conformations).

Performance regarding recovery of the ground-truth rigid-body parameters

The MDSPACE analysis of the synthetic dataset shows that the rotation and translation parameters obtained at the pre-alignment step get refined over the iterations (the first two plots of **Figure 2d,** from left to right). **Figure 2d** shows the angular and shift error distributions over the iterations, i.e., per-iteration statistics on the errors between the ground-truth and estimated angles (the first plot from the left) and shifts (the second plot from the left). Recall here that the estimates of the angles and shifts at each iteration are obtained by combining the initial angles and shifts at that iteration and the angles and shifts extracted from the MD simulation at the same iteration (**Eq. 13**). 0.

The median angular error of the pre-alignment is 11.3 degrees and the median shift error is 3.6 Å, meaning that the projection matching resulted in a relatively poor rigid-body pre-alignment. Even with such a poor pre-alignment, the recovery of the ground-truth angles and shifts at the 4^{th} iteration of MDSPACE is satisfactory (the median angular and shift errors dropped to 3.3 degrees and 0.25 Å, respectively).

Performance regarding recovery of the ground-truth continuous conformational transition

The MDSPACE analysis of the synthetic dataset shows that the conformational space gets refined over the iterations (**Figure 2e**). **Figure 2e** shows the evolution of the conformational space estimation over four MDSPACE iterations, together with the ground-truth conformations (the samples of a line-form trajectory in the normal mode space in this experiment) and the initial conformation used for the 3D-to-2D flexible fitting. For simplicity of the comparison over the iterations, we here show the conformational space reduced to two dimensions (2D PCA space), whereas the first three principal components were used for the PCA-based refinement (from the second MDSPACE iteration on). Also, it should be recalled that the PCA is recalculated at each MDSPACE iteration (PCA components change over the iterations) and that the PCA components calculated at one iteration are used to empower the MD-based flexible fitting at the next iteration. However,

for the purpose of a visual comparison between the fitted and target (ground-truth) conformations, **Figure 2e** shows the fitted conformations in the same space as the ground-truth conformations. The principal components actually used to empower the MD-based flexible fitting (in the PCA-based refinement) differ from those shown in **Figure 2e** as they do not comprise the ground-truth information.

At the first MDSPACE iteration, we observe that a part of the fitted conformations is close to the ground-truth conformations, but many are far (approximatively 50 % are close to the initial conformation used for the 3D-to-2D flexible fitting and they correspond to the images with the projection direction with poor conformational variability information) (**Figure 2e**). After the second MDSPACE iteration (first PCA refinement), we observe a clear improvement of the conformational space as the fitted conformations get closer to the target conformations (ground-truth trajectory). This tendency continues over the iterations and, at the last iteration, most fitted conformations are close to the target conformations and only a few remain close to the initial conformation.

The analysis of the CCs between the particle images and the projections of the fitted conformations and the analysis of the RMSDs between the fitted and target (ground-truth) conformations, shown in the last two plots of **Figure 2d** (from left to right), confirm that the conformations get refined over the iterations. The CC distribution shows a significant CC increase from the initial pre-alignment iteration to the first iteration (the median CC value increased from 0.33 to 0.42), then the CC increases slowly over the following iterations (the median CC values of 0.42, 0.43 and 0.44 for iterations 2, 3 and 4). The RMSD distribution shows that the RMSD decreases significantly up to the third iteration (the median RMSD values of 8.4 Å and 2.6 Å at iterations 1 and 3) and then changes slowly.

Recovery of the ground-truth continuous conformational transition in terms of animations and 3D reconstructions from the conformational space

We analyzed the conformational distribution in the PCA space obtained at the 4th iteration of MDSPACE in terms of 3D reconstructions from clusters of close points in this space, i.e., from the corresponding synthetic particle images, in order to check whether 3D reconstructions along the point distribution in the PCA space follow the ground-truth conformational transition trajectory.

The 3D reconstructions were performed from clusters determined in the PCA space as described in the **Methods** section (**Figure 3a**). First, we interactively defined a 5-point linear trajectory that approximatively fits the point distribution in the PCA space. Then, we performed an automated clustering based on the closest points to each of the 5 points and interactively removed the outlier points. Finally, we performed a 3D reconstruction from each of the 5 clusters without outliers, using the rigid-body parameters obtained after the 4th iteration of MDSPACE, which resulted in five reconstructed EM maps.

Figure 3b shows two atomic conformations corresponding to two extremums of the ground-truth synthetic trajectory (designated as IF1 and IF2) and **Figure 3c** shows the EM maps obtained from the first and last

clusters along the 5-point linear trajectory (**Figure 3a**). By comparing these two 3D reconstructed maps (**Figure 3c**) with the corresponding ground-truth synthetic atomic conformations (**Figure 3b**), we observe that the method was able to capture the ground-truth synthetic conformational change.

Supplementary Movies S1 and S2 show animations of the fitted atomic conformations and the reconstructed EM maps from clusters along the 5-point linear trajectory in the PCA space, respectively.

