Combined Bayesian and Normal Mode Flexible Fitting with Hamiltonian Monte Carlo Sampling for Cryo Electron Microscopy

Rémi Vuillemot and Slavica Jonić*
IMPMC - UMR 7590 CNRS, Sorbonne Université, MNHN, Paris, France
remi.vuillemot@upmc.fr, slavica.jonic@upmc.fr (corresponding*)

***Abstract*—Density volumes obtained by three-dimensional reconstruction of biomolecular complexes from cryogenic electron microscopy (cryo-EM) images (also known as cryo-EM maps) can be interpreted in terms of atomic positions by flexible fitting. The fitting modifies an available atomic structure to match the target EM map. The most accurate fitting methods are based on atomic-coordinate degrees of freedom (e.g. Bayesian flexible fitting) but come with high computational cost for large required displacements. To reduce the computational cost, methods based on Normal Modes Analysis (NMA) decrease the number of degrees of freedom to only several collective motions (described by normal modes). The NMA-based methods are well-suited for global atomic displacements (large collective motions) but are suboptimal regarding local atomic displacements. To take advantages of both methods, we propose to combine them. We tested our method using synthetic and experimental cryo-EM maps of a complex with large-scale conformational changes (p97 ATPase). We show that the combination of both approaches efficiently performs global and local atomic displacements and that it can be more efficient and precise than any of the two approaches alone. To the best of our knowledge, this is the first method combining Bayesian and normal mode flexible fitting approaches.**

***Index Terms*—Cryo-EM, Flexible fitting, Normal Modes, Bayesian model, Hamiltonian Monte Carlo, Molecular Dynamics**

I. Introduction

For many years, cryogenic electron microscopy (cryo-EM) has been drawing attention for its capacity to image structures of biological macromolecular complexes in their close-to-native conditions. Until recently, the resolution of the three-dimensional (3D) maps reconstructed from two-dimensional cryo-EM images of biomolecular complexes (also known as cryo-EM density maps or cryo-EM maps) has been lower than the resolution of structures obtained by X-ray crystallography (the standard technique to obtain structures at high resolution until then [1]). Recent progress in cryo-EM instruments (particularly detector technology advances), high performance computing technologies, and in image processing algorithms and software enabled cryo-EM structure reconstruction at near-atomic resolutions on a more routine basis [2].

To accomplish specific biological functions, biomolecular complexes change their conformations. The study of conformational variability of complexes is the key to deciphering their biological functions and to structure-based drug development. Unlike X-ray crystallography, cryo-EM allows imaging and reconstructing multiple different conformational states of a complex from the same sample [3]. It requires vitrified samples instead of high quality crystals that are required for X-ray crystallography. The possibility of achieving near-atomic resolution of 3D reconstruction and studying conformational variability of complexes in near-physiological conditions make cryo-EM a highly interesting technique complementary to X-ray crystallography. Indeed, a cryo-EM map reconstructed from cryo-EM images is a density volume and the atomic coordinates of the complex are usually obtained by flexible fitting of an available atomic X-ray crystallography structure of a similar complex or the same complex but in a different conformation (e.g., structures available in the public database named Protein Data Bank - PDB, www.rcsb.org).

Flexible fitting methods displace the atoms of the reference atomic PDB structure to achieve the best match between the target cryo-EM map and the density map simulated from the displaced atoms [4]–[6]. The resulting atomic structure, corresponding to the best fit, may in its turn be deposited in the PDB database, which is usually done when the procedure results in the discovery of new conformations. Such procedures are computationally challenging. Most accurate methods simulate the physical interaction between atoms by estimating the force field for each atom [7], [8]. Molecular Dynamics (MD) simulation estimates deterministic trajectories based on classical mechanics principles [6], [7]. Monte-Carlo (MC) is a stochastic method that generates physically probable conformations [5], [9]. In these methods, the number of the conformational degrees of freedom is equal to the number of the atomic coordinates, which allows well-fitting both global dynamics (collective displacement of atoms) and local dynamics (local displacement of atoms). This high number of degrees of freedom comes with a high computational cost, in particular for large-scale conformational changes (consisting of large collective atomic motions).

