



HAL
open science

AFMfit : Deciphering conformational dynamics in AFM data using fast nonlinear NMA and FFT-based search

Rémi Vuillemot, Jean-Luc Pellequer, Sergei Grudinin

► **To cite this version:**

Rémi Vuillemot, Jean-Luc Pellequer, Sergei Grudinin. AFMfit : Deciphering conformational dynamics in AFM data using fast nonlinear NMA and FFT-based search. 2024. hal-04801568

HAL Id: hal-04801568

<https://hal.science/hal-04801568v1>

Preprint submitted on 25 Nov 2024

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

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

AFMfit : Deciphering conformational dynamics in AFM data using fast nonlinear NMA and FFT-based search

Rémi Vuillemot¹, Jean-Luc Pellequer², Sergei Grudinin^{1*}

¹Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

²Univ. Grenoble Alpes, CEA, CNRS, IBS, 38000 Grenoble, France.

*Corresponding author(s). E-mail(s):

sergei.grudinin@univ-grenoble-alpes.fr;

Contributing authors: remi.vuillemot@univ-grenoble-alpes.fr;

jean-luc.pellequer@ibs.fr;

Abstract

Atomic Force Microscopy (AFM) offers a unique opportunity to study the conformational dynamics of proteins in near-physiological conditions at the single-molecule level. However, interpreting the two-dimensional molecular surfaces of multiple molecules measured in AFM experiments as three-dimensional conformational dynamics of a single molecule poses a significant challenge. Here, we present AFMfit, a flexible fitting procedure that deforms an input atomic model to match multiple AFM observations. The fitted models form a conformational ensemble that unambiguously describes the AFM experiment. Our method uses a new fast fitting algorithm based on the nonlinear Normal Mode Analysis (NMA) method NOLB to associate each molecule with its conformational state. AFMfit processes conformations of hundreds of AFM images of a single molecule in a few minutes on a single workstation, enabling analysis of larger datasets, including high-speed (HS)-AFM. We demonstrate the applications of our methods to synthetic and experimental AFM/HS-AFM data that include activated factor V and a membrane-embedded transient receptor potential channel TRPV3. AFMfit is an open-source Python package available at <https://gricad-gitlab.univ-grenoble-alpes.fr/GruLab/AFMfit/>.

Keywords: Atomic Force Microscopy, Conformational dynamics, Structural heterogeneity, Flexible fitting, Nonlinear normal modes,

Introduction

Observing biomolecules at work in their native environment is crucial for understanding their biological function and developing therapeutic applications. Only a handful of experimental methods (such as SAXS or solution NMR) can provide this information at the atomic scale. Among these, Atomic Force Microscopy (AFM) is a unique imaging technique that captures snapshots of biomolecules at nanometer resolution in near-physiological conditions with a signal-to-noise ratio (SNR) high enough to see individual molecules. In the AFM experiments, one scans the molecules with a sharp tip that measures the topography of their surface. The surface profiles are not resolved sufficiently to observe atomic details, but they inform on the domain-level arrangement of single molecules in solution. Moreover, one can collect AFM recordings successively with a frame rate inferior to one second in high-speed (HS)-AFM. This technique has been widely used to observe various biomolecules at work [1–3].

One of the main challenges in AFM is the interpretation of the resolution-limited 2D topographic surfaces of flexible biomolecules into understandable conformational dynamics in 3D. Although the direct reconstruction of a 3D contour of the molecule and its dynamics from the 2D surfaces is theoretically possible in case of sufficient observations, it has severe limitations. Firstly, the molecule in an AFM experiment needs to be attached to a flat surface of the stage to be imaged and tends to adopt the same orientation. Such a preferred viewing direction would cause anisotropy in the 3D reconstruction, a well-known problem in cryo-EM [4]. Secondly, the spatial resolution in AFM imaging is typically limited by the size of the tip to the nanometer scale [5]. Therefore, the AFM images are resolved only to the domain level, and atomic-level dynamics are not directly observable and would require prior structural knowledge.

Due to these limitations, computational methods to interpret AFM data in 3D typically use a model-based approach [6–16]. In such an approach, one or several available 3D structures at atomic resolution, such as those available in the Protein Data Bank (PDB), are docked under the AFM topographic images to estimate their position, orientation, and, possibly, their conformation. Most available computational methods focus on the rigid-body fitting of one or a few static candidate structures to AFM images [8, 10–13]. These candidate structures are, typically, stable conformations solved by X-ray crystallography or cryo-EM. These methods reconstruct the conformational variability using multiple molecular states that must be preliminary available to the user. This requirement has several limitations. Firstly, there is no guarantee that a sufficient number of states is available to describe the data. Secondly, due to crystallographic contacts, the states solved in crystals may not correspond to those observed in solution when probed by the AFM. Finally, the flexibility of biomolecules is typically better described by continuous motions rather than a few discrete states. One can partly avoid these limitations by enriching the number of candidate structures by sampling new states with molecular dynamics (MD) simulations [11, 12] at the expense of sufficiently long simulation trajectories that cover all the possible states encountered in the data. Another way to infer the flexibility of biomolecules is by fitting multiple rigid structures corresponding to the macromolecular subunits under study to the AFM images, allowing rigid-body motions between the subunits [7, 9].

The problem of reconstructing the continuous flexibility of biomolecules, as opposed to discrete states, from heterogeneous 2D views has been widely studied in cryo-EM and has undergone extensive developments in recent years [17–31]. In particular, continuous-state approaches emerged, which intend to treat each single molecule (designated as a particle in the cryo-EM field) as a different conformation and try to place each particle onto a continuous conformational space. This space typically contains 10^5 to 10^6 data points corresponding to each particle and gives an overall understanding of the conformational distribution in the sample. Some recent methods use model-based approaches that infer molecular flexibility using computational simulations from an initial atomic model, including MD simulations [21, 22, 29], the normal mode analysis (NMA) [17, 27, 28], and Zernike polynomials [25].

One can apply a similar approach to AFM. Recently, researchers have proposed two methods to infer the conformational state from a single-molecule AFM image using molecular simulations. One method uses a coarse-grained model represented as a mixture of Gaussians and reconstructs the structural heterogeneity with Monte Carlo simulations constrained by AFM data [14, 16]. The other method performs flexible fitting of an initial model using MD simulations biased with AFM-derived forces [15]. These methods have been applied to describe the conformation of one or a few particles (up to twelve [16]) but have not been tested on larger datasets.

However, AFM images often contain hundreds to thousands of single molecules, and this number may increase even further when recording multiple frames, such as in HS-AFM. Processing a larger number of images can result in a more accurate and comprehensive description of the conformational dynamics in the sample, in particular in the context of continuous motions of highly flexible biomolecules.

In the present work, we propose AFMfit, a new method for reconstructing conformational dynamics from AFM experiments. Our algorithm uses a novel flexible fitting procedure that scales to many single molecules in AFM images. AFMfit explores the flexible degrees of freedom with the nonlinear NMA method NOLB [32]. Our approach has the advantage of being very fast (e.g., a few tens of minutes on a personal desktop), requiring only one atomic model, and avoiding the structural distortions typically induced by linear normal modes at large amplitudes [32]. The method outputs a conformational landscape that describes the principal structural variations and their distribution over the data. We present our approach and demonstrate its applications to synthetic and experimental AFM and HS-AFM data. We show that the method accurately predicts the conformational dynamics given a set of noisy synthetic AFM images. Then, we apply the method to experimental an AFM dataset of activated coagulation factor V (FVA) [33] and an HS-AFM dataset of a transient receptor potential, TRPV3 [34]. The results highlight the conformational dynamics of the studied biomolecules and corroborate the previously published findings [33, 34].

Results

AFMfit’s rigid and flexible fitting

AFMfit aims to find the conformational ensemble of the sample molecule that best describes a set of AFM topographic images. More specifically, for each image in the

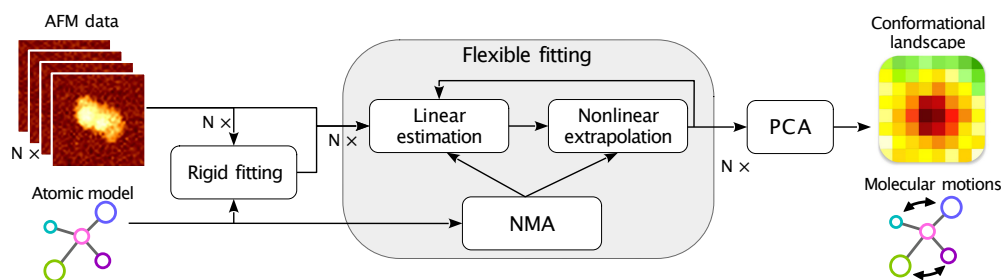


Fig. 1 Schematic workflow of the AFMfit fitting procedure. The method takes as input an available atomic model and a set of N AFM images of single molecules. The fitting consists of three main steps. In the first step, we perform a global rigid fitting of the atomic model to each AFM image. In the second step, we determine a set of continuous flexible and rigid deformations that improve the fit. The flexible deformations are obtained by performing NMA, and the resulting normal modes are fit in an iterative two-step procedure. Firstly, the optimal set of deformations along the linear normal modes is estimated. Secondly, a non-linear extrapolation of the deformations is carried out following the NOLB method [32] that preserves the model structure. In the last step, we assemble the flexibly fitted models into a low-dimensional representation of the conformational ensemble and extract molecular motions.

set, the algorithm searches for the best rigid (3D rotation and translation) and flexible (amplitudes of the normal mode deformations) alignments of a given initial atomic model that maximizes its similarity with the experimental data.

We conduct the rigid and flexible fitting procedures in the image space, meaning that we compute 2D AFM-like projections of the initial 3D model and compare them with the raw experimental observations to find the best match. In the first step of our procedure, we perform a global rigid fitting to estimate the global orientation of the molecule in each AFM image. Then, a flexible fitting algorithm based on nonlinear normal modes [32] locates the best flexible deformations of the initial model to describe the image, given the global rigid alignment found in the previous step. Note that the flexible fitting is also associated with local continuous rigid alignment, which allows for rigid rearrangement during the fitting. Finally, the deformations associated with each image form a conformational ensemble. We can further interpret this ensemble by projecting it onto a low-dimensional space of conformations using Principal Component Analysis (PCA). Figure 1 schematically shows the fitting procedure.