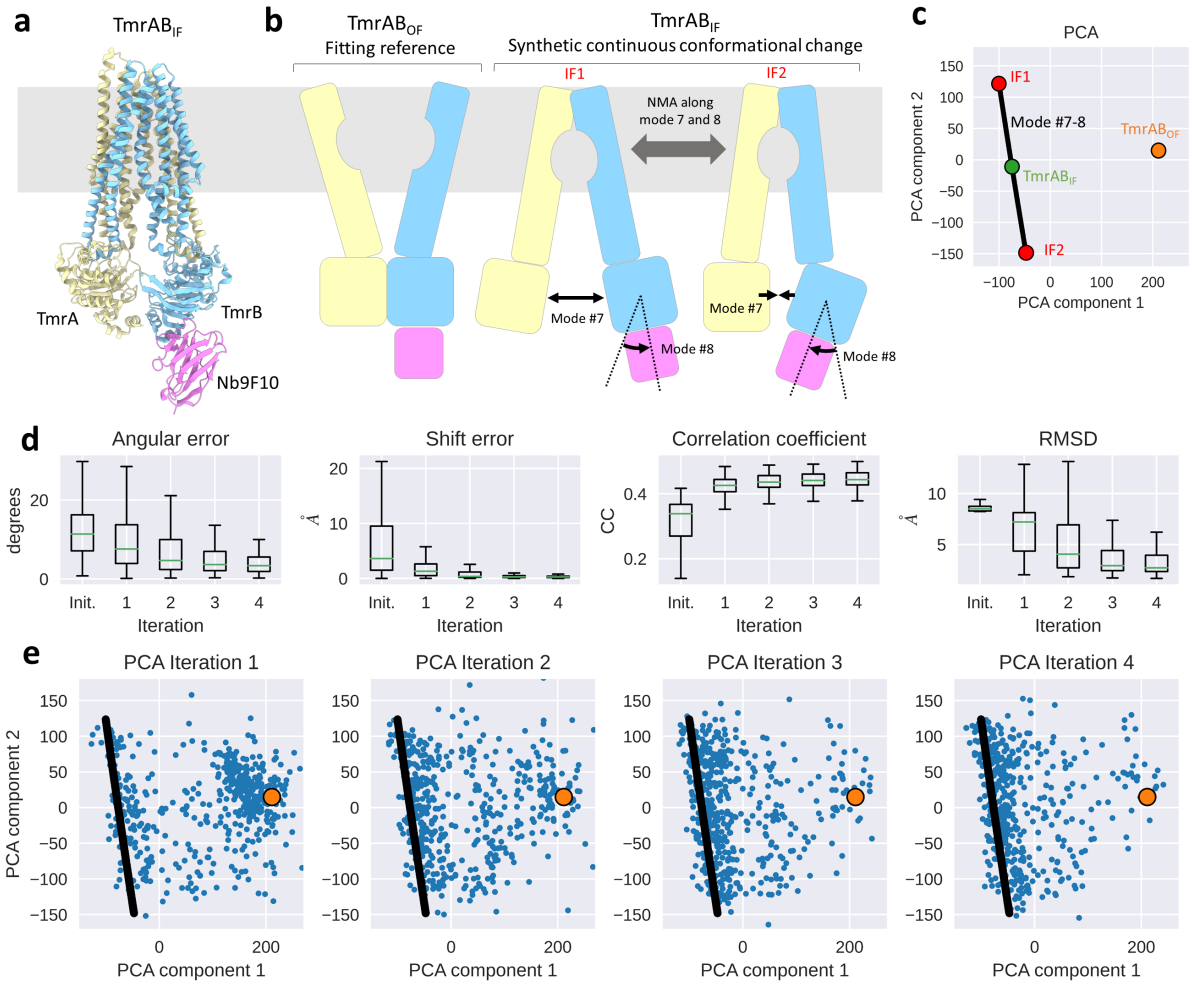

Figure 2 : MDSPACE analysis of the synthetic dataset of TmrAB. (a) Structure of TmrAB in inward-facing conformation (TmrAB$_{IF}$). (b) Diagram of TmrAB in outward-facing conformation (TmrAB$_{OF}$, the initial conformation for fitting) and the synthetic continuous conformational change simulated from TmrAB$_{IF}$ using modes 7 and 8. IF1 corresponds to the conformation with negative mode amplitudes and IF2 with positive mode amplitudes. (c) Two-dimensional PCA space of the ground-truth synthetic conformations (black line: synthetic conformational transition trajectory), the initial conformation for the fitting (orange point), and the conformation used to generate the conformational variability (green point). (d) Accuracy of MDSPACE analysis of 500 particles measured at each MDSPACE iteration shown as box plots (black) and median (green). From left to right: errors between the ground-truth and estimated angles; errors between the ground-truth and estimated shifts; correlation coefficients between the images and the projections of the estimated (fitted) conformations; and RMSDs between the ground-truth and estimated (fitted) conformations. (e) Evolution of the principal component space over the iterations (from left to right: iteration 1 to 4); blue points represent 500 fitted conformations, each of which was obtained by fitting the initial conformation (orange color) to one of the images synthesized from 500 ground-truth conformations (black color).

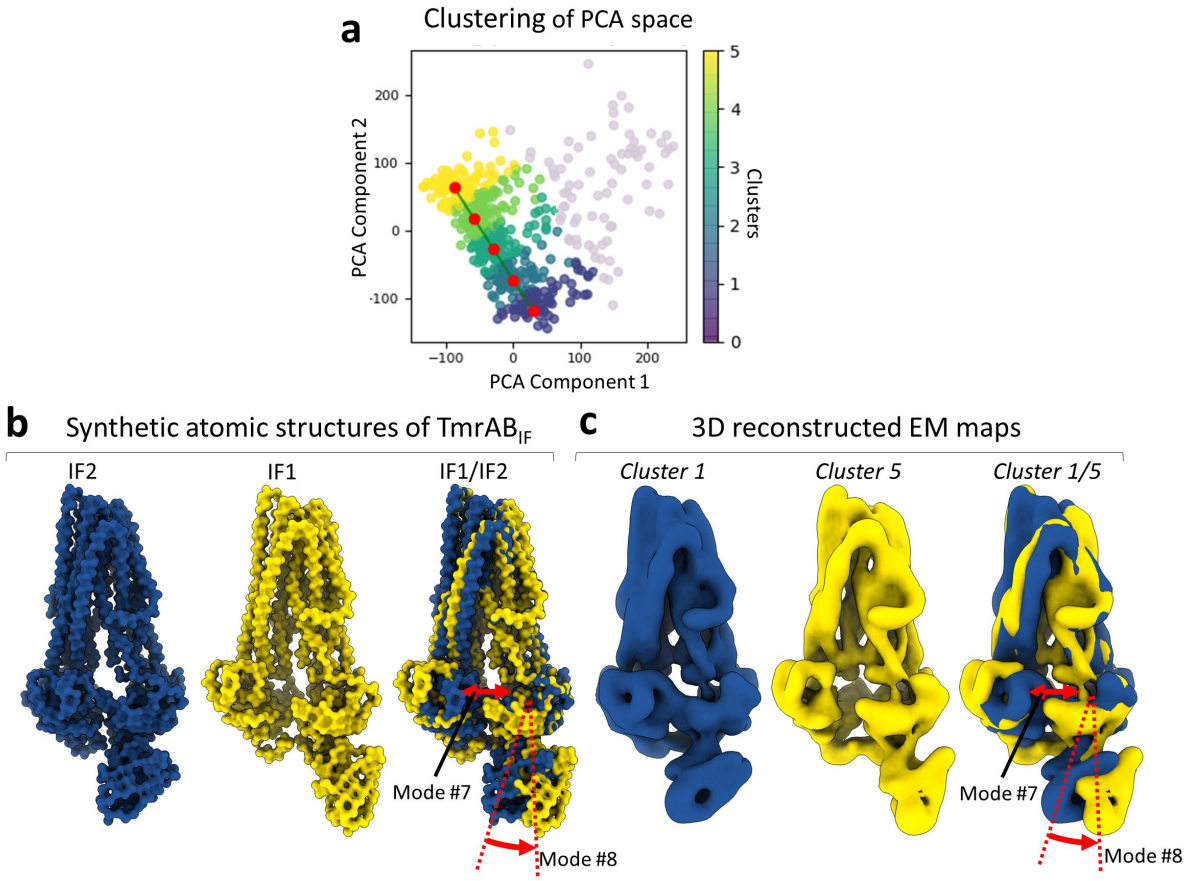

Figure 3 : Recovery of the ground-truth synthetic conformational transition trajectory of TmrAB based on PCA space clustering and 3D reconstructions from the clusters. (a) Five PCA space clusters obtained automatically along an interactively defined trajectory (red points) and colored from yellow to blue. (b) Two synthetic atomic conformations of TmrAB in inward-facing conformations (TmrAB$_{IF}$) corresponding to the two extremums of the ground-truth synthetic trajectory, denoted as IF1 and IF2. (c) 3D reconstructed EM maps from the yellow and blue clusters shown in (a). The conformational transition is visible on the superposed yellow and blue atomic structures (b) and EM maps (c). The color code is the same for the clusters in the PCA space (a), the atomic structures (b), and EM maps (c). See also **Supplementary Movies S1 and S2.**

3.2 Experiment with cryo-EM data from EMPIAR

In this subsection, we show the performance of MDSPACE using experimental cryo-EM data of yeast 80S ribosome-tRNA complexes available in EMPIAR database under the code EMPIAR-10016 [21].