To speed up global dynamics simulations, Normal Modes Analysis (NMA) [10] is used in some flexible fitting methods to reduce the number of the conformational degrees of freedom to a small number of vectors (known as normal modes) along which atoms move with the corresponding amplitudes (one

amplitude per normal mode) [4]. NMA only calculates the directions of motions (normal modes) but not the normal-mode amplitudes. In flexible fitting methods based on NMA, these amplitudes are determined by displacing the reference atomic structure along normal modes until it matches the target cryo-EM map. Flexible fitting using low-frequency normal modes proved to be well-suited for fitting between conformations of complexes with large-scale conformational transitions (large global dynamics). This suggests that low-frequency normal modes describe collective (global) atomic motions well and it was also used in other fitting contexts (e.g., NMA-based fitting of cryo-EM images with a reference atomic or pseudoatomic structure to determine continuous conformational variability of complexes [11]). However, additional exploring of local motions with NMA-based fitting methods would require including high-frequency normal modes, either using an entire set of modes (computationally challenging) or a subset of modes (none of the selection criteria proposed so far is optimal and may miss to select some relevant modes).

In this article, we propose a method that combines exploring global dynamics with low-frequency normal modes and exploring local dynamics with Bayesian fitting. Our approach is built using a Bayesian model and sampled with Hamiltonian Monte Carlo (HMC) [9], [12], [13]. HMC combines MD trajectory and MC stochastic sampling to obtain a more efficient sampling. Normal-mode amplitude displacement is added to reduce the computational cost of the fitting. The tests of the method were performed using synthetic and experimental cryo-EM maps of the p97 ATPase complex [14]. We show that the combination of the Bayesian and normal mode approaches efficiently performs both global and local atomic displacements and that it can be more efficient and precise than any of the two approaches alone. To the best of our knowledge, this is the first method combining fitting based on Bayesian inference and fitting based on normal modes.

II. Methods

A. Bayesian Model

Flexible fitting of a given atomic structure (reference structure) into a cryo-EM map (target map) can be defined using the Bayesian formalism. Notably, the posterior distribution can be written as follows:

$$\mathrm{P}(\mathbf{r}|\rho_{exp}) \propto \mathrm{P}(\rho_{exp}|\mathbf{r})\mathrm{P}(\mathbf{r}), \tag{1}$$

where $\mathbf{r}$ is the vector of $N \times 3$ Cartesian atomic coordinates, N is the number of atoms, and ρ_{exp} is the target map.

The most common choice for the prior distribution $\mathrm{P}(\mathbf{r})$ is the Boltzmann distribution [7]:

$$\mathrm{P}(\mathbf{r}) = e^{-\frac{U_p(\mathbf{r})}{k_B T}}, \tag{2}$$

where k_B is the Boltzmann constant, T is the temperature, and U_p is the potential energy of the system. The potential energy defines the physical interaction between atoms, which determines energetically accessible conformations. It corresponds to a sum of bonded and non-bonded potentials.

The likelihood distribution $\mathrm{P}(\rho_{exp}|\mathbf{r})$ is the function that assess the goodness of the match between the target map and the reference structure that is modified (atomic positions changed) during the fitting. To this end, a density map is simulated from the atomic structure and compared with the target map. The simulated map is obtained by placing a 3D isotropic Gaussian function at the position of each atom, $\mathbf{r}_n$ $(n = 1, N)$, and by integrating these Gaussian functions over the center of each voxel. The value of the voxel with the coordinates (i, j, k) in the simulated map is as follows:

$$\rho^{\mathbf{r}}_{sim}(i,j,k) = \sum_{n=1}^{N} \frac{1}{(2\pi\sigma^2)^{3/2}} \exp\big(-\frac{1}{2\sigma^2}\|[i\ j\ k] - \mathbf{r}_n\|^2\big), \tag{3}$$

where σ is the standard deviation of the 3D Gaussian functions. Note here that $\mathbf{r}_n$ is the position of the n-th atom and that $\mathbf{r}$ contains the positions of all atoms.

The likelihood is assumed to be normally distributed and centered at the simulated map, which results in the following:

$$\mathrm{P}(\rho_{exp}|\mathbf{r}) = \frac{1}{\sqrt{2\pi\sigma_\rho^2}} \prod_{l=1}^{N_{vox}} \exp\big(-\frac{1}{2\sigma_\rho^2}(\rho_{exp}(l) - \rho^{\mathbf{r}}_{sim}(l))^2\big), \tag{4}$$

where the voxel coordinates (i, j, k) of ρ_{exp} and $\rho^{\mathbf{r}}_{sim}$ are, for simplicity reasons, replaced by the voxel indexes l, N_{vox} is the number of voxels $(l = 1, N_{vox})$, and σ_ρ is the standard deviation of the likelihood distribution that determines the desired fitness precision.