Evaluation of AFMfit on synthetic data of EF2

As a first step in evaluating our technique, we created a synthetic AFM data to test its performance in a controlled environment. We used a protein model of Elongation Factor 2 (EF2) that is available in the PDB (code 1N0V) [36]. This protein is suitable for testing our fitting pipeline because it has an asymmetric structure that consists of several globular domains, shown in Fig. 2a. When we deform these domains with NMA, they could adopt a cylinder-like shape with a pseudo-symmetry axis, which makes the fitting ambiguous at the AFM resolution. This ambiguity makes the rigid fitting challenging and requires the fitting pipeline to disentangle the conformation and the orientations to achieve a good fit. We created a conformational ensemble of 100

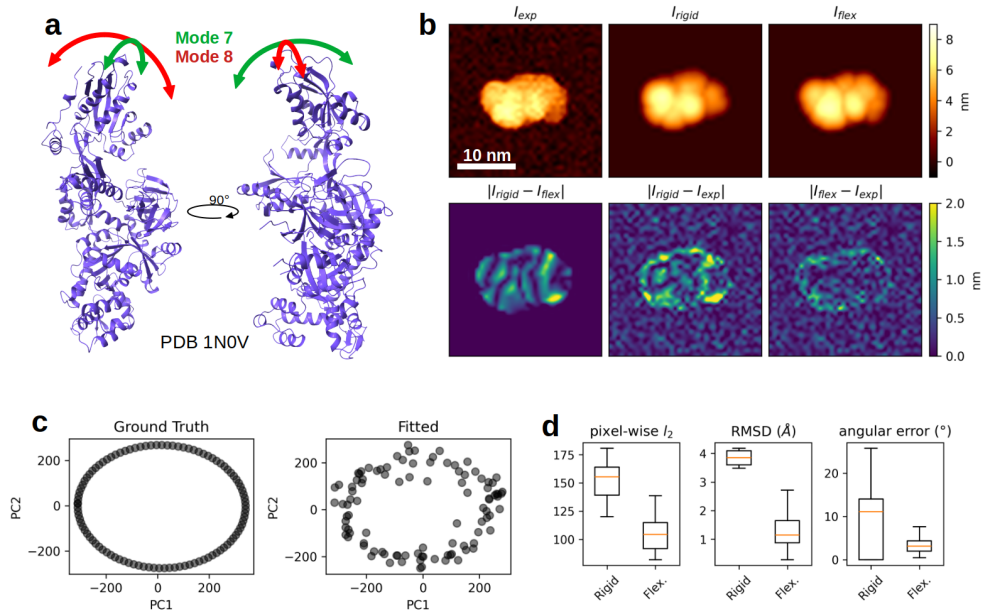


Fig. 2 AFMfit applied to synthetic data of EF2. **a.** 3D atomic structure of EF2 (PDB 1NOV), the arrows show the motion along the normal modes used to generate the synthetic dataset (modes 7 and 8). Molecular visualizations: ChimeraX [35]. **b.** Fitting results for a selected particle designated as I_{exp} . The best-fitted projections of the rigid and the flexible fitting are designated as I_{rigid} , and I_{flex} , respectively. The second row shows the absolute value of the difference (in nm) between I_{rigid} , I_{flex} and I_{exp} . **c.** Conformational space obtained by PCA on the ground truth (left) and the fitted (right) atomic models. **d.** Statistics of the flexible fitting compared to the rigid fitting, from left to right, the pixel-wise l_2 norm used to drive the fitting, the RMSD between the ground truth and the fitted atomic models, and the angular error between the ground truth and the estimated particle orientations.

conformations by applying flexible deformations to the EF2 model using the nonlinear normal mode analysis method NOLB [32]. We used the first two nonzero-frequency normal modes (number 7 and 8 shown in Fig. 2a) to generate these deformations. We selected the protein conformations along these deformations to be on average of 4 Å RMSD and describe a circle in the conformational space obtained through PCA on the conformational ensemble (as shown in Fig. 2c). We then applied 3D rotations to the conformational ensemble so that the models lie approximately flat on the xy -plane. We then generated pseudo-AFM images from the conformational ensemble of 40×40 pixels with a pixel size of 7 Å using the collision-detection method of the "afmize" software [37]. We set the simulated probe angle to 10° and the probe radius to 2 nm. We added Gaussian noise with a standard deviation of 3 nm to match the typical SNR of experimental AFM data. It is important to note that this simulation method is conceptually different from the one used in the flexible fitting pipeline to recover the flexible deformations. We used AFMfit with a $\sqrt{\lambda}$ value of 5, a σ value of 4.2 and included the five lowest-frequency normal modes in the fitting.

AFMfit was able to recover the ground truth conformational ensemble from the synthetic AFM data of EF2. Figure 2b shows an example of fitting for one selected particle in the set. The top left panel shows the particle, and the top middle and top right panels demonstrate the best match for rigid and flexible fits, respectively. If we examine this selected example, we observe that the rigid fitting alone misaligns the model with the image. Indeed, the best orientation obtained by the rigid fitting has an angular error close to 180° compared to the ground truth, meaning that the model flipped around the pseudo-symmetry axis. However, the flexible fitting corrected the misalignment and found the correct conformation, reducing the angular error to 3° (see the differences between rigid and flexible fitting, bottom row of Fig. 2b).

We further analyzed the fitted conformational ensemble (together with the ground-truth ensemble) with PCA. Figure 2c shows its first two components, with the left panel showing the samples from the ground truth and the right panel – the ones from the fitted ensemble. We can observe that both the fitted and the ground-truth ensembles closely match in terms of the amplitude of the deformations and the distribution of the conformations, as seen by the circle shape in the first two principal components.

Figure 2d provides a further comparison between the rigid and flexible fits. We found that the pixel-wise l_2 , used to drive the fitting, was significantly decreased between the rigid and flexible fitting procedures, with an average decrease of 34% from its initial value. Additionally, the RMSD between ground truth and fitted structures decreases from an average of 3.9 \AA to 1.1 \AA , and the angular error between the ground-truth orientation and the fitted orientation decreases from an average of 11.1° to below 3.1° . Overall, we can conclude that the presented method successfully recovered the correct conformational ensemble from a set of noisy synthetic AFM images of EF2.

Analysis of experimental AFM data of FVa

In the second stage of our method assessment, we applied the fitting pipeline to experimental AFM data of activated factor V (FVa). FVa is a protein involved in the blood coagulation cascade. In 2014, Chaves et al. obtained a high-resolution AFM measurement of isolated human FVa in a liquid environment containing multiple particles of FVa [33].

FVa is a heterodimer composed of a heavy chain with two A domains (A1-A2) and a light chain with one A domain and two C domains (A3-C1-C2). The C domains are known to be implicated in the binding of FVa with phospholipid membranes, which is crucial to its function, as shown by mutagenesis studies [38, 39]. The 2014 study revealed dynamic variations of the C1 and C2 domains [33]. At the time of the study, the structure of full human FVa was unknown. To assemble crystal structures of A trimers and two of C domains, the authors used a computational reconstruction pipeline known as AFM-assembly [7]. They reported conformational changes in the angle formed by C1 and C2 domains on a selected set of 20 particles [33].

More recently, Ruben et al. derived a complete structural model of human FVa (PDB code 7KXY) from a cryo-EM map of the Prothrombin-FVa-FXa complex at 4.1 \AA resolution [40]. We used this model as the initial state to apply our fitting algorithm to the AFM data of FVa. We performed NMA and selected the six zero-frequency and eight lowest-frequency (nonzero) normal modes, which mainly correspond to motions

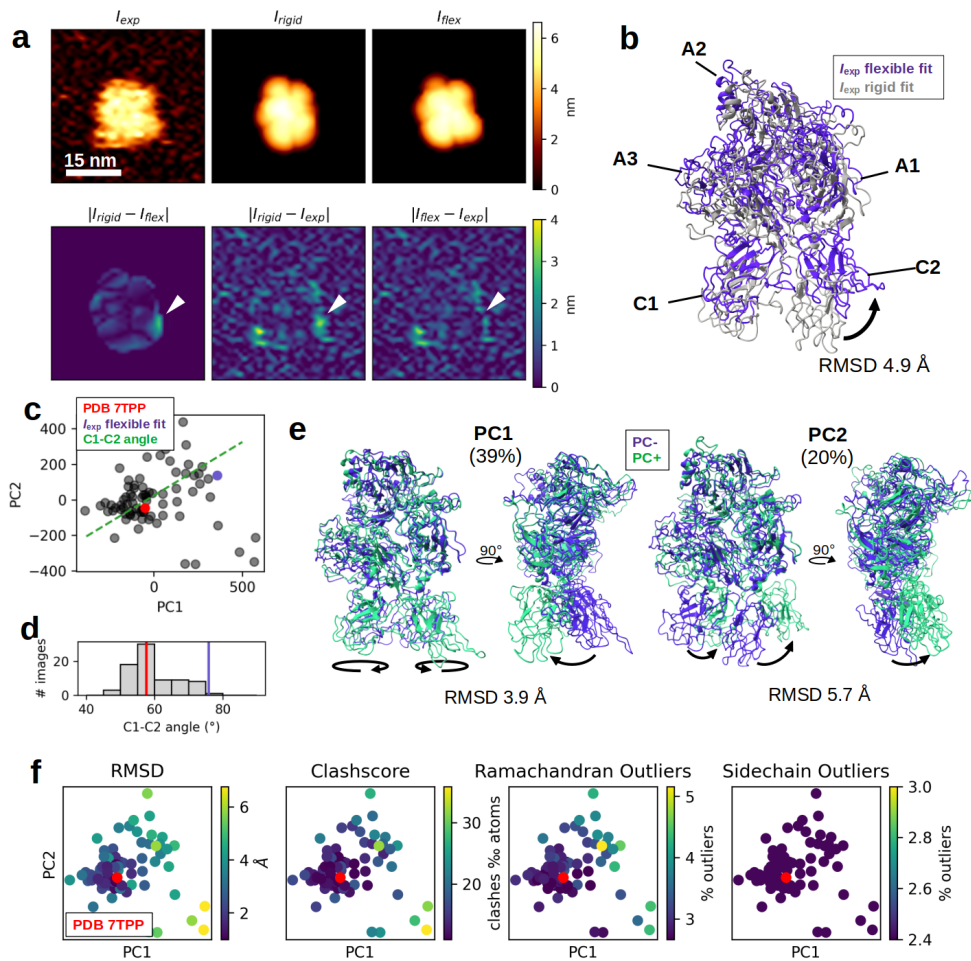


Fig. 3 AFMfit applied to the experimental AFM data of FVa. **a.** Fitting results for a selected AFM image designated as I_{exp} . The best-fitted projection of the rigid and the flexible fitting are designated as I_{rigid} and I_{flex} , respectively. The second row shows the absolute value of differences (in nm) between I_{rigid} , I_{flex} , and I_{exp} . The arrow heads show the location of C2. **b.** 3D structure of the best-fitted atomic models for I_{exp} (see panel (a.)) for the rigid fitting, PDB 7TPP (grey) and the flexible fitting (blue). **c.** Conformational space obtained by PCA on the set of flexibly-fitted atomic models. The first two principal components are shown, the red dot corresponding to the location of the initial model and the purple dot corresponding to the flexibly fitted model for the selected image I_{exp} . The green dashed line corresponds to the C1-C2 angle motion direction obtained by linear regression of the C1-C2 angle with respect to PC1 and PC2. **d.** Histogram of the C1-C2 angles of the flexibly-fitted atomic models. The red line corresponding to the C1-C2 angle of the initial model and the purple line corresponding to the fitted model for I_{exp} . **e.** The conformational changes associated with PC1 (39 % of explained variance) and PC2 (20 % of explained variance) are shown as a superposed atomic model associated with positive (purple) and negative (green) motions along the respective principal components. The arrows outline the conformational changes. **f.** Structural quality measures are shown on the first two principal components. From left to right, the RMSD between the initial and the fitted models, the clashscore corresponding to the number of steric clashes per 1000 atoms, the Ramachandran outliers corresponding to non-favorable dihedral angles, and the sidechain outliers corresponding to non-favorable rotamers. The red dots correspond to the location of the initial model.

of the C domains. We increased the number of extracted particles in the dataset from 20 to 77 by a manual selection and applied a masking operation composed of low-pass filtering and threshold to remove most of the background noise.