EMPIAR-10016 dataset (yeast 80S ribosome-tRNA complexes)

During protein synthesis, tRNAs are translocated from the A (aminoacyl) to P (peptidyl) to E (exit) sites of the ribosome. During this process, two tRNAs adopt hybrid A/P and P/E states while progressing from the A-A and P-P states to the P-P and E-E states, which also involves a rotation between the two subunits of the ribosome [67-69].

In the original study of yeast 80S ribosome-tRNA complexes, resulting in the EMPIAR-10016 dataset publication [21], two cryo-EM maps were obtained using the FREALIGN likelihood-based image classification

[20, 70] into 5 classes from an initial set of 86,866 particle images, and deposited in the EMDB (accession codes: EMD-5976 and EMD-5977). The EMPIAR-10016 dataset contains a stack of particle images (image size: 360 × 360 pixels; pixel size: 1.05 Å × 1.05 Å) and 5 metadata files containing the orientation and translation parameters for 5 image classes. Two of the metadata files contain the parameters of 23,726 and 22,369 images, which correspond to the classes that yielded the two reconstructed cryo-EM maps (EMD-5976 at the resolution of 6.2 Å and EMD-5977 at the resolution of 6.3 Å, respectively). The EMD-5976 map corresponds to the rotated conformation (the ribosome inter-subunit rotation of around 9°) with one tRNA in a hybrid P/E state (designated as 80S-tRNA). The EMD-5977 map corresponds to the nonrotated conformation with two tRNA in the classical P-P and E-E states (designated as 80S-2tRNA). Two atomic models were derived from these cryo-EM maps (**Figure 4**) and deposited in the PDB (accession codes: PDB-3J77 and PDB-3J78) [21].

MDSPACE analysis of EMPIAR-10016 dataset

For the conformational variability analysis with MDSPACE, we used the particle images that yielded the EMD-5976 and EMD-5977 maps (all available 46,095 particle images), and we used their available orientation and translation parameters (determined by FREALIGN) as rigid-body pre-alignment parameters. We down-sampled the original particle images by a factor of two in each of the two dimensions, yielding images of size 180 × 180 pixels with a pixel size of 2.1 Å × 2.1 Å. The other 3 classes identified in the original publication [21] in the EMPIAR-10016 dataset appear less clean and less populated, and were therefore not used in our analysis with MDSPACE.

We performed 2 iterations of MDSPACE using 30-picosecond MD simulations, a force constant value of 10,000 kcal/mol, and the coarse-to-fine data processing scheme described in the **Methods** section. As the initial conformation, we used the atomic model of the 80S-tRNA conformation (the rotated conformation with one tRNA in a hybrid P/E state) (**Figure 4b**). In the first iteration, we analyzed 10,000 particles using normal-mode empowered MD-based flexible fitting (10 lowest-frequency normal modes were added to MD simulations as in NMMD). In the second iteration, we analyzed the entire set of 46,095 particles using principal-component empowered MD-based flexible fitting (PCA refinement) and the first 3 principal components (3 most dominant motions) obtained at the previous iteration.

The CC distribution shown in **Figure 5c** follows the same behavior as in the experiment with synthetic data. We observe a strong increase of the CC from the rigid-body pre-alignment to the first iteration (the median CC increases from 0.121 to 0.149), which is followed by a light increase at the next iteration (the median CC increases from 0.149 to 0.161 between the first and second iterations). This indicates that the conformations obtained at the first iteration get refined at the second iteration of MDSPACE.

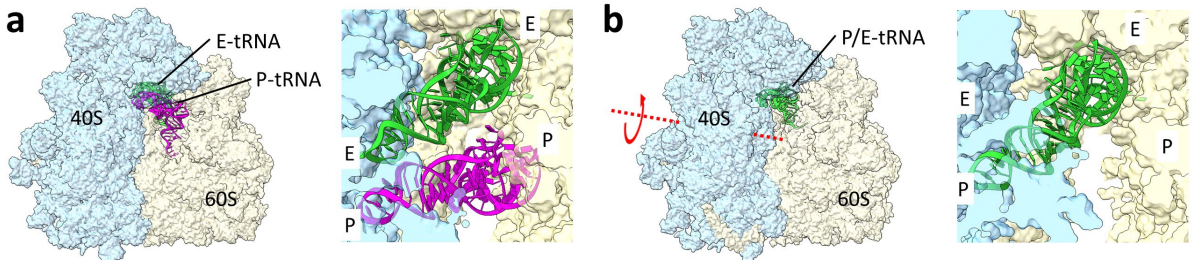

Figure 4 : Atomic models of yeast 80S ribosome-tRNA complexes derived from cryo-EM maps obtained with FREALIGN likelihood-based image classification of EMPIAR-10016 set of single particle images [21]. (a) Nonrotated conformation with two tRNAs in the classical E/E and P/P states (80S-2tRNA). (b) Rotated conformation with one tRNA in a hybrid P/E state (80S-tRNA). The 40S and 60S subunits are represented as light blue and light-yellow surfaces, whereas tRNAs are displayed as ribbons. The boxes at the right in (a) and (b) show close-up top views of the tRNAs.

MDSPACE recovers 80S-tRNA and 80S-2tRNA conformational states

The analysis of the conformational space obtained with MDSPACE shows a continuum of conformational states (**Figure 5a**). The singular values of the principal components (**Figure 5b**) decrease by 20% between the first and third components. To observe the extracted conformational variability in 3D, we performed clustering of the conformations in the PCA space, using 5 clusters linearly distributed along the first principal component (**Figure 5a**). **Figure 5d** shows the 3D reconstructions from the first and last clusters along the first principal component. These two clusters contain a relatively low number of particles (approximatively 3,000 particles per cluster) as they correspond to the extremums of the trajectory along the principal axis, which may explain a low resolution of the two reconstructed maps. Additionally, the maps in **Figure 5d** were low-pass filtered for the sake of reducing noise for visualization (low-pass cutoff frequency: 10 Å). We identified the first and last clusters (cluster 1 and 5) to correspond to the 80S-tRNA and 80S-2tRNA states, respectively, based on the presence of a single tRNA in a hybrid P/E state in the 3D reconstruction from cluster 1, two tRNA in the E/E and P/P states in the 3D reconstruction from cluster 5, and the inter-subunit rotation in the 3D reconstruction from cluster 1 (pre-translocation state) with respect to the 3D reconstruction from cluster 5 (post-translocation state). The remaining clusters (clusters 2 to 4) correspond to intermediate states of the inter-subunit rotation (**Supplementary Movie S3**).