Finally, the resulting posterior distribution can be expressed by combining the prior (2) and likelihood (4) distributions. It is a common practice to use the logarithm of the posterior distribution, which simplifies the expression as follows:

$$\log \mathrm{P}(\mathbf{r}|\rho_{exp}) = -\frac{1}{2\sigma_\rho^2}\|\rho_{\mathbf{exp}} - \rho^{\mathbf{r}}_{\mathbf{sim}}\|^2 - \frac{1}{k_B T}U_p(\mathbf{r}) + C, \tag{5}$$

where $C = \log \frac{1}{\sqrt{2\pi\sigma_\rho^2}}$, and $\rho_{\mathbf{exp}}$ and $\rho^{\mathbf{r}}_{\mathbf{sim}}$ are vector versions of the target and simulated maps, respectively (e.g., $\rho_{\mathbf{exp}}$ is a vector of voxel values in the target map, $\rho_{exp}(l), l = 1, N_{vox}$).

In this model, the conformational degrees of freedom are $N \times 3$ atomic coordinate displacements $\Delta\mathbf{r}$ defined with respect to their position in the reference structure, $\mathbf{r}_{init}$:

$$\mathbf{r} = \mathbf{r}_{init} + \Delta\mathbf{r}. \tag{6}$$

This gives the model a flexibility for global and local dynamics but also comes with a high computational cost of sampling.

B. Normal Modes Analysis

Instead of $N \times 3$ conformational degrees of freedom (atomic displacements $\Delta\mathbf{r}$), one can consider using a less detailed model, which is particularly interesting when dealing with lower-resolution data. One popular model is the elastic network model [10], where the potential energy is described by simple harmonic potentials between close atoms and the dynamics is estimated by Normal Mode Analysis (NMA) of the reference structure [10]. NMA consists of diagonalizing a

Hessian matrix of second derivatives of the potential energy function of size $(N \times 3)^2$. This produces a matrix of normal modes and their associated frequencies. The total number of normal modes is $N \times 3$ and the length of each normal mode is $N \times 3$. Usually, a small subset of M lowest-frequency normal modes (describing global, collective motions) is selected to displace atoms to fit the target map. The atomic displacement with respect to the reference structure is determined by a linear combination of the selected modes, as follows:

$$\Delta\mathbf{r}(\mathbf{q}) = \mathbf{q} \cdot \mathbf{A}, \tag{7}$$

where $\mathbf{A} = \{\mathbf{a}_i\}$ is the matrix of the selected M normal modes (size $M \times (N \times 3)$) and $\mathbf{q} = \{q_i\}$ is the vector of M coefficients of the linear combination (M normal-mode amplitudes). The displaced coordinates are $\mathbf{r} = \mathbf{r}_{init} + \Delta\mathbf{r}(\mathbf{q})$.

The optimal value of $\mathbf{q}$ to fit the target map is usually obtained by optimizing a cost function like the Correlation Coefficient (CC) [4], [8].

$$CC = \frac{\sum_{l=1}^{N_{vox}} \rho^{\mathbf{r}}_{sim}(l)\rho_{exp}(l)}{\sqrt{\sum_{l=1}^{N_{vox}} \rho^{\mathbf{r}}_{sim}(l)^2}\sqrt{\sum_{l=1}^{N_{vox}} \rho_{exp}(l)^2}}. \tag{8}$$

NMA-based fitting is much faster than fitting with other methods as the number of degrees of freedom is reduced to M (M normal-mode amplitudes q_i, where $M << N$ and, usually, $M < 10$). However, the selected lowest-frequency normal modes fit well global motions but not local motions.

C. Combined Bayesian and Normal Mode Flexible Fitting

In this article, we propose to combine normal mode flexible fitting (small number of degrees of freedom describing well global motions) with Bayesian flexible fitting (large number of degrees of freedom describing well local motions). With this model, we expect to reduce the computational cost of fitting large-scale conformational transitions, thanks to NMA-based fitting, while maintaining the precision of fitting local dynamics with Bayesian fitting.

