Published in final edited form as:

Methods. 2009 October ; 49(2): 174–180. doi:10.1016/j.ymeth.2009.04.005.

Molecular Dynamics Flexible Fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography

Leonardo G. Trabuco^{a,b,1}, Elizabeth Villa^{a,b,1,2}, Eduard Schreiner^a, Christopher B. Harrison^a, and Klaus Schulten^{*,a,c}

^aBeckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

^bCenter for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

^cDepartment of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Abstract

Hybrid computational methods for combining structural data from different sources and resolutions are becoming an essential part of structural biology, especially as the field moves toward the study of large macromolecular assemblies. We have developed the molecular dynamics flexible fitting (MDFF) method for combining high-resolution atomic structures with cryo-electron microscopy (cryo-EM) maps, that results in atomic models representing the conformational state captured by cryo-EM. The method has been applied successfully to the ribosome, a ribonucleoprotein complex responsible for protein synthesis. MDFF involves a molecular dynamics simulation in which a guiding potential, based on the cryo-EM map, is added to the standard force field. Forces proportional to the gradient of the density map guide an atomic structure, available from X-ray crystallography, into high-density regions of a cryo-EM map. In this paper we describe the necessary steps to set up, run, and analyze MDFF simulations and the software packages that implement the corresponding functionalities.

Keywords

MDFF; flexible fitting; cryo-EM; X-ray crystallography; NAMD; VMD; docking

1. Introduction

The field of structural biology is evolving rapidly, turning its focus from single proteins or protein domains to large macromolecular assemblies formed by many components working together. Most biological functions come about through the orchestrated interaction between many of such components. Recent advances have positioned cryo-electron microscopy (cryo-EM) single-particle reconstruction as a central structure determination technique for studying macromolecular assemblies. In some cases, cryo-EM is the only source of structural

*Corresponding author. E-mail address: kschulte@ks.uiuc.edu (K. Schulten).

¹These authors contributed equally to this work.

²Current address: Department of Structural Molecular Biology, Max Plank Institute for Biochemistry, 82152 Martinsried, Germany

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

information; however, it is increasingly common to have low- to medium-resolution cryo-EM reconstructions of macromolecular complexes for which high-resolution structures of the individual components are available. Furthermore, cryo-EM can provide images of macromolecular assemblies at different functional states usually elusive to X-ray crystallography. A notable example is the ribosome, for which cryo-EM has provided important insights into its function [1].

Computational methods for fitting atomic structures into cryo-EM maps bridge the resolution gap between cryo-EM and X-ray crystallography, allowing for interpretation of cryo-EM data beyond the nominal resolution. Moreover, structures obtained by fitting methods are sometimes the only source of atomic models of macromolecular assemblies. Attempts to fit crystal structures into cryo-EM density maps have been pursued for more than two decades. For low-resolution EM maps, locations of proteins inside a complex can be inferred by their overall shape compared to the density by simple visual inspection, or by systematic rigid-body docking, where an exhaustive search in six dimensions (three translational and three rotational degrees of freedom) is performed to find the best position and orientation of a given molecule inside the density. Rigid-body docking is a well-established technique with several software packages available. A comprehensive comparison of rigid-body docking methods, as well as a discussion on flexible fitting techniques, can be found in [2].

The advent of subnanometer-resolution structures stemming from single-particle cryo-EM studies has boosted the development of flexible fitting techniques. Unlike rigid-body docking, flexible fitting methods allow for conformational changes of the atomic structure in order to better represent the conformational state captured by the EM density. Early attempts to flexibly fit atomic structures into EM maps include reciprocal space refinement described by Mendelson and Morris [3]. A commonly used flexible fitting approach consists of dividing the macromolecule into rigid pieces which are then rigid-body docked and optimized with the real-space refinement method [4,5,6]. The Situs package implements a flexible fitting method based on vector quantization, where coarse-grained representations of both the atomic structure and the density map are used to drive the fitting using molecular mechanics refinement [7,8] or spatial interpolation [9]. Another popular approach is to use a linear combination of the low-frequency normal modes to deform the atomic structure, optimizing its correlation to the density map [10,11]. Bioinformatic analyses have also been used for flexible fitting, including iterative comparative modeling [12] and use of structural variability in protein superfamilies [13]. Other recently proposed approaches include identifying structures generated by molecular dynamics (MD) that correspond to the conformation in the EM map [14]; use of a deformable elastic network model with restraints imposed by the EM map [15]; a Gaussian network model combined with an empirical potential derived from the map [16]; an empirical potential for both the atomic structure and the EM map combined with a Monte Carlo search [17]; MD simulations with added forces proportional to the gradient of the cross-correlation coefficient between the atomic structure and the density map [18]; constrained geometric simulations based on rigidity analysis [19]; and damped-dynamics flexible fitting, in which the atomic structure is attracted toward incompletely occupied regions of the EM map [20].