AFMfit recovers dynamics of the C domains of FVa

Overall, the fitting procedure decreased the loss function significantly by an average of 7.4 % between the rigid and the flexible fitting. Figure 3a-b shows a comparison between the rigid and flexible fitting results for a selected particle from the dataset. In this example, we observe that the C2 domain of FVa rotates (see red arrows in Figure 3a-b), decreasing the pixel-wise l_2 by 18.5 % (see bottom panels of Figure 3a).

We further processed the 77 fitted models with PCA to obtain the conformational space shown in Figure 3c. The first two principal components account for 60 % of the total variance (39 % for PC1 and 20 % for PC2). The conformational changes associated with PC1 and PC2 involve motions of the C domains, as shown in Fig. 3e. For PC1, the C domains are displaced collinearly to the pseudo-symmetry axis formed by the A trimer (Supplementary Video 1). The conformational change described by PC1 is also associated to a rotation of C1 and C2 that opens the angle between the two C domains. For PC2, the C domains rotate in the plane formed by the A trimer (Supplementary Video 2).

We note that the motion associated with a the combination of PC1 and PC2 reassembles the reconstructed models from the assembly protocol in [33]. More precisely, in the 2014 study, the authors reported a variation in the angle formed by the C1 and C2 domains. Here, the two most dominant reconstructed motion are associated with a variation of the angle between C1 and C2. Figure 3d shows the distribution of the C1-C2 angle over the fitted models. The distribution closely matches a linear combination of PC1 and PC2 (green dashed line in Figure 3c).

AFMfit conserves the structural quality of the initial model

To ensure that the produced conformational ensemble conserves the structural elements of the initial model, we calculated several standard metrics for the structural quality over the entire ensemble. We selected metrics similar to those used to assess cryo-EM-derived models in the PDB: the *clashscore* that measures the number of steric clashes (including hydrogens) per 1000 atoms, the *Ramachandran outliers* that count the percentage of disallowed dihedral angles and the *sidechain outliers*, and the percentage of disallowed rotamers. We computed all the metrics using the MolProbity software [41].

We shall note that the initial model of FVa used in this study was reconstructed from a medium-resolution (4.1 Å) EM map with some regions poorly resolved. Therefore, some of these metrics are initially of average quality, and the objective of using these metrics is not to give an absolute value of the quality of the model but rather to ensure that the metrics remain consistent and the quality does not deteriorate.

Figure 3d shows the value of these metrics, together with the RMSD between the initial and fitted models (first panel) projected on the two first principal components. For low to medium RMSD values (below 6 Å), corresponding to 93 % of the ensemble, the three metrics are close to the initial values, suggesting that the structural quality

of the initial model remains stable. For high RMSD values (above 6 Å), which account for solely 7 % of the ensemble, the structural quality decreases (clashscore above 30 %, Ramachandran outliers above 5 %). We shall expect this behavior, as using normal modes to approximate molecular deformations is valid only in the close vicinity of the equilibrium conformation (i.e., the initial model, in this case).

AFMfit recovers compositional and structural heterogeneity from HS-AFM movies of TRPV3

In the last stage of assessment of the method, we processed HS-AFM movies of a membrane-embedded transient receptor potential channel, TRPV3 [34]. We processed two movies of TRPV3 present in two oligomeric states, pentameric and tetrameric. The first HS-AFM movie contains multiple TRPV3 single molecules that show dynamic transitions between two oligomeric states. The second movie has a higher magnification and captures two single molecules at high resolution, one in the pentameric state and the other in the tetrameric state.

Apart from the apparent flexibility, the movies contain compositional heterogeneity (different oligomeric states). As AFMfit relies on the flexible deformation of an initial model, by default, it does not consider compositional heterogeneity. To tackle this issue, we used two initial models (TRPV3 pentamer, PDB code 8GKG; TRPV3 tetramer, PDB code 8GKA) and carried out the AFMfit protocol over all the particles with the two models. To assign the best model to each single molecule, we selected those that achieved the highest correlation on each particle.

We analyzed the first HS-AFM movie and identified up to 13 single molecules over 27 frames, with a maximum of 10 single molecules in one frame present simultaneously, resulting in 172 single-molecule frames. The AFMfit results, starting from the two oligomeric states (PDB codes 8GKG and 8GKA), allowed us to identify that 11 % of the molecules were pentamers. Supplementary Video 3 shows the result of the fitting for the entire movie and Figure 4a shows the result for one selected frame.

Apart from the compositional heterogeneity, we analyzed the flexibility of each single molecule in the first HS-AFM movie with PCA over the fitted models. Figure 4b shows the first three principal components of the 255 fitted models. We can see that the first component (80 % of explained variance) separates the two oligomeric states. The second, third, and fourth components (8.1 %, 4.4 % and 2.5 % of explained variance, respectively) are associated with flexible motions of the tetramers. The fifth component is associated with flexible motions of the pentamer (1.5 % of explained variance). The most dominant flexible motion for the tetramer (second principal component) is an inplane flattening (Figure 4b left and Supplementary Video 4) and for the pentamer (fifth principal component) is an asymmetric up and down motion in the z -axis (Figure 4b right and Supplementary Video 5). We should note that these flexible motions are relatively low-amplitude, especially for the pentamer (RMSD 4.3 Å and 2.9 Å between extremums, respectively).

We analyzed these motions further in the second HS-AFM movie at a higher magnification (pixel size of 1.25 Å). The movie contains 92 frames with a sampling rate of 0.3 seconds with two single molecules with stable oligomeric states (one pentamer and one tetramer). Supplementary Video 6 shows the result of the fitting for this movie.

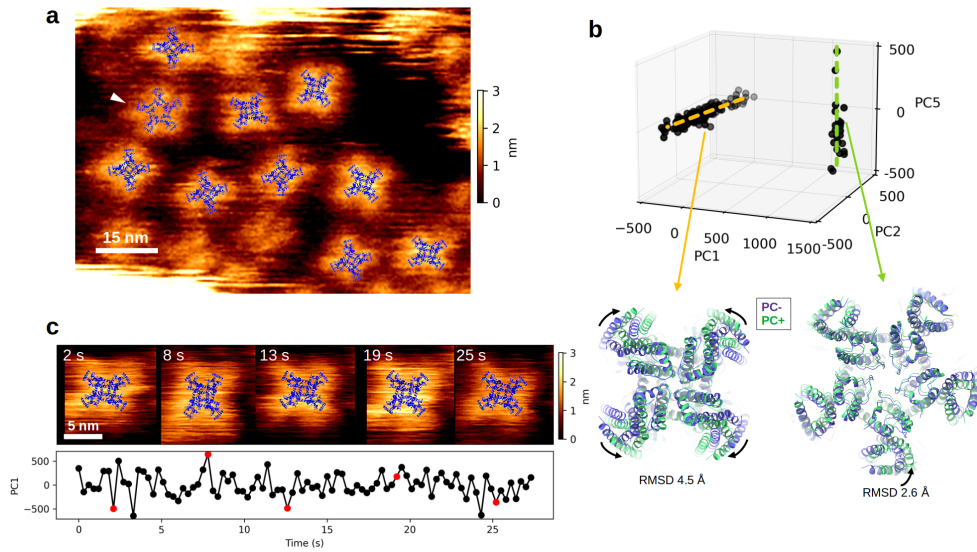


Fig. 4 AFMfit applied to HS-AFM data of TRPV3. **a.** An example of a fitted HS-AFM frame with several pentameric states of TRPV3 (arrowheads). The models fitted by AFMfit are superimposed with the HS-AFM frame for each single molecule. **b.** The first three components of the PCA decomposition from the 255 models fitted by AFMfit on the HS-AFM movie are shown in (a.). The first components separate pentameric and tetrameric states. The second component is mainly associated with motions in the pentameric state (green dashed line); the corresponding structural changes are shown in the right-most structure. The third component is mainly associated with motions in the tetrameric state (yellow dashed line); the corresponding structural changes are shown in the left-most structure. **c.** Flexible motions observed in high-magnification HS-AFM movie of TRPV3 in the tetrameric state. The top row shows the relevant HS-AFM frame superimposed with the corresponding models fitted with AFMfit. The bottom row shows the time series of the state along the first principal component (red dots correspond to the frames shown in the top row).

The flexible motions extracted from these two single molecules are the same as the flexible motions described for the first movie (Figure 4b), i.e., small-amplitude inplane flattening for the tetramer and up and down motions of the monomers in the z-axis for the pentamer. Figure 4c shows the fitting results for the tetramer TRPV3 of the second movie over time. We selected relevant states at different locations along the main flexible component (PC1). A positive displacement along PC1 corresponds to an inplane flattening of the TRPV3 tetramer in the x-axis direction, and a negative displacement corresponds to the same motion in the y-axis direction. Most interestingly, there is no apparent time coordination between these small-amplitude flexible motions. Overall, we conclude that the AFMfit protocol separated the different oligomeric states in the HS-AFM movies of TRPV3 and assigned small-amplitude flexible deformations to each state.

Table 1 Execution time of AFMfit for the presented AFM/HS-AFM datasets on a single workstation Intel(R) Core(TM) i7-13700 CPU @ 5.2 GHz

Dataset	# image	Pixel size (\AA)	Time (min)
EF2	100	7.0	15
FVa	77	9.75	19
TRPV3 (1 st)	172	4.0	9
TRPV3 (2 nd)	184	1.25	23

Discussion

One advantage of AFMfit is its speed, which allows the processing of hundreds to thousands of AFM images in just a few minutes on a personal desktop. Table 1 reports the computational time to run the AFMfit experiments presented in this work.

To approximate low-dimensional motion manifolds of the input models, we employed NMA. We must notice, however, that using NMA for simulating the flexibility of biomolecules may have limitations. For example, one shortcoming of linear normal modes is that at high deformation amplitudes, the interatomic distances are no longer respected, and some parts of the molecular structure tend to get distorted. In our model, we have utilized a nonlinear NMA approach, NOLB, that brings a solution to this issue [32]. Another limitation of NMA is that its energy model predetermines the motions, and it can be challenging to pick the relevant ones. The elastic network model inside NOLB and many other NMA approaches accelerates the computations but may restrict the accuracy of the normal modes to the lowest frequencies. In our case, selecting a few low-frequency normal modes is suitable as the resolution of the AFM images is limited, and using more degrees of freedom would be prone to overfitting. Therefore, we recommend using less than ten lowest-frequency normal modes, and selecting the lowest ones would give acceptable results for most models. We shall note that we did not employ any restriction on the elastic network model regarding the interaction between the molecule and the surface to which it is attached. Such interactions can be included in a future analysis, as they may influence the flexibility of the structural model.