To this end, we propose to modify the atomic coordinate displacement $\Delta\mathbf{r}$ in equation (6) as follows:

$$\mathbf{r} = \mathbf{r}_{init} + \Delta\mathbf{r}_{global}(\mathbf{q}) + \Delta\mathbf{r}_{local}, \tag{9}$$

where $\Delta\mathbf{r}_{global}(\mathbf{q})$ is the atomic displacement by normal mode fitting ($\Delta\mathbf{r}(\mathbf{q})$ in equation (7), M unknown parameters, $M << N$), $\Delta\mathbf{r}_{local}$ is the atomic displacement by Bayesian fitting ($\Delta\mathbf{r}$ in equation (6), $N \times 3$ unknown parameters).

In this model, the total number of parameters is $M + N \times 3$. These parameters correspond to $\mathbf{q}$ and $\Delta\mathbf{r}_{local}$ and will be estimated simultaneously by sampling from the resulting posterior distribution $\mathrm{P}(\mathbf{r}|\rho_{exp})$. In the next section, we focus on methods for sampling such posterior distributions.

D. Sampling Conformational Space

a) Monte-Carlo methods: Monte-Carlo (MC) are stochastic approaches that are designed to generate samples from posterior distributions. One of the most popular algorithm is the Metropolis algorithm [15]. This method is an iterative algorithm that performs the sampling in two steps. First, random displacements are applied to atomic positions $\mathbf{r}^n$ at the current n-th iteration, from a transition distribution $\mathcal{T}$ (generally a Gaussian distribution), and a candidate structure $\tilde{\mathbf{r}}$ is generated. Then, the transition kernel is adjusted to the target distribution by accepting or rejecting this candidate structure with the acceptation probability α:

$$\alpha = \min\left(1, \frac{\mathrm{P}(\tilde{\mathbf{r}}|\rho_{exp})/\mathcal{T}(\tilde{\mathbf{r}}|\mathbf{r}^n)}{\mathrm{P}(\mathbf{r}^n|\rho_{exp})/\mathcal{T}(\mathbf{r}^n|\tilde{\mathbf{r}})}\right). \tag{10}$$

This acceptation probability means that the candidate structure is always accepted if it increases the posterior probability, and sometimes accepted if the posterior probability is not too much decreased.

The main problem of the Metropolis algorithm comes from his random walk behavior. In many cases, random displacement of atoms is very likely to be rejected, especially in high density regions where atoms are very close and small displacements make them overlap. In such cases, the acceptation rate becomes dramatically low, which results in high inefficiency of the sampling.

b) Molecular Dynamics: One of the most widely used simulation approaches to exploring conformational dynamics is Molecular Dynamics (MD) [7]. Unlike MC methods, MD generates deterministic trajectories based on classical mechanics, which describe the evolution of the structure over time. In a typical MD simulation, the motion of atoms is estimated by numerical integration of Newton equation by setting the force field to be the gradient of the potential energy, $U_p(\mathbf{r})$:

$$\mathbf{F} = \partial U_p(\mathbf{r})/\partial\mathbf{r}. \tag{11}$$

MD simulations have been applied to flexible fitting by adding a biasing potential to $U_p(\mathbf{r})$ i.e., by modifying the force field $\mathbf{F}$ [6]. In these approaches, the biasing potential is a measure of the goodness of the fit that leads the simulation to the target density map. The biasing potential measures include CC (8) and mean square error. When the mean square error is used as the biasing potential, the MD-based fitting approach is equivalent to the one with the force field set to be the gradient of the log-posterior distribution in the Bayesian model (5):

$$\mathbf{F} = \partial \log \mathrm{P}(\mathbf{r}|\rho_{exp})/\partial\mathbf{r}. \tag{12}$$

MD is very time consuming. The time step of the integrator must remain very small in order to maintain the stability of the trajectory. This results in high computation times, especially for large-scale conformational changes [9].

c) Hamiltonian Monte Carlo: Hamiltonian Monte-Carlo (HMC) method aims at generating samples more efficiently than MD and MC methods. HMC was originally called Hybrid Monte-Carlo [12] since it combines deterministic MD trajectories and stochastic Metropolis acceptation scheme. At each iteration of the algorithm, a standard MD simulation is performed and the candidate structure is accepted with the probability acceptance α as in Metropolis algorithm (10). As