Our approach consists in applying forces proportional to the gradient of the EM density map in MD simulations, driving the atomic structure into the high-density regions resulting in conformations representing the functional state captured by the experimental data [21]. The MD force field describing the interactions between the individual components at the atomic level ensures that the resulting structure is physically sound. The molecular dynamics flexible fitting (MDFF) method has been successfully applied to the study of the ribosome [21,22,23] and protein-induced membrane curvature [24,25]. As an illustration of the method, Fig. 1 shows the a tRNA from a ribosome-free crystal structure flexibly fitted into a cryo-EM density map of the tRNA being presented to the ribosome by EF-Tu during decoding. For details on the

derivation of the method, as well as on the application of MDFF to several test cases for validation purposes, the reader is referred to [21]. Here, we describe in detail the practical steps to be followed in the application of MDFF and in the analysis of the results obtained.

2. The MDFF method

The MDFF method is an extension of the molecular dynamics simulation technique, in which each atom is represented by a point particle with assigned mass and charge. The interaction between every pair of atoms is described by the force field – an empirical potential energy function U_{MD} . Based on the forces calculated from U_{MD} , the position and velocity of every atom are propagated in time according to classical mechanics, providing a dynamical description of the system. For details on the MD technique, the reader is referred to [26]. A recent review discussing MD simulation of nucleic acids can be found in an earlier issue of Methods [27].

MDFF builds on the MD method by extending U_{MD} by two additional terms. The first term takes into account the experimental information. The cryo-EM map is converted into a potential energy function, U_{EM} , that drives the atoms into the cryo-EM density, while the MD potential, U_{MD} , preserves the correct physical parameters of the model (e.g., no high-energy prohibitive structural features such as long covalent bonds or overlapping atoms are introduced). The second term, U_{SS} , maintains the integrity of secondary structure elements.

The preparation of MDFF simulations and the analysis of the resulting trajectories are performed using the `mdff` plugin of the visualization and analysis software VMD [28]. The MDFF simulations are conducted with the software package NAMD [29].

Section 2.1 describes the necessary data to perform MDFF simulations, namely a cryo-EM map and atomic models of the composing elements represented by the EM data. Section 2.2 details the form of the potential energy term U_{EM} derived from the cryo-EM density map that drives atoms into high-density regions. The atomic models have to be prepared for MD simulations, as described in Section 2.3.1. Section 2.3.2 describes the U_{SS} term aimed to prevent so-called overfitting, i.e., unphysical deformations of the atomic structure that result in a better fit to the cryo-EM data. A number of simulation parameters can be tuned in an MDFF simulation in order to achieve a better fit; the MDFF simulation protocol is described in detail in Section 2.4. The simulation can be performed in several steps, which is particularly useful when working with nucleoprotein complexes, as discussed in Section 2.4.3.

2.1. Data required for starting MDFF simulations

To carry out an MDFF simulation, two types of input data are required: a cryo-EM map of the complex under study, and atomic models of the components to be fitted into it. The source of the initial atomic models can be any structural technique such as X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, or a modeling technique such as *ab initio* or homology modeling. It is important to note that any errors contained in the initial atomic models may be propagated into the atomic models obtained through MDFF. In the same way, errors in the cryo-EM data arising from the optics, sample, numerical imprecision, small number of particles, and others, will be included in the potential energy function driving the fit and, thus, will affect the final model. It is therefore crucial to keep in mind the nature of the errors contained in the data when interpreting the results. For more information on cryo-EM data, the user is referred to [30]; a brief discussion is also included in the supplemental material of [21].