MDSPACE reveals multiple conformational changes of the 80S ribosome

Additional conformational changes were observed by clustering along the second and third principal components (in 5 clusters by repeating the clustering procedure that was used for the first principal component). The conformational change along the second principal component shows a displacement of ribosomal protein P0 (**Figure 5e**), which with other P-stalk ribosome proteins plays a role in the interaction with translation initiation factors and promotion of translation initiation [71]. The conformational change along the third principal component shows a rotation of the head of the small subunit (**Figure 5f**), playing a role in facilitating the translocation process [72]. Other, but finer, conformational changes have been observed, as described in the next subsection.

MDSPACE reveals gradual transitions with many intermediate states at atomic scale

MDSPACE yielded the conformational space at atomic scale. In this subsection, we analyze this space in terms of atomic models along the first principal component. We selected a 9-point linear trajectory along this principal component (**Figure 6a**) and performed an inverse PCA for these points, which allowed us to visualize the atomic coordinate displacements along this 9-point trajectory (**Supplementary Movie S4**). The 9 obtained atomic models were compared with the 80S-tRNA and 80S-2tRNA models, which indicated a continuum of conformational states and agreement with the expected conformational changes (**Figure 6c-e**). **Figure 6c** shows the conformational change of the L1 stalk (25S rRNA single loop) which is known to interact with the E-site tRNA [21, 73, 74]. **Figure 6d** shows the conformational change of ribosomal protein uS12, located on the head of the small subunit, emphasizing the inter-subunit rotation. **Figure 6e** shows the conformational change of the section of 18S rRNA at the P-site of the small subunit. We can note that MDSPACE resolves the 80S-tRNA and 80S-2tRNA models, which approximately correspond to the 4th and 9th points of the trajectory, respectively (**Figure 6c-e**).

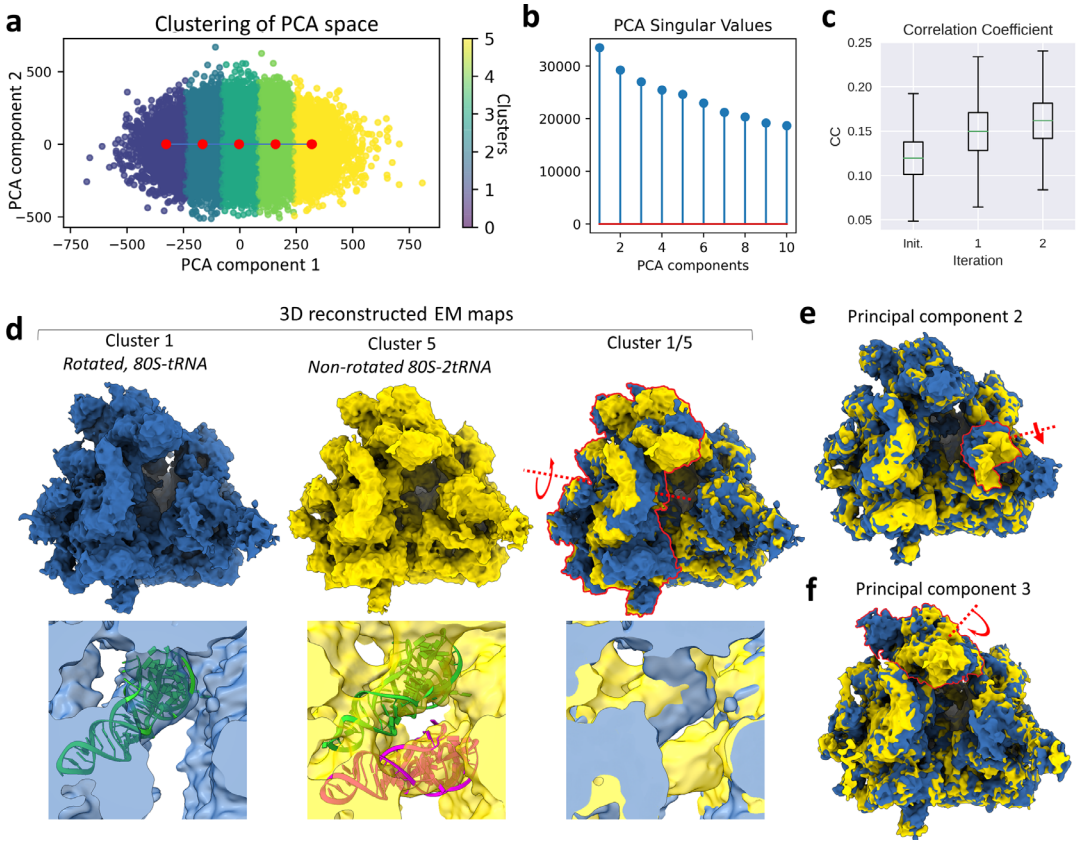

Figure 5: MDSPACE analysis of 80S ribosome-tRNA complexes from EMPIAR-10016 cryo-EM dataset [21]. (a) Conformational space obtained by MDSPACE, described by the first two PCA axes. Five PCA-space clusters colored from blue to yellow were automatically obtained along an interactively defined trajectory along the first PCA axis (red points). (b) Singular values of the PCA components. (c) Distribution of the correlation coefficient over the MDSPACE iterations. (d) 3D reconstructed EM maps from cluster 1 (blue points in (a)) and cluster 5 (yellow points in (a)). At the bottom, the first two panels from left to right show close-up top views of the E and P sites on the 3D reconstructed EM maps from cluster 1 and cluster 5 overlapped with the tRNAs (ribbon representation) of the 80S-tRNA model (rotated) and the 80S-2tRNA model (non-rotated), respectively. The remaining panel at the bottom shows a close-up top view of the overlapped EM maps from clusters 1 and 5. The red arrow shows the rotation of the 40S subunit. (e) 3D reconstructed EM maps from clusters 1 and 5 along the second PCA axis. (f) 3D reconstructed EM maps from clusters 1 and 5 along

the third PCA axis. The red arrows in (e) and (f) indicate the conformational changes. The clustering procedure used for the first PCA axis was repeated for the second (e) and third (f) PCA axes. See also **Supplementary Movie S3**.