HMC is based on MD, it uses the gradient of the log-posterior distribution in the Bayesian model (12) to generate the trajectories. HMC has been successfully applied to flexible fitting [5] but only using all ($N \times 3$) atomic degrees of freedom (search for unknown parameters $\Delta\mathbf{r}$ in equation (6)). This article shows the first HMC application to a combined Bayesian and normal mode flexible fitting (search for $M + N \times 3$ unknown parameters, which includes M normal-mode amplitudes). We have implemented the log-posterior gradient calculation and the HMC sampler for this combined model in Python. Our method searches for $\mathbf{q}$ and $\Delta\mathbf{r}_{local}$ (9) so as to efficiently fit global and local dynamics.

III. Results

In this section, we use synthetic and experimental data of the human p97 ATPase complex, involved in various cellular processes, to compare the proposed combined HMC-based Bayesian and normal mode flexible fitting approach (atomic coordinates determined by $\mathbf{r} = \mathbf{r}_{init} + \Delta\mathbf{r}_{global}(\mathbf{q}) + \Delta\mathbf{r}_{local}$ (9)) with the HMC-based Bayesian and HMC-based normal mode flexible fitting approaches alone (atomic coordinates determined by $\mathbf{r} = \mathbf{r}_{init} + \Delta\mathbf{r}_{local}$ and $\mathbf{r} = \mathbf{r}_{init} + \Delta\mathbf{r}_{global}(\mathbf{q})$, respectively). To reduce the computational complexity of each approach, we here use Carbon Alpha (CA) atoms only, which is a common practice (coarse-grain approaches).

The three approaches are compared using the data of two out of several p97 conformations solved by cryo-EM [14]. The most striking difference between the two conformations is in the position of one of the p97 domains (known as N domain), which is clearly "up" in one and "down" in the other [14]. We use the following data of these two conformations, publicly available in the PDB database and the EMDB database (www.ebi.ac.uk/pdbe/emdb): 1) atomic structure of the "down" conformation (PDB code: *5ftm*); and 2) cryo-EM map of the "up" conformation (EMDB code: *EMD-3299*).

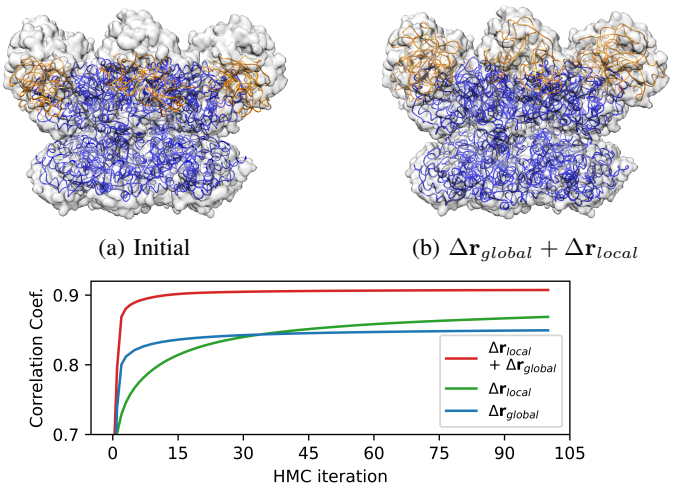
The results of the following two tests are shown here: 1) fitting of the PDB-*5ftm* CA-atom structure of the p97 ATPase "down" conformation into a synthetic target map obtained from the *5ftm* structure (the mode that lifts up the N domain of *5ftm* was identified and *5ftm* was displaced along this mode, namely mode 9, with the amplitude of -1500 and, then, energy-minimized with MD, to synthesize large global and small local displacements, respectively, and the resulting structure was used to synthesize the map); and 2) fitting of the PDB-*5ftm* CA-atom structure of the "down" conformation into the experimental *EMD-3299* map of the "up" conformation. In both cases, flexible fitting was performed using the first four non-rigid-body modes (modes 7-10), arbitrarily chosen so as to include mode 9 (N domain lifting). Note that the first six normal modes are related to combinations of rigid-body motions and are not used for flexible fitting. The current implementation of our approach does not perform rigid-body alignment and, thus, the reference (initial) atomic structure and the target map are required to be rigid-body aligned before the flexible fitting can be performed with our approach.