2.2. Converting the cryo-EM density map into a potential

The VMD plugin `mdff` can be used to generate a potential energy function U_{EM} defined on a 3-D grid based on the cryo-EM density map

$$U_{EM}(\mathbf{R}) = \sum_j \omega_j V_{EM}(\mathbf{r}_j), \quad (1)$$

where

$$V_{EM}(\mathbf{r}) = \begin{cases} \xi \left[1 - \frac{\Phi(\mathbf{r}) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}} \right] & \text{if } \Phi(\mathbf{r}) \geq \Phi_{thr}, \\ \xi & \text{if } \Phi(\mathbf{r}) < \Phi_{thr}. \end{cases} \quad (2)$$

Note that the Coulomb potential, $\Phi(\mathbf{r})$, represented by the cryo-EM data, as well as its maximum value, Φ_{max} , are clamped below the threshold value Φ_{thr} to remove the solvent contribution. The choice of Φ_{thr} is simplified by the analysis of the EM density histogram with the `mdff` plugin. Generally, cryo-EM maps will display a large density peak corresponding to the solvent; a threshold value at or above the solvent peak should be chosen, yielding a flat potential in the solvent regions (see [21] for a detailed discussion). The scaling factor ξ can be used to uniformly adjust the strength of the influence of the cryo-EM map on the molecular system. In addition, $V_{EM}(\mathbf{r})$ has a weight ω_j for each atom j present at position $\mathbf{r}_j$. Generally, ω_j is set to the atomic mass; this weighting avoids strong differences in the acceleration of atoms due to mass disparities, ensuring stability of the simulation. This choice of weighting factor is also in line with the rough correspondence between the mass of the atoms and their density in a cryo-EM map.

The potential energy function U_{EM} defined by Eqs. (1, 2) is incorporated into an MD simulation using NAMD's `gridForces` feature [31]. The `gridForces` feature allows an arbitrary potential defined on a 3-D grid (in DX file format) to be added to any MD simulation. The gradient of the potential is calculated by finite difference methods and forces are applied to each atom depending on its position on the grid using an interpolation scheme. In the case of MDFF, for each atom i in the system, the resulting force is given by

$$\mathbf{f}_i^{EM} = - \frac{\partial}{\partial \mathbf{r}_i} U_{EM}(\mathbf{R}) = - \omega_i \frac{\partial}{\partial \mathbf{r}_i} V_{EM}(\mathbf{r}_i). \quad (3)$$

The force $\mathbf{f}_i^{EM}$ can be tuned via the scaling factor ξ (Eq. (2)), which is the same for all atoms, and the weight ω_i , which can be defined on a per-atom basis. Fig. 2 shows a cryo-EM map of the ribosome and a 2-D slice of a region of V_{EM} illustrating how the calculated forces drive the atomic model into high-density regions of the EM map, which correspond to low-energy regions of the potential.

VMD provides many capabilities for visualization and manipulation of volumetric data. Currently, maps can be manipulated with the `volutil` plugin; inclusion of map manipulation functions in the core code of VMD with added support for graphics processing unit (GPU) acceleration is under way. Examples of map manipulations that can be performed with `volutil` include adding, subtracting, multiplying, averaging, filtering, cropping, down-sampling, supersampling, scaling, and clamping, among others. All map operations are

performed in Cartesian space and do not rely on the native grid of the map. This approach makes manipulations like combining maps represented on grids with different dimensions and spacing possible; the resampling of the maps to a common grid is transparent to the user. VMD can read all file formats commonly used in electron microscopy, and currently can write DX files, supported by NAMD `gridForces`; Situs files, useful for rigid-body docking with the Situs software [32]; and the commonly used CCP4 format. Even though the intended use of the Situs package within the MDFF workflow is limited to rigid-body docking, we note that Situs also provides map manipulation capabilities and flexible fitting based on vector quantization [7,8,9].

2.3. Structure preparation

Since MDFF is an extension of the MD simulation method, the steps required for setting up a regular MD simulation are also required by MDFF. The general requirements for setting up an MD simulation are described in Section 2.3.1. The additional potential energy term U_{SS} , introduced in MDFF to preserve the secondary structure of proteins and nucleic acids, is described in Section 2.3.2.

2.3.1. Building an all-atom structure—As for any all-atom MD simulation, a structure containing all the atoms, including hydrogens, is required for MDFF. Such a structure, together with the associated MD force field, determines the connectivity (bonds, angles, and dihedrals), and all the interactions in the system. The VMD plugin `autopsf` provides an easy and automated way to build an all-atom structure for MD simulation. For a thorough introduction on how to set up a system for MD simulation the reader is referred to the NAMD tutorial¹.