Conclusion

To conclude, we propose a new fitting method to interpret conformational dynamics from AFM/HS-AFM using an input atomic model. To the best of our knowledge, it is the first approach that applies NMA-based fitting to the analysis of AFM images. We confirmed its ability to recover a given conformational space from synthetic data. We also tested the method on experimental datasets of AFM and HS-AFM and demonstrated that our results corroborate with previously published results [33, 34].

Methods

Simulating pseudo-AFM images

AFMfit compares experimental AFM images with pseudo-AFM images synthetically generated from the atomic models in different conformations and orientations. To simulate a pseudo-AFM image from a 3D atomistic model, a conventional approach consists of calculating the height in the z -direction at which a tip perpendicular to the xy -plane collides with the molecular structure, the so-called collision-detection method [10]. However, this method is not differentiable, which is a limitation for our flexible fitting purposes. Instead, we use a smooth differentiable function proposed in [15] that demonstrated a fair approximation to the collision-detection method.

The smooth function combines a set of Gaussian functions in the xy -plane that define the distance between each atom and the pixel center and a soft-maximum log-sum-exp function in the z -direction. Given an atomic model X composed of N atoms $X_i = \{x_i, y_i, z_i\}, i = 1, N$, the smooth function $\Pi : \mathbb{R}^{3N} \rightarrow \mathbb{R}^{N_P \times N_P}$ for the p th pixel $p = \{p_x, p_y\}$ is given by :

$$\Pi_p(X) = \gamma \log \left(1 + \sum_{i=1}^N \exp \left(-\frac{(x_i - p_x)^2 + (y_i - p_y)^2}{2\sigma^2} + \frac{z_i + r_i}{\gamma} \right) \right), \quad (1)$$

where r_i is the radius of the i th atom, γ is a parameter that makes the function $\Pi_p(X)$ smooth and σ is the spatial range (resolution) of the function. In our experiments, we use a fixed value of the resolution parameter $\gamma = 1 \text{ \AA}$. For this value of γ , the authors in [15] determined the optimal value of σ that best approximates the collision-detection method. This value depends on two parameters, the radius and the angle of the tip, and Table 1 in [15] describes the optimal σ given a range of values of these two parameters. We shall note that it can be challenging to accurately determine these parameters in the experimental setup. Therefore, this table provides a first estimate of σ , which can be adjusted later by running multiple experiments around this value and choosing the best fit. For the experiments, we used a value of σ of 3.2, 5.0 and 4.0 for the dataset of EF2, FVa and TRPV3, respectively.

Global rigid fitting

The objective of the rigid-body fitting in AFMfit is to find the optimal rigid alignment (3D rotation and translation) of an atomic model X and the experimental AFM images preliminary to the flexible fitting. We will then refine the rotation and the translation obtained at this stage by the flexible fitting described below.

The algorithm exhaustively searches for the best rigid-body transformed model $\mathcal{X}(\theta, \tau)$ rotated by the 3D rotation described by the three Euler angles $\theta \in \mathbb{R}^3$ and translated by three translation parameters $\tau \in \mathbb{R}^3$ by minimizing the squared pixel-wise l_2 norm between the experimental AFM image Y and a pseudo-AFM image generated from $\mathcal{X}(\theta, \tau)$ with the function described in Eq. 1 :

$$L(\theta, \tau) = \|Y - \Pi(\mathcal{X}(\theta, \tau))\|_2^2. \quad (2)$$

We use the pixel-wise l_2 norm as the similarity measure (similarly as for the flexible fitting in Eq. 12). We have chosen the pixel-wise l_2 over other measures (e.g., correlation coefficient widely used in cryo-EM), notably because it changes by uniform scaling $\Pi(X) \rightarrow s\Pi(X)$ and uniform shifting $\Pi(X) \rightarrow \Pi(X) + s$. In comparison, the correlation coefficient is invariant to uniform scaling and uniform shifting. The changes to uniform shifting may help place the atomic model at the correct height in the topographic AFM image, and the changes to uniform scaling ensure the conservation of the geometric distances between molecular structures at different heights.

The algorithm exhaustively generates a pseudo-AFM image for discretized values of the three rotation parameters θ and one translation parameter τ_z (the translation in the z -direction). We accelerate the calculation of the last two translation parameters, translations in the xy -plane, via the fast Fourier transform calculation of L described later in this section.

To evenly discretize the space of 3D rotations given an angular increment $\Delta\theta$, we start by evenly distributing the viewing directions along the 3D sphere with a spherical Fibonacci lattice, which samples the two first Euler angles. Then, the remaining Euler angle describing an in-plane rotation (rotation about the z -axis) is discretized linearly with a $\Delta\theta$ step. We linearly discretize the translation along the z -axis with an increment $\Delta\tau_z$.

Finally, we obtain the last two translation parameters τ_x and τ_y (translation in the xy -plane) by an FFT-accelerated calculation of L . Concretely, we can rewrite L as

$$L(\theta, \tau) = \|Y\|_2^2 + \|\Pi(\mathcal{X}(\theta, \tau))\|_2^2 - 2C(\theta, \tau), \quad (3)$$

where

$$C(\theta, \tau) = \sum_{x=1}^{N_P} \sum_{y=1}^{N_P} Y(x, y) \Pi(x + \tau_x, y + \tau_y; \mathcal{X}(\theta, \tau_z)). \quad (4)$$

Note that we apply the translation in the xy -plane in Eq. 4 in the image space rather than the 3D Cartesian space, therefore $C(\theta, \tau)$ is also the 2D cross-correlation between Y and $\Pi(\mathcal{X}(\theta, \tau_z))$. For a given value of θ and τ_z , considering that $\|Y\|_2^2$ is constant and $\|\Pi(\mathcal{X}(\theta, \tau))\|_2^2$ is invariant to translation in the xy -plane, the values of τ_x, τ_y that maximize C also minimize L and can be calculated for all values of τ_x and τ_y in the discrete image space $\{1, \dots, N_P\}$ via the fast Fourier transform :

$$C(\tau_x, \tau_y; \theta, \tau) = Y * \Pi(\mathcal{X}(\theta, \tau_z)) = \mathcal{F}^{-1} \left(\overline{\mathcal{F}(Y)} \mathcal{F}(\Pi(\mathcal{X}(\theta, \tau_z))) \right), \quad (5)$$

where the $*$ sign denotes the 2D cross-correlation, \mathcal{F} is the 2D Fourier transform, \mathcal{F}^{-1} is the inverse 2D Fourier transform, and $\overline{\mathcal{F}(\cdot)}$ is its complex conjugate.

Normal Mode Analysis and NOLB

The core behind the flexible fitting part of our method is a low-rank representation of molecular flexibility computed by the NOLB method [32]. In a nutshell, NOLB

represents the molecular system using an all-atom elastic network model with an anisotropic potential function of the following form,

$$V = \sum_{d_{ij}^0 < R_c} \frac{\gamma}{2} (d_{ij} - d_{ij}^0)^2, \quad (6)$$

where d_{ij} is the distance between the i^{th} and the j^{th} atoms, d_{ij}^0 is the reference distance between these atoms, as found in the original model, γ is the stiffness constant, and R_c is a cutoff distance, which we fix to 8 Å by default. Then, using this potential function, it constructs a mass-weighted Hessian matrix and projects it into the space where each residue is represented as a rigid body. After, the method computes several lowest-frequency eigenvectors in the reduced rigid-residue space. We fix this number to a maximum of ten, and one can select the most relevant vectors manually, e.g., by visualizing each of them to get a finer description of the flexible motions. Note that using a higher number of normal modes is prone to overfitting. Each eigenvector has six components per residue: three correspond to the instantaneous linear velocities \vec{v} and the other three to the instantaneous angular velocities $\vec{\omega}$. At the same time, we also compute an all-atom linearization A of the normal modes projecting the eigenvector back to the all-atom space. Finally, following the NOLB methodology, we extrapolate coordinates $X(\alpha)$ of each atom in a residue at an amplitude α as

$$X(\alpha) = R(\alpha \|\vec{\omega}\|_2, \vec{\omega} / \|\vec{\omega}\|_2) (X(0) - \vec{r}) + \vec{r} + \alpha \vec{\omega} (\vec{v} \cdot \vec{\omega}) / \|\vec{\omega}\|_2^2, \quad (7)$$

where we parameterize a rigid rotation R using an angle-axis notation. In this equation, the center of rotation \vec{r} for a residue with the center of mass \vec{c} is given by

$$\vec{r} = \vec{c} + (\vec{\omega} \times \vec{v}) / \|\vec{\omega}\|_2^2. \quad (8)$$

Flexible fitting of an AFM image

Our flexible fitting procedure operates in two steps that are iterated until the algorithm convergence. In the first step, we estimate the optimal change in the linear normal mode amplitudes $\Delta\alpha \in \mathbb{R}^M$ using M lowest frequency normal modes, such that

$$X^n = X^0 + (\alpha^{n-1} + \Delta\alpha)A, \quad (9)$$

where X^n is the atomic model estimated at the n -th iteration of the algorithm, $\alpha^{n-1} \in \mathbb{R}^M$ are the normal mode amplitudes at the previous iteration, X^0 is the known initial atomic model, and $A \in \mathbb{R}^{M \times 3N}$ is the set of M low-frequency normal modes associated to X^0 . In the second step, we update the normal mode amplitudes $\alpha^n = \Delta\alpha + \alpha^{n-1}$ and extrapolate the atomic model \tilde{X}^n with the nonlinear NOLB procedure explained above. We then use the obtained model \tilde{X}^n to calculate the change in normal mode amplitudes in the first step of the next iteration. Then, we repeat the two steps until convergence. Typically, the algorithm converges in less than ten iterations. Note that

during the process, the normal modes A do not need to be updated and can remain constant, which reduces the computational cost.

We shall note that the set of M low-frequency normal modes includes the six modes A_i with $i = \{1, \dots, 6\}$ with *zero* frequency corresponding to rigid-body translations and rotations. We include these six additional degrees of freedom in the fitting procedure to allow for continuous rigid-body rearrangement during the flexible deformations. These rigid rearrangements are crucial for the accuracy of our method, as the global rigid fitting might not be valid when the conformation of the model is changed. We compute the *zero* frequency modes analytically according to [42]. The three modes associated with a rigid translation are straightforward to express as a displacement in the Cartesian space. The three other modes correspond to three infinitesimal rotations about the x, y , and z axes. We can compute an infinitesimal rotation $A_R(u, \theta)$ about a unit vector u by an angle θ , for small enough θ , as a linear approximation according to

$$A_R = [X^0]_{\times} \theta u, \quad (10)$$

where X^0 is the initial model used for the NMA computations, and $[\cdot]_{\times}$ denotes the cross-product matrix,

$$[v]_{\times} = \begin{pmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{pmatrix}. \quad (11)$$

We determine the unknown normal mode amplitude $\Delta\alpha$ at each iteration by minimizing the loss function L defined as

$$L = \|\Pi(\tilde{X}^n) - Y\|_2^2 + \lambda \text{RMSD}_i^2, \quad (12)$$

where Y is the experimental AFM image, Π is the smooth pseudo-AFM image function defined in Eq. 1, $\|\cdot\|_2$ denotes the pixel-wise l_2 norm, RMSD_i defines the *root-mean-square-deviation* between the initial model X^0 and the fitted model \tilde{X}^n , and λ is a parameter that controls the amplitude of the flexible deformation. The leftmost term in the loss function describes the similarity between the fitted atomic model and the experimental image. The right-most term is a regularization, which keeps the deformation amplitude of the initial model finite. Parameter λ controls the balance between data fitting and deformation amplitude and needs to be tuned carefully. Small λ values would attenuate constraints on the deformation amplitude and might result in model overfitting. At large λ values, the fitting would minimize the deformation amplitude and not deform the model sufficiently to achieve a good fit. We recommend scanning λ values between $\sqrt{\lambda} = 5$ and $\sqrt{\lambda} = 20$ over multiple images (e.g., on a small subset of images) to find an optimal value. We used values of $\sqrt{\lambda}$ of 5, 20 and 10 for the experiments with EF2, FVa and TRPV3, respectively. To evaluate if the models are overfitted, one can use software such as Molprobit [41] that assesses their structural quality.