For the first test (test 1), the synthetic target map was obtained by converting the modified PDB-*5ftm* structure into a map of size of 128^3 voxels using the method of atomic scattering factors [16]. For the second test (test 2), the *EMD-3299* map was slightly low-pass filtered and size-reduced to 128^3 voxels, which reduces noise and speeds up calculations. Recall that the simulated maps during the fitting are obtained using our Gaussian kernel method (3). The use of the atomic scattering factors [16] for the synthesis of the target map was decided to get better resolution of the density in the target map (closer to the density resolution in the experimental cryo-EM maps) and, at the same time, to make the fitting more difficult as trying to fit a map whose density resolution is slightly different from the one that can be simulated with our Gaussian kernel method (3) during the fitting. The target maps used in test 1 and test 2 are superposed with the reference (initial) CA-atom structure in Fig. 1a and Fig. 2a, respectively. The N domain in both figures is displayed in orange color and is clearly in "down" conformation.

We performed 10 HMC runs over 100 iterations. The structures fitted to the target maps in test 1 and test 2, using the proposed method (averaging over 10 HMC runs), are shown in Fig. 1b and Fig. 2b, respectively. We can observe that, in both tests, the N domain moves from "down" conformation at the beginning (initial structure) to "up" conformation at the end of the fitting (fitted structure). Fig. 1c and Fig. 2c show the evolution of the CC (8), averaged over 10 HMC runs, as a function of HMC iterations, for test 1 and test 2, respectively. These figures show that the two approaches explicitly fitting global dynamics (incorporating normal-mode displacements) converge faster to the equilibrium than the approach based on fitting local dynamics. This can be interpreted by a fast large-scale displacement generated by normal modes in the first iterations, which we observed in the resulting normal-mode amplitudes. More importantly, Fig. 1c and Fig. 2c show that the method proposed here (combining fitting global and local dynamics) achieves the highest CC value in the smallest number of HMC iterations compared to the other two methods, in both tests (with synthetic and experimental target maps). However, in the case of the experimental map, the local fitting method achieves a slightly worse CC value than the proposed method, which suggests that the fitting by the proposed method strongly depends on local fitting in this specific data case. An explanation for this is that the selected normal modes may not describe the global dynamics sufficiently well in this particular data case. Other sets of normal modes may be used in the future to investigate their contributions. It should be noted that the search for normal-mode amplitudes takes negligible time with respect to the search for local fitting parameters.

IV. Conclusion

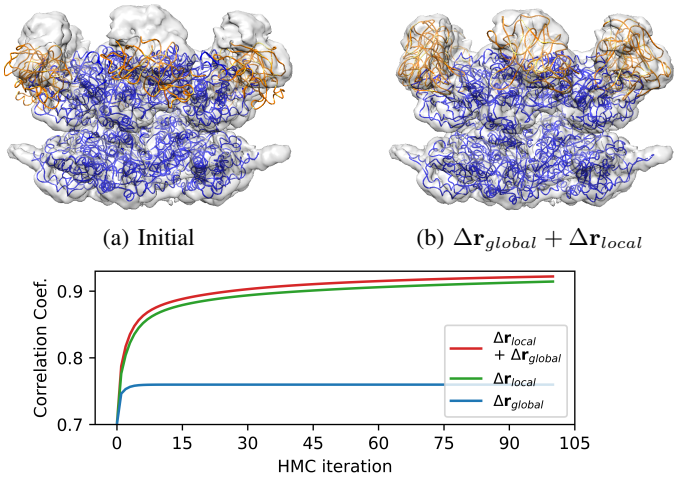
In this article, we proposed a method for flexible fitting of atomic structures into cryo-EM maps that combines global and local atomic displacements to speed up the fitting. Global displacements are performed with normal modes while local displacements are the three Cartesian coordinate displacements



(a) Initial (b) $\Delta\mathbf{r}_{global} + \Delta\mathbf{r}_{local}$

(c) Evolution of CC

Fig. 1: Fitting of the PDB-*5ftm* CA-atom structure of the p97 ATPase "down" conformation into a synthetic map using three approaches (the details on the map synthesis are in the text). (a,b) Superposition of the target map (transparent gray) with (a) the initial structure and (b) the fitted structure using the method proposed here that combines global and local fitting (100 HMC iterations). (c) Evolution of the CC (average of 10 HMC runs over 100 iterations) for the three fitting approaches.