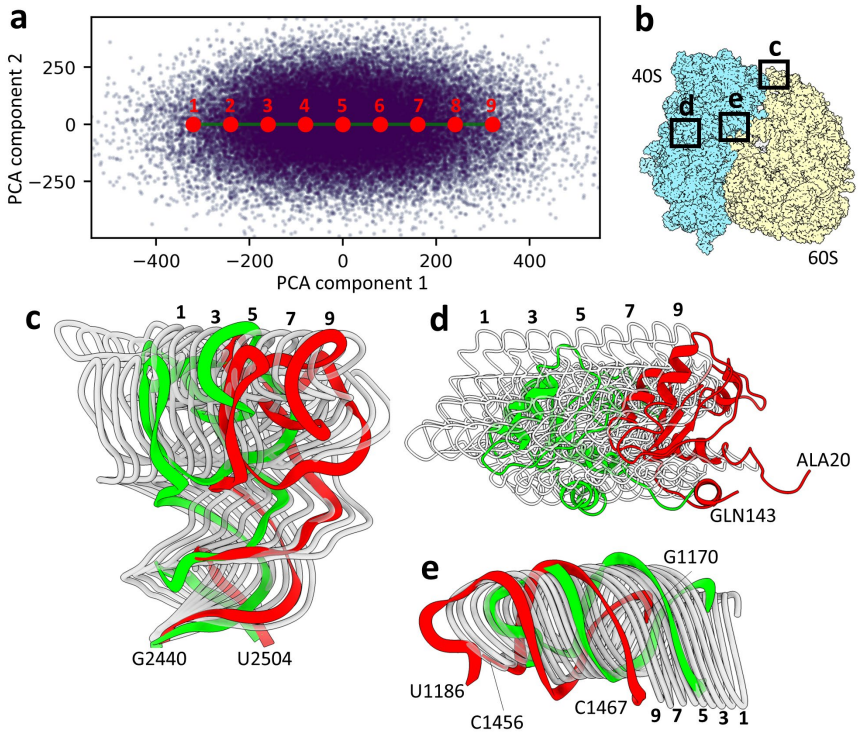

Figure 6 : Continuous conformational trajectory at atomic scale identified in the PCA space. (a) Conformational space obtained with MDSPACE (also shown in **Figure 5a**). The red points correspond to a 9-point linear atomic-coordinate trajectory along the first PCA axis, selected interactively in the PCA space. (b) Approximate locations of the 9 atomic models shown in (c)-(e) corresponding to the 9 red points in (a). (c) Motion of 25S rRNA single loop of the large-subunit L1 stalk (white) along the 9-point trajectory shown in (a). (d) Motion of ribosomal protein S12 (white), part of the head of the small subunit, along the 9-point trajectory shown in (a). (d) Motion of the section of the small-subunit 18S rRNA at the P-site (white) along the 9-point trajectory shown in (a). The models in (c)-(e) are superposed with the 80-tRNA and 80S-2tRNA models (green and red, respectively). See also **Supplementary Movie S4.**

4. Discussion and Conclusion

In this article, we presented a new approach for analyzing continuous conformational variability in cryo-EM single particle images, MDSPACE. MDSPACE uses, to the best of our knowledge, the first 3D-to-2D MD-based flexible fitting method and estimates the conformational landscape of the cryo-EM data at atomic scale, using a single input atomic model. The conformational landscape allows obtaining single atomic-scale conformational models, atomic-model animations, and 3D reconstructions. MDSPACE is a reference-based method, which implies that it relies on the structural composition of the reference to extract the conformational features from the images. For this reason, MDSPACE is generally not suitable for analyzing datasets with compositional heterogeneities, such as those caused by disassembly or cofactor binding of molecular complexes. However, in the 80S ribosome case studied in this article, we could capture binding of the second tRNA because it is associated with a conformational change of the 80S ribosome (inter-subunit rotation). In some cases of datasets, MDSPACE could help to identify a disassembly of large subunits of the complexes because the images containing single subunits may look like outliers in the resulting low-dimensional conformational space (under the condition that the entire complex is used as the reference and

that the fraction of the entire complexes is much larger than the fraction of the single subunits). However, if such strong compositional heterogeneity is suspected, we instead recommend using classical discrete-classification methods before using MDSPACE in order to filter out the compositional heterogeneity.

MDSPACE obtains an atomic-scale conformational model per particle

Methods development for conformational variability analysis based on mapping each cryo-EM single particle image onto a low-dimensional conformational space (an estimated essential conformational landscape) is currently an active field of research. Such methods belong to the class of continuous conformational variability methods. They cluster particle images with similar conformations in the low-dimensional conformational space determined beforehand, contrary to the methods that use discrete classification of the particle images into a number of classes decided prior to image analysis.

Most of the existing continuous conformational variability methods result in 3D reconstructions from the estimated low-dimensional conformational space [1, 11, 14]. Atomic models can then be derived from these 3D reconstructions. In contrast to these methods, our previously developed method HEMNMA additionally obtains the particle conformation in each particle image in terms of the model coordinates flexibly displaced along its normal modes so as to fit the particle image [2, 75]. These coordinates can be all-atom, coarse-grain, or artificial (the centers of 3D Gaussian functions that approximate a given EM map that is used as the initial conformation). Therefore, HEMNMA not only allows obtaining 3D reconstructions from clusters of particles in the low-dimensional conformational space, but also allows obtaining models of single atomic conformations (when a given atomic structure is used as the initial conformation) and molecular movies in this space based on these conformations. Conformational modeling using normal modes is fast (a fast displacement of the structure along its modes) but can suffer from the structural distortions for large amplitudes of conformational changes. Additionally, for speed reasons, the conformational modeling using normal modes is in practice limited to global motions, described by lowest-frequency normal modes.

In this article, we presented MDSPACE method for continuous conformational variability analysis that obtains one atomic conformation per particle image by flexibly fitting a given atomic model to images based on MD simulations, yielding good quality atomic models refined against images. MDSPACE refines the given atomic model against each image using normal-mode empowered MD simulations in the first iteration and principal-component empowered MD simulations in the remaining iterations. The normal modes speed up the MD-based fitting of the initial conformation to the target conformation in the particle image. The principal-components extracted from the ensemble conformational information at the end of each MDSPACE iteration refine the flexible fitting at the next MDSPACE iteration, especially for the images with difficult particle views and low SNR.

Additionally, MDSPACE allows obtaining atomic-resolution molecular movies and 3D reconstructions in the conformational space along its principal axes or along other directions ("trajectories") identified through the densest regions of this space, which may correspond to the most probable conformational transitions.

MDSPACE strategy to reduce the computation time

Each iteration of MDSPACE requires running one MD simulation per particle image. The MDSPACE code is MPI parallelized to process one particle image per CPU thread. MD simulations can be computationally expensive and their computational costs depends on multiple factors, such as simulation length, biomolecular size, and model type (all-atom or coarse-grained). Therefore, the large size of cryo-EM single particle datasets is a challenge for MDSPACE.

In this article, we tackle the speed issue of MD-based 3D-to-2D flexible fitting by using short-length simulations with coarse-grained models, within an original strategy combining (1) an integrated MD and normal mode simulation (previously used in the 3D-to-3D flexible fitting method called NMMD [52]); (2) an iterative refinement of the conformational space estimation via an iterative improvement of learning the principal axes of the conformational variability; and (3) coarse-to-fine data processing strategy (processing a small subset of images in the first iteration, followed by processing of an entire dataset in the next iteration).