To obtain an analytical solution for $\Delta\alpha$, we perform the following linear approximation,

$$\Pi(\tilde{X}^n) \approx \Pi(\tilde{X}^{n-1}) + (\tilde{X}^n - \tilde{X}^{n-1}) \nabla \Pi(\tilde{X}^{n-1}), \quad (13)$$

where $\nabla\Pi(X)$ is the gradient of the function $\Pi(X)$ such as :

$$\nabla\Pi_p(X) = \begin{pmatrix} \frac{\partial\Pi_p(X)}{\partial X_1} \\ \vdots \\ \frac{\partial\Pi_p(X)}{\partial X_N} \end{pmatrix}, \quad (14)$$

$$\frac{\partial\Pi_p(X)}{\partial X_i} = \begin{pmatrix} -\frac{\gamma}{\sigma^2}(x_i - p_x) \\ -\frac{\gamma}{\sigma^2}(y_i - p_y) \\ 1 \end{pmatrix} \exp\left(-\frac{(x_i - p_x)^2 + (y_i - p_y)^2}{2\sigma^2} + \frac{z_i + r_i}{\gamma} - \frac{1}{\gamma}\Pi_p(X)\right). \quad (15)$$

We also consider that the nonlinear extrapolation between each step is close to the linear transition :

$$\tilde{X}^n - \tilde{X}^{n-1} \approx \Delta\alpha A \quad (16)$$

By using Eq. 16 and substituting Eqs. 9 and 13 in our loss function in Eq. 12, we obtain :

$$L = \|\Pi(\tilde{X}^{n-1}) + \Delta\alpha A \nabla\Pi(\tilde{X}^{n-1}) - Y\|_2^2 + \frac{\lambda}{N} \|(\alpha_{n-1} + \Delta\alpha)A\|_2^2. \quad (17)$$

Finally, by equating partial derivatives of the Loss function with respect to the unknown amplitudes $\Delta\alpha$ to zero,

$$0 = \frac{\partial L}{\partial \alpha} \equiv 2(\Pi(\tilde{X}^{n-1}) + \Delta\alpha A \nabla\Pi(\tilde{X}^{n-1}) - Y)(A \nabla\Pi(\tilde{X}^{n-1}))^T + \frac{2\lambda}{N}(\Delta\alpha + \alpha^{n-1})AA^T, \quad (18)$$

we obtain the final result:

$$\Delta\alpha = -((\Pi(\tilde{X}^{n-1}) - Y)\nabla\Pi(\tilde{X}^{n-1})^T A^T + \frac{\lambda}{N}\alpha^{n-1}AA^T) (A \nabla\Pi(\tilde{X}^{n-1})\nabla\Pi(\tilde{X}^{n-1})^T A^T + \frac{\lambda}{N}AA^T)^{-1}. \quad (19)$$

Data availability

The source code of the method is freely available at <https://gricad-gitlab.univ-grenoble-alpes.fr/GruLab/afmfit> under the GPL-3.0 license. AFMfit comes as a Python package that can be installed with the *pip* package manager. We ship it with the NOLB normal modes package, which runs on Linux and MacOS operating systems. The package provides code examples with GUI in the form of Jupyter notebooks. We also provide the reconstructed models for experiments with EF2, FVA and TRPV3, along with the synthetic AFM images of EF2 and the experimental AFM images of

FVA. The raw HS-AFM images from SI movies 7 and 11 from [34] were kindly provided by Simon Scheuring. The original raw and corrected AFM images of FVA are deposited by Jean-Luc Pellequer at <http://doi.org/10.5281/zenodo.11234725>.

Acknowledgements

We are very grateful to Simon Scheuring from Cornell University for providing us raw HS-AFM data of TRPV3. This work is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02).

Supplementary information

Supplementary Video 1: Reconstructed motion of FVA along the first principal component.

Supplementary Video 2: Reconstructed motion of FVA along the second principal component.

Supplementary Video 3: Fitting results of the first HS-AFM movie of TRPV3. Top left: the experimental HS-AFM movie of TRPV3. Bottom left: the pseudo-HS-AFM movie reconstructed from the fitted models. Top right: the difference between the experimental and the reconstructed HS-AFM movies. Bottom right: the fitted models.

Supplementary Video 4: Reconstructed motion of the TRPV3 tetramer along the second principal component for the first HS-AFM movie of TRPV3.

Supplementary Video 5: Reconstructed motion of the TRPV3 pentamer along the fifth principal component for the first HS-AFM movie of TRPV3.

Supplementary Video 6 Fitting results of the second HS-AFM movie of TRPV3. Top left: the experimental HS-AFM movie of TRPV3. Bottom left: the pseudo-HS-AFM movie reconstructed from the fitted models. Top right: the difference between the experimental and the reconstructed HS-AFM movies. Bottom right: the fitted models.

References

- [1] Kodera, N., Yamamoto, D., Ishikawa, R., Ando, T.: Video imaging of walking myosin v by high-speed atomic force microscopy. *Nature* **468**(7320), 72–76 (2010)
- [2] Shibata, M., Nishimasu, H., Kodera, N., Hirano, S., Ando, T., Uchihashi, T., Nureki, O.: Real-space and real-time dynamics of crispr-cas9 visualized by high-speed atomic force microscopy. *Nature communications* **8**(1), 1430 (2017)
- [3] Kodera, N., Noshiro, D., Dora, S.K., Mori, T., Habchi, J., Blocquel, D., Gruet, A., Dosnon, M., Salladini, E., Bignon, C., *et al.*: Structural and dynamics analysis of intrinsically disordered proteins by high-speed atomic force microscopy. *Nature Nanotechnology* **16**(2), 181–189 (2021)

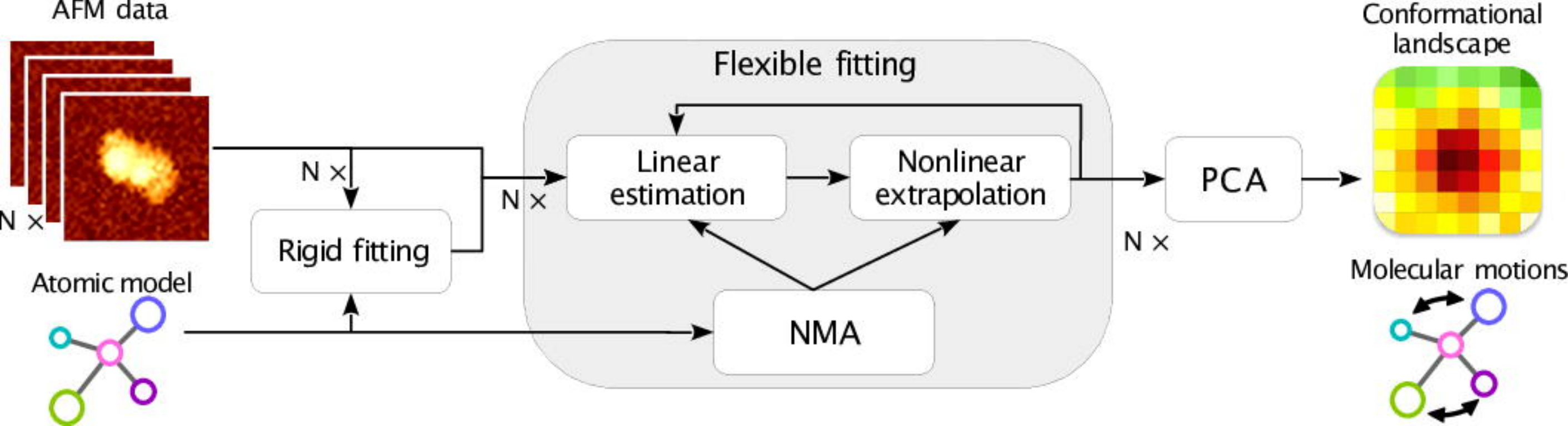
- [4] Naydenova, K., Russo, C.J.: Measuring the effects of particle orientation to improve the efficiency of electron cryomicroscopy. *Nature communications* **8**(1), 629 (2017)
- [5] Ando, T., Uchihashi, T., Kodera, N.: High-speed afm and applications to biomolecular systems. *Annual review of biophysics* **42**, 393–414 (2013)
- [6] Flechsig, H., Ando, T.: Protein dynamics by the combination of high-speed afm and computational modeling. *Current Opinion in Structural Biology* **80**, 102591 (2023)
- [7] Trinh, M.-H., Odorico, M., Pique, M.E., Teulon, J.-M., Roberts, V.A., Ten Eyck, L.F., Getzoff, E.D., Parot, P., Shu-wen, W.C., Pellequer, J.-L.: Computational reconstruction of multidomain proteins using atomic force microscopy data. *Structure* **20**(1), 113–120 (2012)
- [8] Scheuring, S., Boudier, T., Sturgis, J.N.: From high-resolution afm topographs to atomic models of supramolecular assemblies. *Journal of Structural Biology* **159**(2), 268–276 (2007)
- [9] Chaves, R.C., Teulon, J.-M., Odorico, M., Parot, P., Chen, S.-w.W., Pellequer, J.-L.: Conformational dynamics of individual antibodies using computational docking and afm. *Journal of Molecular Recognition* **26**(11), 596–604 (2013)
- [10] Niina, T., Matsunaga, Y., Takada, S.: Rigid-body fitting to atomic force microscopy images for inferring probe shape and biomolecular structure. *PLOS Computational Biology* **17**(7), 1009215 (2021)
- [11] Ogane, T., Noshiro, D., Ando, T., Yamashita, A., Sugita, Y., Matsunaga, Y.: Development of hidden markov modeling method for molecular orientations and structure estimation from high-speed atomic force microscopy time-series images. *PLOS Computational Biology* **18**(12), 1010384 (2022)
- [12] Amyot, R., Marchesi, A., Franz, C.M., Casuso, I., Flechsig, H.: Simulation atomic force microscopy for atomic reconstruction of biomolecular structures from resolution-limited experimental images. *PLoS computational biology* **18**(3), 1009970 (2022)
- [13] Amyot, R., Nakamoto, K., Kodera, N., Flechsig, H.: Predicting the placement of biomolecular structures on afm substrates based on electrostatic interactions. *Frontiers in Molecular Biosciences* **10** (2023)
- [14] Dasgupta, B., Miyashita, O., Tama, F.: Reconstruction of low-resolution molecular structures from simulated atomic force microscopy images. *Biochimica et Biophysica Acta (BBA)-General Subjects* **1864**(2), 129420 (2020)
- [15] Niina, T., Fuchigami, S., Takada, S.: Flexible fitting of biomolecular structures