(a) Initial (b) $\Delta\mathbf{r}_{global} + \Delta\mathbf{r}_{local}$

(c) Evolution of CC

Fig. 2: Ftting of the PDB-*5ftm* CA-atom structure of the p97 ATPase "down" conformation into the experimental *EMD-3299* map of the "up" conformation using three approaches. (a,b) Superposition of the target map (transparent gray) with (a) the initial structure and (b) the fitted structure using the method proposed here that combines global and local fitting (100 HMC iterations). (c) Evolution of the CC (average of 10 HMC runs over 100 iterations) for the three fitting approaches.

of each atom. The model is described in the Bayesian formalism and sampled using Hamiltonian Monte-Carlo sampler. To the best of our knowledge, this is the first combined Bayesian and normal mode fitting method. We demonstrated its performance with synthetic and experimental data. We showed that it is faster and achieves better fitting precision than the classical method that does not take into account normal modes. In the future, we will perform tests with experimental cryo-EM maps of other biomolecular complexes, investigating particularly the cases where normal modes may be more critical to use to accelerate fitting. The software will be available as open-source after it is tested with other data.

Acknowledgment

We acknowledge the support of the French National Research Agency — ANR (ANR-19-CE11-0008-01 and ANR-20-CE11-0020-03 to SJ) and cooperation between the CNRS and the University of Melbourne (CNRS PRC 2889 to SJ).

References

[1] W. Kühlbrandt. The resolution revolution. *Science*, 343(6178):1443–1444, 2014.

[2] J. Ognjenovic, R. Grisshammer, and S. Subramaniam. Frontiers in cryo electron microscopy of complex macromolecular assemblies. *Annual Review of Biomedical Engineering*, 21(1):395–415, 2019.

[3] S. Jonić. Computational methods for analyzing conformational variability of macromolecular complexes from cryo-electron microscopy images. *Current Opinion in Structural Biology*, 43:114 – 121, 2017.

[4] F. Tama, O. Miyashita, and C.L. Brooks III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-em. *Journal of Structural Biology*, 147(3):315 – 326, 2004.

[5] M. Habeck. Bayesian modeling of biomolecular assemblies with cryo-em maps. *Frontiers in molecular biosciences*, 4:15, 2017.

[6] O. Miyashita, C. Kobayashi, T. Mori, Y. Sugita, and F. Tama. Flexible fitting to cryo-em density map using ensemble molecular dynamics simulations. *Journal of computational chemistry*, 38(16):1447–1461, 2017.

[7] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*, volume 1. Elsevier, 2001.

[8] O. Miyashita and F. Tama. Hybrid methods for macromolecular modeling by molecular mechanics simulations with experimental data. *Integrative Structural Biology with Hybrid Methods*, pages 199–217, 2018.

[9] A. Brass, B.J. Pendleton, Y. Chen, and B. Robson. Hybrid monte carlo simulations theory and initial comparison with molecular dynamics. *Biopolymers: Original Research on Biomolecules*, 33(8):1307–1315, 1993.

[10] M.M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Physical review letters*, 77(9):1905, 1996.

[11] Q. Jin, C.O.S. Sorzano, J.M. De La Rosa-Trevín, J.R. Bilbao-Castro, R. Núñez-Ramírez, O. Llorca, F. Tama, and S. Jonić. Iterative elastic 3d-to-2d alignment method using normal modes for studying structural dynamics of large macromolecular complexes. *Structure*, 22(3):496–506, 2014.

[12] S. Duane, A.D. Kennedy, B.J. Pendleton, and D. Roweth. Hybrid monte carlo. *Physics letters B*, 195(2):216–222, 1987.

[13] R.M. Neal et al. Mcmc using hamiltonian dynamics. *Handbook of markov chain monte carlo*, 2(11):2, 2011.

[14] S. Banerjee, A. Bartesaghi, A. Merk, P. Rao, S.L. Bulfer, Y. Yan, N. Green, B. Mroczkowski, R.J. Neitz, P. Wipf, et al. 2.3 å resolution cryo-em structure of human p97 and mechanism of allosteric inhibition. *Science*, 351(6275):871–875, 2016.

[15] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[16] L-M Peng, G Ren, SL Dudarev, and MJ Whelan. Robust parameterization of elastic and absorptive electron atomic scattering factors. *Acta Crystallographica Section A: Foundations of Crystallography*, 52(2):257–276, 1996.