With coarse-grained models, MDSPACE was able to retrieve the expected conformational changes while reducing the computational cost by several orders of magnitude, in both synthetic and experimental data cases. The computational cost reduction was especially important in the case of large experimental set of ribosome complexes. We have chosen Cα Gō model for our simulation to reduce the computational cost maximally. However, MDSPACE could be used with other coarse-grained force fields such as Gō-MARTINI [76]. Gō-MARTINI combines Go-like models with the MARTINI model and allows modeling more realistic environments and simulating larger amplitudes of structural dynamics (potentially unachievable with the Cα Gō model) while being less computationally expensive than all-atom simulations. For the moment, the only coarse-grained force field available in MDSPACE is the Cα Gō model. The use of MARTINI-like models will be implemented in the future. Nevertheless, when studying smaller systems than ribosomes, one could use all-atom simulations as they may results in more accurate conformational landscapes.

The computing time of MDSPACE could additionally be reduced by determining the adequate values of the MD simulation parameters that ensure the fastest convergence to the correct conformation, which may be challenging. The two most important parameters are the length of simulation and the force constant, whose values depend on other factors such as the molecular system or the chosen force field. A high value of the force constant speeds up the fitting as it helps the simulation to go faster through energy barriers, but the value of the force constant must be tuned with caution as too high values result in structural distortions. Concerning the length of the simulation, if the simulation is too short, the fitting gets closer but do not have enough time to converge. On the contrary, if the simulation is too long, the computational cost increases

without improving the results. Currently, there is no satisfying measure for assessing automatically the convergence. In the experiments described in this article, the force constant and the simulation length were determined based on preliminary experiments on a small number of images (around ten images). First, the optimal force constant was determined by running long simulations (up to 1 ns) with different values of the force constant and selecting the value that induces the fastest conformational changes (the fastest increase in the *CC*) without inducing structural distortions. Then, the optimal simulation length for the obtained optimal value of the force constant was determined as the time after which the *CC* does not increase anymore in all the simulations.

Even with a short simulation and a coarse-grained model, one simulation of 80S ribosome-tRNA complex takes one hour of CPU time on one core of Intel Xeon 6248 processor, meaning that one iteration of MDSPACE with the whole dataset (46,095 particles) would require about 46,000 CPU hours on one core of Intel Xeon 6248. Note here that the wall-clock time of the data processing presented in this article was much shorter as the processing was distributed over multiple CPU cores in parallel. The set of 46,095 particles was processed using 16 nodes with two Intel Xeon 6248 processors per node (20 CPU cores per processor), leading to approximatively 3 days of computation on 640 CPU cores used without multithreading (i.e., 1 thread was used on each core). To further reduce the computing time in the large experimental data case, we used only about one fourth (10,000 particles) of the entire dataset in the first iteration of MDSPACE. This subset of images ("training" set) was sufficient to estimate the global aspect of the conformational space, which was then refined in the next iteration using the entire dataset. This approach will be more investigated in the future, by further reducing the "training" data subset, as it may further increase the computational efficiency of MDSPACE.

Initial conformation and conformational space refinement

The conformational space depends on the chosen initial conformation. Different initial conformations will result in different conformational spaces in the sense that the same conformational state will be visible in different regions of the different conformational spaces. However, the principal motions that can be extracted from these different conformational spaces will be similar, unless the initial conformations are totally unrelated to the particles in the given dataset. If several potential initial conformations are available and all of them are coherent with the dataset, any of them can be used as the initial conformation. In case of doubts, the conformation that is the closest to the global average conformation (reconstructed from all the images) should be chosen as the initial conformation.

If the initial conformation is chosen correctly (it is coherent with the dataset, as explained in the previous paragraph) and the dataset analyzed with NMMD is large enough, only a fraction of the images will result in wrongly fitted conformations (those with more difficult views). The remaining set of images (with correctly

fitted conformations) is expected to be large enough to determine the initial conformational space (the principal axes) correctly enough, in order to drive the conformational space refinement in the next iteration.

Acknowledgment

We acknowledge the support of the French National Research Agency - ANR (ANR-19-CE11-0008 to SJ, BK, FT, and IR); 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); FOCUS for Establishing Supercomputing Center of Excellence (to FT); Nagoya University fund (to FT); JSPS (KAKENHI Grant No. JP20H05453 to OM); the Sorbonne University (2019 "Interface pour le Vivant" PhD scholarship grant to MH); and access to HPC resources of CINES and IDRIS granted by GENCI (A0100710998R, A0100710998, A0070710998, AP010712190, AD011012188 to SJ).

Credit author statement

Rémi Vuillemot: Conceptualization, Methodology, Software, Investigation, Validation, Writing- Original draft preparation. **Alex Mirzaei**: Conceptualization, Methodology, Software, Investigation, Validation, Writing - Review & Editing. **Mohamad Harastani**: Methodology, Software, Writing - Review & Editing. **Ilyes Hamitouche**: Validation. **Léo Fréchin**: Validation. **Bruno P. Klaholz**: Validation, Funding acquisition. **Osamu Miyashita**: Conceptualization, Methodology, Software, Validation, Writing - Review & Editing, Funding acquisition. **Florence Tama**: Conceptualization, Methodology, Validation, Writing - Review & Editing, Funding acquisition. **Isabelle Rouiller**: Methodology, Validation, Writing - Review & Editing, Funding acquisition. **Slavica Jonic**: Conceptualization, Methodology, Investigation, Validation, Writing - Review & Editing, Project administration, Funding acquisition.

Data availability statement

The original contributions presented in the study are included in the article/supplementary files. The software code of MDSPACE will be available on Github (https://github.com/scipion-em/scipion-em-continuousflex) and will also be part of the open-source ContinuousFlex plugin of Scipion V3. The synthetic data used in this study will be available on Zenodo (DOI: 10.5281/zenodo.7415104). All questions regarding the software or data availability can be addressed to the corresponding author.

Conflict of interest

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

References

[1] Dashti A, Schwander P, Langlois R, Fung R, Li W, Hosseinizadeh A, Liao HY, Pallesen J, Sharma G, Stupina VA, Simon AE, Dinman JD, Frank J, Ourmazd A. Trajectories of the ribosome as a Brownian nanomachine. Proc Natl Acad Sci U S A. 2014;111:17492-7.

[2] Jin Q, Sorzano COS, De La Rosa-Trevín JM, Bilbao-Castro JR, Núñez-Ramírez R, Llorca O, Tama F, Jonić S. Iterative elastic 3D-to-2D alignment method using normal modes for studying structural dynamics of large macromolecular complexes. Structure. 2014;22:496-506.

[3] Jonić S. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. Curr Opin Struct Biol. 2017;43:114-21.

[4] Sorzano COS, Jiménez A, Mota J, Vilas JL, Maluenda D, Martínez M, Ramírez-Aportela E, Majtner T, Segura J, Sánchez-García R, Rancel Y, Del Caño L, Conesa P, Melero R, Jonic S, Vargas J, Cazals F, Freyberg Z, Krieger J, Bahar I, Marabini R, Carazo JM. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy. Acta crystallographica Section F, Structural biology communications. 2019;75:19-32.