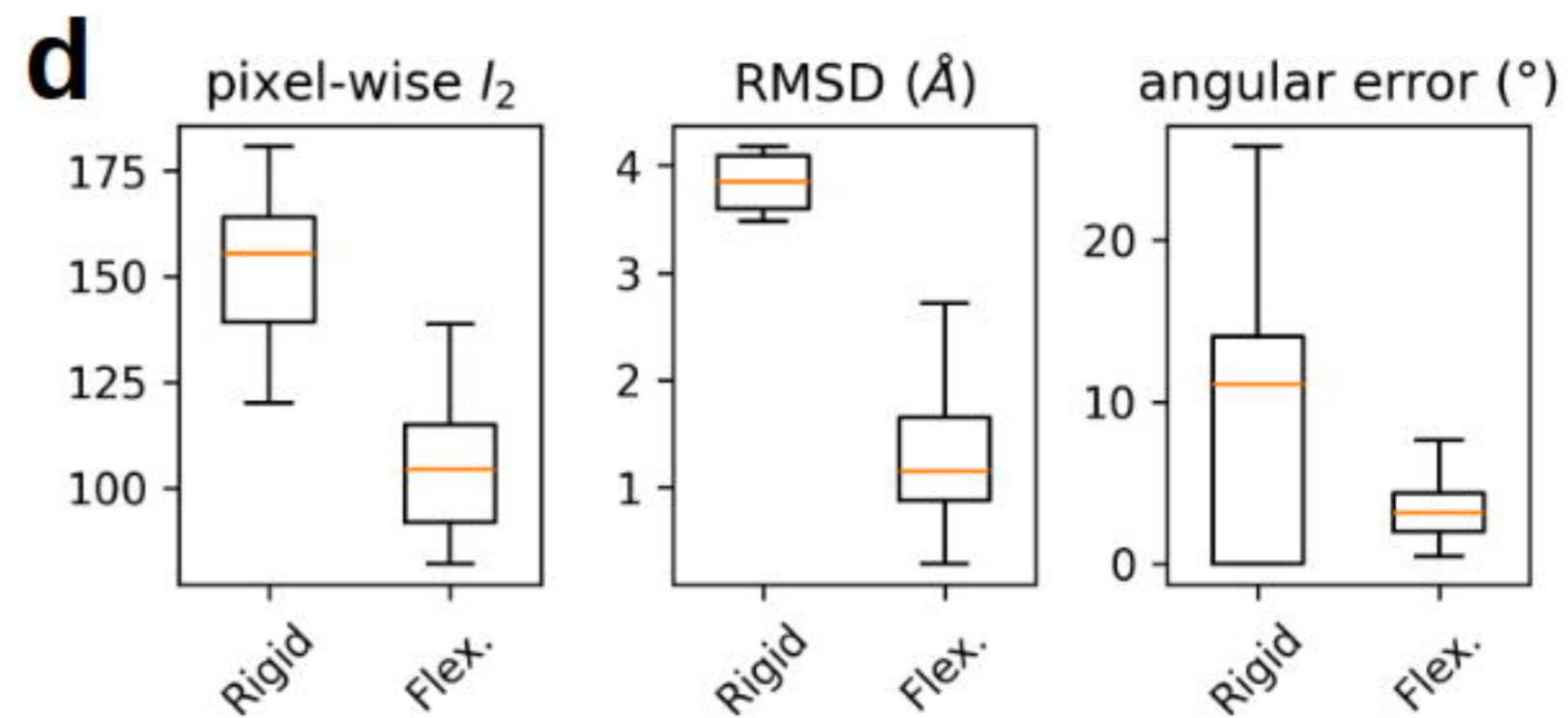
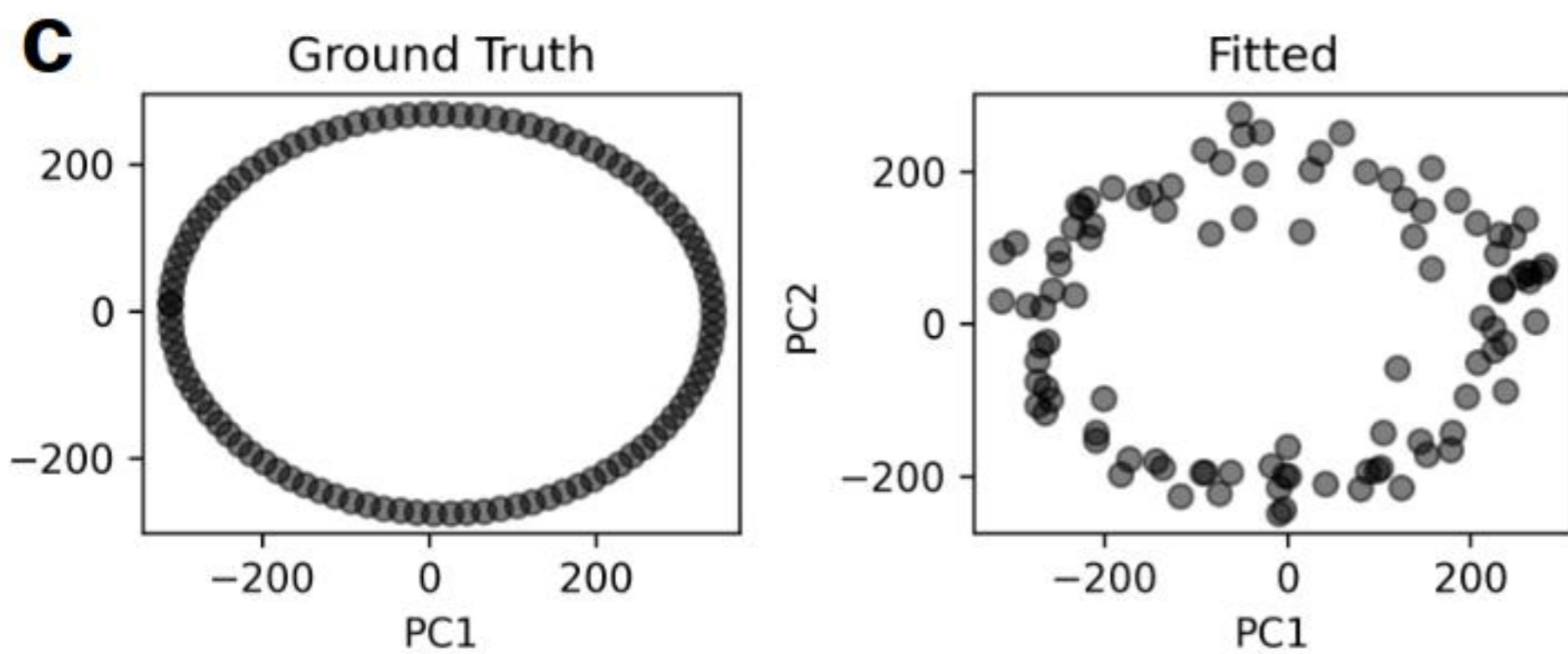
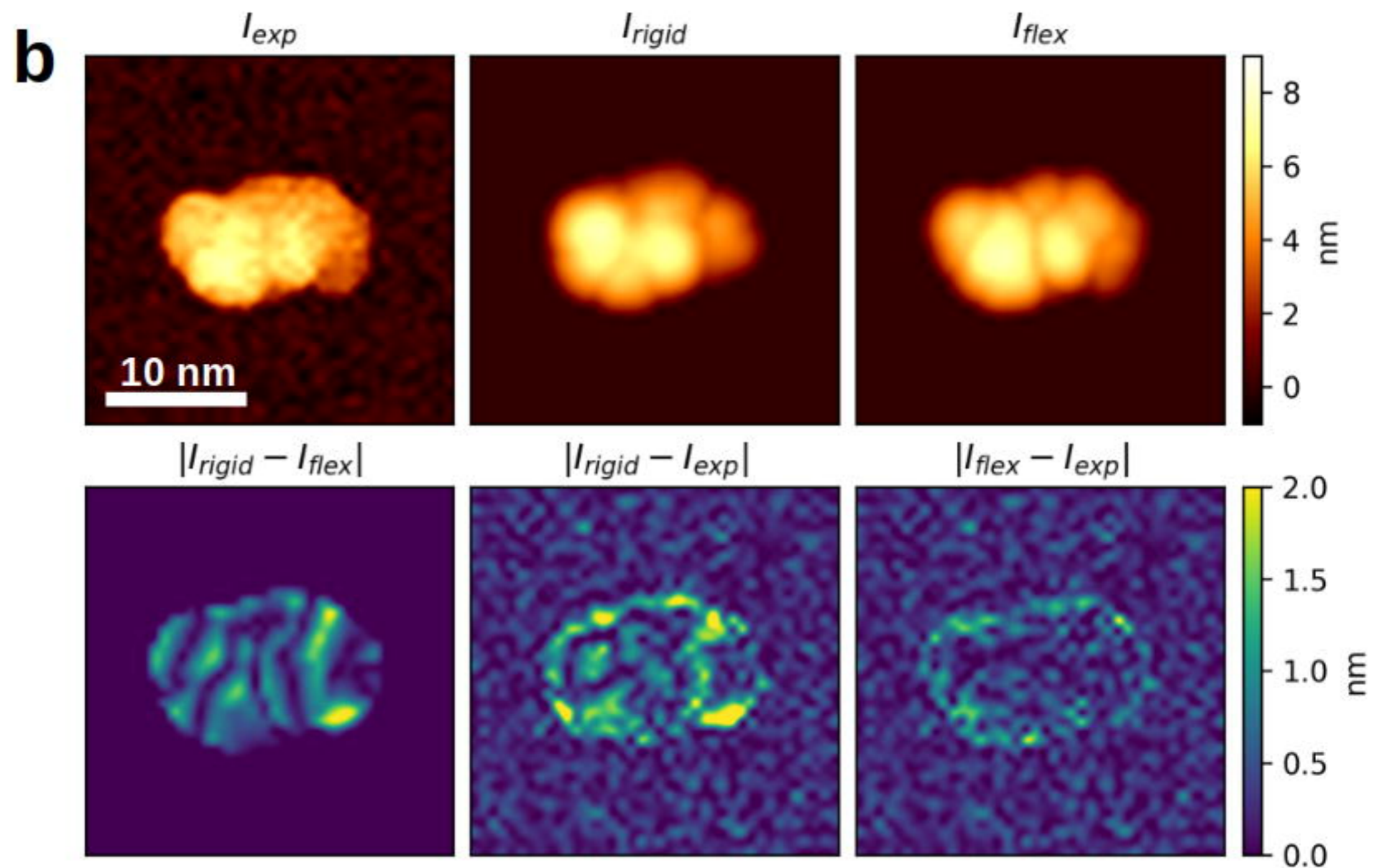
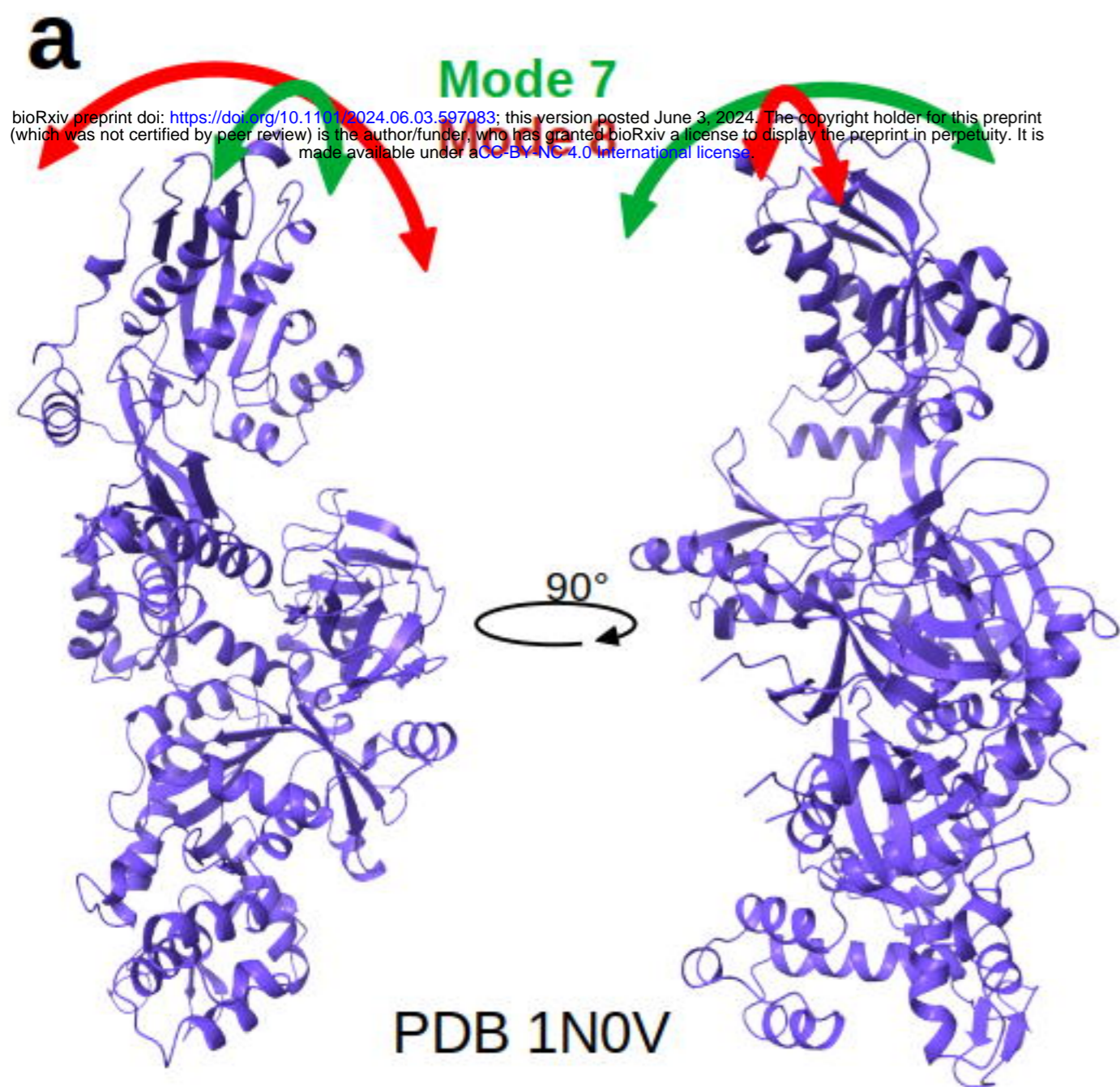
- to atomic force microscopy images via biased molecular simulations. *Journal of Chemical Theory and Computation* **16**(2), 1349–1358 (2020)
- [16] Dasgupta, B., Miyashita, O., Uchihashi, T., Tama, F.: Reconstruction of three-dimensional conformations of bacterial clpb from high-speed atomic-force-microscopy images. *Frontiers in Molecular Biosciences* **8**, 704274 (2021)
- [17] Jin, Q., Sorzano, C.O.S., La Rosa-Trevín, J.M., Bilbao-Castro, J.R., 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* **22**(3), 496–506 (2014)
- [18] Frank, J., Ourmazd, A.: Continuous changes in structure mapped by manifold embedding of single-particle data in cryo-em. *Methods* **100**, 61–67 (2016)
- [19] Punjani, A., Fleet, D.J.: 3d variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-em. *Journal of structural biology* **213**(2), 107702 (2021)
- [20] Zhong, E.D., Bepler, T., Berger, B., Davis, J.H.: Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks. *Nature methods* **18**(2), 176–185 (2021)