[5] Giraldo-Barreto J, Ortiz S, Thiede EH, Palacio-Rodriguez K, Carpenter B, Barnett AH, Cossio P. A Bayesian approach to extracting free-energy profiles from cryo-electron microscopy experiments. Scientific Reports. 2021;11:13657.

[6] Dashti A, Mashayekhi G, Shekhar M, Ben Hail D, Salah S, Schwander P, des Georges A, Singharoy A, Frank J, Ourmazd A. Retrieving functional pathways of biomolecules from single-particle snapshots. Nat Commun. 2020;11:4734.

[7] Hamitouche I, Jonic S. DeepHEMNMA: ResNet-based hybrid analysis of continuous conformational heterogeneity in cryo-EM single particle images. Front Mol Biosci. 2022;9:965645.

[8] Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, Urlaub H, Kastner B, Lührmann R, Stark H. Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. Cell. 2018;172:454-64.e11.

[9] Lederman RR, Andén J, Singer A. Hyper-molecules: on the representation and recovery of dynamical structures for applications in flexible macro-molecules in cryo-EM. Inverse Problems. 2020;36:044005.

[10] Moscovich A, Halevi A, Andén J, Singer A. Cryo-EM reconstruction of continuous heterogeneity by Laplacian spectral volumes. Inverse Problems. 2020;36:024003.

[11] Punjani A, Fleet DJ. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. Journal of Structural Biology. 2021;213:107702.

[12] Tagare HD, Kucukelbir A, Sigworth FJ, Wang H, Rao M. Directly reconstructing principal components of heterogeneous particles from cryo-EM images. Journal of structural biology. 2015;191:245-62.

[13] Chen M, Ludtke SJ. Deep learning-based mixed-dimensional Gaussian mixture model for characterizing variability in cryo-EM. Nature Methods. 2021;18:930-6.

[14] Zhong ED, Bepler T, Berger B, Davis JH. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. Nature Methods. 2021;18:176-85.

[15] Katsevich E, Katsevich A, Singer A. Covariance Matrix Estimation for the Cryo-EM Heterogeneity Problem. SIAM J Imaging Sci. 2015;8:126-85.

[16] 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. Frontiers in molecular biosciences. 2021;8:663121.

[17] 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. Journal of molecular biology. 2022;434:167381.

[18] Scheres SH, Gao H, Valle M, Herman GT, Eggermont PP, Frank J, Carazo JM. Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat Methods. 2007;4:27-9.

[19] Scheres SH. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012;180:519-30.

[20] Lyumkis D, Brilot AF, Theobald DL, Grigorieff N. Likelihood-based classification of cryo-EM images using FREALIGN. J Struct Biol. 2013;183:377-88.

[21] Svidritskiy E, Brilot AF, Koh CS, Grigorieff N, Korostelev AA. Structures of yeast 80S ribosome-tRNA complexes in the rotated and nonrotated conformations. Structure. 2014;22:1210-8.

[22] Zhou A, Rohou A, Schep DG, Bason JV, Montgomery MG, Walker JE, Grigorieff N, Rubinstein JL. Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM. Elife. 2015;4:e10180.

[23] Bai XC, Rajendra E, Yang G, Shi Y, Scheres SH. Sampling the conformational space of the catalytic subunit of human gamma-secretase. Elife. 2015;4.

[24] Abeyrathne PD, Koh CS, Grant T, Grigorieff N, Korostelev AA. Ensemble cryo-EM uncovers inchworm-like translocation of a viral IRES through the ribosome. Elife. 2016;5.

[25] Banerjee S, Bartesaghi A, Merk A, Rao P, Bulfer SL, Yan Y, Green N, Mroczkowski B, Neitz RJ, Wipf P, Falconieri V, Deshaies RJ, Milne JL, Huryn D, Arkin M, Subramaniam S. 2.3 A resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. Science. 2016;351:871-5.

[26] Hofmann S, Januliene D, Mehdipour AR, Thomas C, Stefan E, Brüchert S, Kuhn BT, Geertsma ER, Hummer G, Tampé R, Moeller A. Conformation space of a heterodimeric ABC exporter under turnover conditions. Nature. 2019;571:580-3.

[27] Nakane T, Kotecha A, Sente A, McMullan G, Masiulis S, Brown PMGE, Grigoras IT, Malinauskaite L, Malinauskas T, Miehling J, Uchański T, Yu L, Karia D, Pechnikova EV, de Jong E, Keizer J, Bischoff M, McCormack J, Tiemeijer P, Hardwick SW, Chirgadze DY, Murshudov G, Aricescu AR, Scheres SHW. Single-particle cryo-EM at atomic resolution. Nature. 2020;587:152-6.

[28] Kato K, Miyazaki N, Hamaguchi T, Nakajima Y, Akita F, Yonekura K, Shen J-R. High-resolution cryo-EM structure of photosystem II reveals damage from high-dose electron beams. Communications Biology. 2021;4:382.

[29] Klaholz BP, Myasnikov AG, Van Heel M. Visualization of release factor 3 on the ribosome during termination of protein synthesis. Nature. 2004;427:862-5.

[30] Simonetti A, Marzi S, Myasnikov AG, Fabbretti A, Yusupov M, Gualerzi CO, Klaholz BP. Structure of the 30S translation initiation complex. Nature. 2008;455:416-20.

[31] Klaholz BP. Structure sorting of multiple macromolecular states in heterogeneous cryo-EM samples by 3D multivariate statistical analysis. Open J Stat. 2015;5:820-36.

[32] Loerke J, Giesebrecht J, Spahn CM. Multiparticle cryo-EM of ribosomes. Methods Enzymol. 2010;483:161-77.

[33] Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, Rodnina MV, Stark H. Structure of the E. coli ribosome-EF-Tu complex at <3 A resolution by Cs-corrected cryo-EM. Nature. 2015;520:567-70.

[34] Penczek PA, Frank J, Spahn CM. A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. Journal of structural biology. 2006;154:184-94.

[35] Ma J. Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. Structure. 2005;13:373-80.

[36] Tama F, Brooks III CL. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu Rev Biophys Biomol Struct. 2006;35:115-33.

[37] Bahar I, Lezon TR, Yang L-W, Eyal E. Global Dynamics of Proteins: Bridging Between Structure and Function. Annual Review of Biophysics. 2010;39:23-42.

[38] López-Blanco JR, Chacón P. New generation of elastic network models. Current Opinion in Structural Biology. 2016;37:46-53.

[39] Grudinin S, Laine E, Hoffmann A. Predicting Protein Functional Motions: an Old Recipe with a New Twist. Biophys J. 2020;118:2513-25.

[40] Harastani M, Vuillemot R, Hamitouche I, Barati Moghadam N, Jonic S. ContinuousFlex: Software package for analyzing continuous conformational variability of macromolecules in cryo electron microscopy and tomography data. Journal of Structural Biology. 2022:107906.

[41] De la Rosa-Trevín J, Quintana A, Del Cano L, Zaldívar A, Foche I, Gutiérrez J, Gómez-Blanco J, Burguet-Castell J, Cuenca-Alba J, Abrishami V. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. Journal of structural biology. 2016;195:93-9.

[42] Zhong ED, Lerer A, Davis JH, Berger B. CryoDRGN2: Ab initio neural reconstruction of 3D protein structures from real cryo-EM images. 2021 IEEE/CVF International Conference on Computer Vision (ICCV)2021. p. 4046-55.

[43] Jonić S, Sorzano CÓS. Coarse-Graining of Volumes for Modeling of Structure and Dynamics in Electron Microscopy: Algorithm to Automatically Control Accuracy of Approximation. IEEE Journal of Selected Topics in Signal Processing. 2016;10:161-73.

[44] Miyashita O, Kobayashi C, Mori T, Sugita Y, Tama F. Flexible fitting to cryo-EM density map using ensemble molecular dynamics simulations. Journal of computational chemistry. 2017;38:1447-61.

[45] Orzechowski M, Tama F. Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. Biophysical journal. 2008;95:5692-705.

[46] Trabuco LG, Villa E, Schreiner E, Harrison CB, Schulten K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods. 2009;49:174-80.

[47] Wu X, Subramaniam S, Case DA, Wu KW, Brooks BR. Targeted conformational search with map-restrained self-guided Langevin dynamics: application to flexible fitting into electron microscopic density maps. Journal of structural biology. 2013;183:429-40.

[48] Igaev M, Kutzner C, Bock LV, Vaiana AC, Grubmüller H. Automated cryo-EM structure refinement using correlation-driven molecular dynamics. Elife. 2019;8:e43542.

[49] Habeck M. Bayesian Modeling of Biomolecular Assemblies with Cryo-EM Maps. Frontiers in Molecular Biosciences. 2017;4.

[50] Bonomi M, Pellarin R, Vendruscolo M. Simultaneous Determination of Protein Structure and Dynamics Using Cryo-Electron Microscopy. Biophys J. 2018;114:1604-13.

[51] Zhang J, Minary P, Levitt M. Multiscale natural moves refine macromolecules using single-particle electron microscopy projection images. Proc Natl Acad Sci U S A. 2012;109:9845-50.

[52] 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. Journal of Molecular Biology. 2022;434:167483.

[53] Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJ. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25:1422-3.

[54] Clementi C, Nymeyer H, Onuchic JN. 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. Journal of molecular biology. 2000;298:937-53.

[55] Takada S. Gō model revisited. Biophysics and physicobiology. 2019;16:248-55.

[56] Hills RD, Jr., Brooks CL, 3rd. Insights from coarse-grained Gō models for protein folding and dynamics. Int J Mol Sci. 2009;10:889-905.

[57] Noel JK, Levi M, Raghunathan M, Lammert H, Hayes RL, Onuchic JN, Whitford PC. SMOG 2: A Versatile Software Package for Generating Structure-Based Models. PLoS Comput Biol2016. p. e1004794.

[58] Kobayashi C, Jung J, Matsunaga Y, Mori T, Ando T, Tamura K, Kamiya M, Sugita Y. GENESIS 1.1: A hybrid-parallel molecular dynamics simulator with enhanced sampling algorithms on multiple computational platforms. Wiley Online Library; 2017.

[59] Suhre K, Sanejouand YH. ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res. 2004;32:W610-4.

[60] Tama F, Gadea FX, Marques O, Sanejouand Y-H. Building-block approach for determining low-frequency normal modes of macromolecules. Proteins: Structure, Function, and Bioinformatics. 2000;41:1-7.

[61] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825–30.

[62] Tipping ME, Bishop CM. Mixtures of Probabilistic Principal Component Analyzers. Neural Computation. 1999;11:443-82.

[63] Střelák D, Jiménez-Moreno A, Vilas JL, Ramírez-Aportela E, Sánchez-García R, Maluenda D, Vargas J, Herreros D, Fernández-Giménez E, de Isidro-Gómez FP, Horáček JM, Myška D, Horáček M, Conesa P, Fonseca-Reyna YC, Jiménez J, Martínez M, Harastani M, Jonić S, Filipovič J, Marabini R, Carazo JM, Sorzano COS. Advances in Xmipp for Cryo–Electron Microscopy: From Xmipp to Scipion. Molecules. 2021;26.

[64] Sorzano COS, Marabini R, Velázquez-Muriel J, Bilbao-Castro JR, Scheres SHW, Carazo JM, Pascual-Montano A. XMIPP: a new generation of an open-source image processing package for electron microscopy. Journal of Structural Biology. 2004;148:194-204.

[65] Peng L-M, Ren G, Dudarev S, Whelan M. Robust parameterization of elastic and absorptive electron atomic scattering factors. Acta Crystallographica Section A: Foundations of Crystallography. 1996;52:257-76.

[66] Sorzano CO, Jonic S, Nunez-Ramirez R, Boisset N, Carazo JM. Fast, robust, and accurate determination of transmission electron microscopy contrast transfer function. J Struct Biol. 2007;160:249-62.

[67] Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature. 2000;406:318-22.

[68] Spiegel PC, Ermolenko DN, Noller HF. Elongation factor G stabilizes the hybrid-state conformation of the 70S ribosome. Rna. 2007;13:1473-82.

[69] Ling C, Ermolenko DN. Structural insights into ribosome translocation. WIREs RNA. 2016;7:620-36.

[70] Grigorieff N. Frealign: an exploratory tool for single-particle cryo-EM. Methods in enzymology. 2016;579:191-226.

[71] Murakami R, Singh CR, Morris J, Tang L, Harmon I, Takasu A, Miyoshi T, Ito K, Asano K, Uchiumi T. The Interaction between the Ribosomal Stalk Proteins and Translation Initiation Factor 5B Promotes Translation Initiation. Mol Cell Biol. 2018;38.

[72] Ratje AH, Loerke J, Mikolajka A, Brünner M, Hildebrand PW, Starosta AL, Dönhöfer A, Connell SR, Fucini P, Mielke T, Whitford PC, Onuchic JN, Yu Y, Sanbonmatsu KY, Hartmann RK, Penczek PA, Wilson DN, Spahn

CMT. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature. 2010;468:713-6.

[73] Natchiar SK, Myasnikov AG, Kratzat H, Hazemann I, Klaholz BP. Visualization of chemical modifications in the human 80S ribosome structure. Nature. 2017;551:472-7.

[74] Khatter H, Myasnikov AG, Natchiar SK, Klaholz BP. Structure of the human 80S ribosome. Nature. 2015;520:640-5.

[75] Harastani M, Sorzano COS, Jonić S. Hybrid Electron Microscopy Normal Mode Analysis with Scipion. Protein Science. 2020;29:223-36.

[76] Poma AB, Cieplak M, Theodorakis PE. Combining the MARTINI and Structure-Based Coarse-Grained Approaches for the Molecular Dynamics Studies of Conformational Transitions in Proteins. Journal of Chemical Theory and Computation. 2017;13:1366-74.