- [21] Vuillemot, R., Mirzaei, A., Harastani, M., Hamitouche, I., Fréchin, L., Klaholz, B.P., Miyashita, O., Tama, F., Rouiller, I., Jonic, S.: Mdspace: Extracting continuous conformational landscapes from cryo-em single particle datasets using 3d-to-2d flexible fitting based on molecular dynamics simulation. *Journal of Molecular Biology* **435**(9), 167951 (2023)
- [22] Vuillemot, R., Rouiller, I., Jonić, S.: Mdtomo method for continuous conformational variability analysis in cryo electron subtomograms based on molecular dynamics simulations. *Scientific Reports* **13**(1), 10596 (2023)
- [23] Gilles, M.A., Singer, A.: A bayesian framework for cryo-em heterogeneity analysis using regularized covariance estimation. *bioRxiv* (2023)
- [24] Punjani, A., Fleet, D.J.: 3dflex: determining structure and motion of flexible proteins from cryo-em. *Nature Methods* **20**(6), 860–870 (2023)
- [25] Herreros, D., Lederman, R.R., Krieger, J.M., Jiménez-Moreno, A., Martínez, M., Myška, D., Strelak, D., Filipovic, J., Sorzano, C.O., Carazo, J.M.: Estimating conformational landscapes from cryo-em particles by 3d zernike polynomials. *Nature Communications* **14**(1), 154 (2023)
- [26] Vuillemot, R., Jonić, S.: Combined bayesian and normal mode flexible fitting with hamiltonian monte carlo sampling for cryo electron microscopy. In: 2021 29th European Signal Processing Conference (EUSIPCO), pp. 1211–1215 (2021).

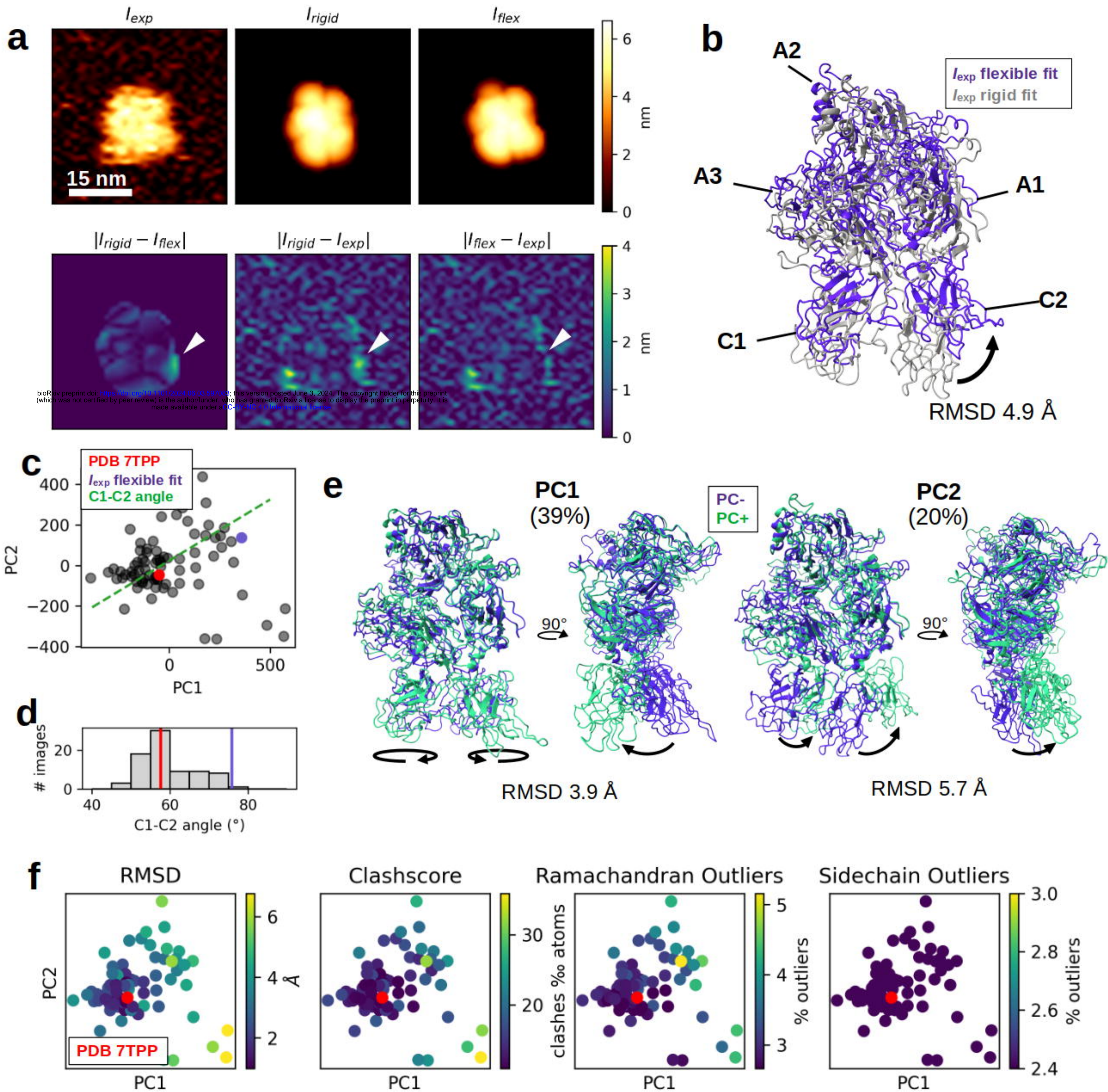
IEEE

- [27] Hamitouche, I., Jonic, S.: DeepHemmma: Resnet-based hybrid analysis of continuous conformational heterogeneity in cryo-em single particle images. *Frontiers in Molecular Biosciences* **9**, 965645 (2022)
- [28] Harastani, M., Eltsov, M., Leforestier, A., Jonic, S.: Hemmma-3d: cryo electron tomography method based on normal mode analysis to study continuous conformational variability of macromolecular complexes. *Frontiers in molecular biosciences* **8**, 663121 (2021)
- [29] Tang, W.S., Silva-Sánchez, D., Giraldo-Barreto, J., Carpenter, B., Hanson, S.M., Barnett, A.H., Thiede, E.H., Cossio, P.: Ensemble reweighting using cryo-em particle images. *The Journal of Physical Chemistry B* **127**(24), 5410–5421 (2023)
- [30] Tang, W.S., Zhong, E.D., Hanson, S.M., Thiede, E.H., Cossio, P.: Conformational heterogeneity and probability distributions from single-particle cryo-electron microscopy. *Current Opinion in Structural Biology* **81**, 102626 (2023)
- [31] Toader, B., Sigworth, F.J., Lederman, R.R.: Methods for cryo-em single particle reconstruction of macromolecules having continuous heterogeneity. *Journal of Molecular Biology* **435**(9), 168020 (2023)
- [32] Hoffmann, A., Grudin, S.: Nolib: Nonlinear rigid block normal-mode analysis method. *Journal of chemical theory and computation* **13**(5), 2123–2134 (2017)
- [33] Chaves, R.C., Dahmane, S., Odorico, M., Nicolaes, G.A., Pellequer, J.-L.: Factor va alternative conformation reconstruction using atomic force microscopy. *Thrombosis and haemostasis* **112**(12), 1167–1173 (2014)
- [34] Lansky, S., Betancourt, J.M., Zhang, J., Jiang, Y., Kim, E.D., Paknejad, N., Nimigean, C.M., Yuan, P., Scheuring, S.: A pentameric trpv3 channel with a dilated pore. *Nature* **621**(7977), 206–214 (2023)
- [35] Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., Ferrin, T.E.: Ucsf chimeraX: Meeting modern challenges in visualization and analysis. *Protein science* **27**(1), 14–25 (2018)
- [36] Jørgensen, R., Ortiz, P.A., Carr-Schmid, A., Nissen, P., Kinzy, T.G., Andersen, G.R.: Two crystal structures demonstrate large conformational changes in the eukaryotic ribosomal translocase. *Nature Structural & Molecular Biology* **10**(5), 379–385 (2003)
- [37] Niina, T., Kato, S.: ToruNiina/afmize: Version 1.1.0. <https://doi.org/10.5281/zenodo.3362044> . <https://doi.org/10.5281/zenodo.3362044>
- [38] Saleh, M., Peng, W., Quinn-Allen, M.A., Macedo-Ribeiro, S., Fuentes-Prior, P.,

- Bode, W., Kane, W.H.: The factor v c1 domain is involved in membrane binding: identification of functionally important amino acid residues within the c1 domain of factor v using alanine scanning mutagenesis. *Thrombosis and haemostasis* **91**(01), 16–27 (2004)
- [39] Peng, W., Quinn-Allen, M., Kane, W.: Mutation of hydrophobic residues in the factor va c1 and c2 domains blocks membrane-dependent prothrombin activation. *Journal of Thrombosis and Haemostasis* **3**(2), 351–354 (2005)
- [40] Ruben, E.A., Summers, B., Rau, M.J., Fitzpatrick, J.A., Di Cera, E.: Cryo-em structure of the prothrombin-prothrombinase complex. *Blood, The Journal of the American Society of Hematology* **139**(24), 3463–3473 (2022)
- [41] Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., *et al.*: Molprobity: More and better reference data for improved all-atom structure validation. *Protein Science* **27**(1), 293–315 (2018)
- [42] Hoffmann, A., Perrier, V., Grudinin, S.: A novel fast fourier transform accelerated off-grid exhaustive search method for cryo-electron microscopy fitting. *Journal of Applied Crystallography* **50**(4), 1036–1047 (2017)

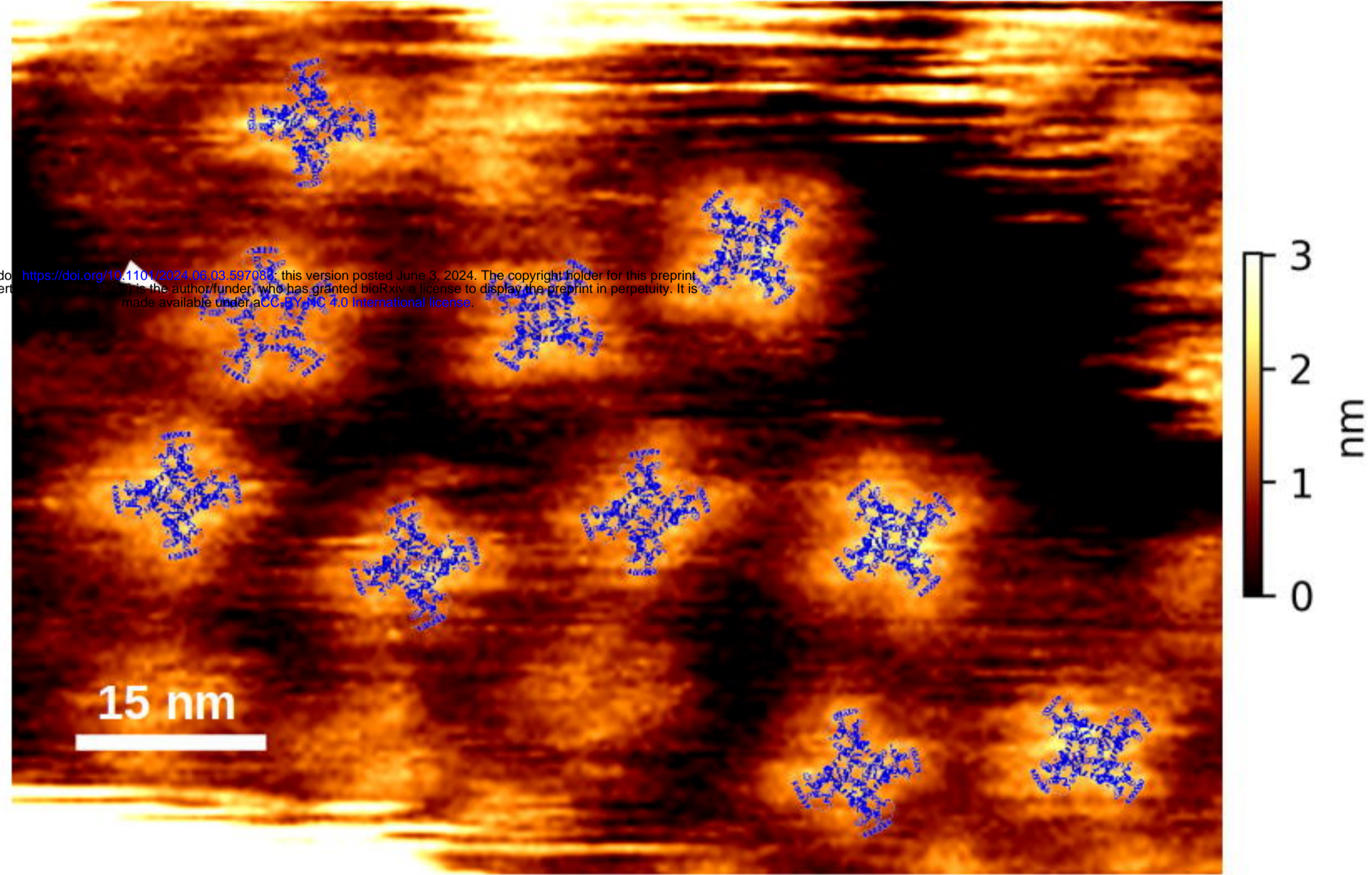
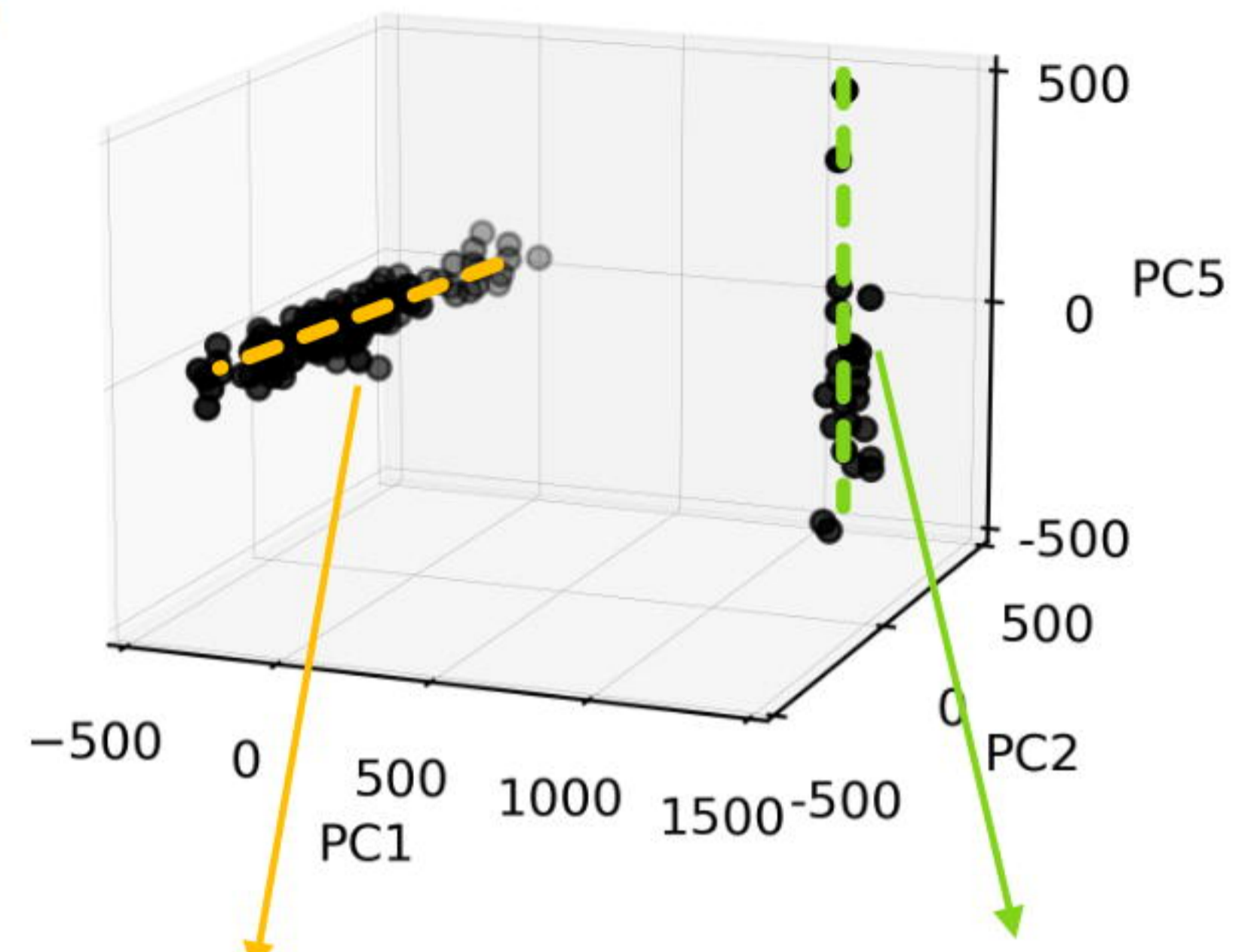






a

bioRxiv preprint doi: <https://doi.org/10.1101/2024.06.03.597057>; this version posted June 3, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**b****c**