For explicit solvent simulations, the structure must be placed in a water box and typically ions are added to neutralize the total charge of the system and to mimic the desired ionic strength. For RNA simulations, it is particularly important to accurately place counterions at energetically favorable locations in order to stabilize the structure [33]; to that end, the highly efficient VMD plugin `cionize` may be employed [34]. Proper solvation of the system can be achieved in several steps. It is advisable to complete the octahedral first solvation shell of Mg^{2+} ions prior to the simulation (see [33] for details). Cavities and crevices can be solvated using the DOWSER program [35]; the VMD plugin `dowser` provides a graphical user interface to DOWSER, and extends the latter to support RNA systems. A water shell can then be constructed around the macromolecule using the Solvate program² [36]. Finally, the VMD plugin `solvate` can be used to insert the system into a water box. The desired ionic strength can be adjusted using the VMD plugin `autoionize`.

Once an all-atom structure is ready, the VMD plugin `mdff` can be used to generate a NAMD configuration file and to define the per-atom force scaling for the external potential derived from the EM map (Eq. (3)) needed for MDFF.

2.3.2. Secondary structure restraints—In order to avoid overfitting during the flexible fitting procedure, it is suitable to use restraints that preserve secondary structure elements in proteins and nucleic acids. The definition and strength of secondary structure restraints is facilitated by the VMD plugin `ssrestraints`. NAMD can then use the defined restraints through its `extraBonds` feature, which allows a user to restrain internal coordinates such as bonds, angles, dihedrals and impropers in an MD simulation.

¹<http://www.ks.uiuc.edu/Training/Tutorials/>

²<http://www.mpibpc.mpg.de/home/grubmueller/downloads/solvate/index.html>

For proteins, `ssrestraints` defines extra φ and ψ dihedral angles, with the equilibrium values taken either from the initial atomic structure or from ideal secondary structure elements. By default, only amino acid residues in helices or β sheets are considered for restraints, but any set of residues can be selected.

For nucleic acids, the VMD plugin `rnaview` provides an interface to the software package RNAView [37], currently used by `ssrestraints` to identify and classify base pairs. It is worth noting that `ssrestraints` can be easily extended to support other packages for analysis of nucleic acid structure. RNA base pairs can be classified according to the interaction edges involved, i.e., Watson-Crick (WC), Hoogsteen (H), sugar (S), or “C-H” edges (Fig. 3A). The following RNA base pairs are selected by `ssrestraints`: WC/WC, WC/H, WC/S, H/H, H/S, and stacked. For all of these six types of base pairs, restraints are applied to the seven dihedral angles (α , β , γ , δ , ϵ , ζ , and χ) and two interatomic distances (d_1 and d_2) depicted in Fig. 3B, the latter to preserve the planarity of the base pair.

The potential energy term U_{ss} has the form

$$U_{ss} = \sum_{\mu} k_{\mu} (X_{\mu} - X_{\mu}^0)^2, \quad (4)$$

where μ denotes the restrained internal coordinate, i.e., protein dihedral angles φ and ψ , RNA dihedral angles α , β , γ , δ , ϵ , ζ , and χ , and RNA distances d_1 and d_2 ; X_{μ} is the instantaneous value of this coordinate. By default, the equilibrium values X_{μ}^0 are set to the values from the initial atomic structure, but can also be automatically set to ideal values or manually set to any user-defined value. Secondary structure restraints are applied in MDFF in order to conservatively prevent overfitting. The user may gradually decrease the secondary structure restraints along the MDFF simulation, in particular after large-scale motions have been accomplished and local structural refinement is desired, as exemplified in the multistep protocol described in Section 2.4.3. Moreover, through the use of the `extrabonds` feature, other internal coordinates of the structure can be restrained, e.g., to enforce rigidity or proximity of certain subdomains.

2.4. MDFF simulation protocol

In this section, we discuss in detail the simulation protocol used by MDFF. First, we describe general MD simulation parameters used, such as force field, temperature control, and integration scheme (Section 2.4.1). Then, we address the main parameters in an MDFF simulation that can be tuned to achieve better results (Section 2.4.2). Finally, we give one concrete example of a multistep MDFF protocol that was applied to experimental cryo-EM maps of the ribosome (Section 2.4.3).

2.4.1. MD simulation protocol—MDFF simulations are performed using NAMD [29], and thus any MD force field or simulation parameter supported by NAMD can be used in MDFF. Thus far, we have employed the CHARMM27 force field with the CMAP correction [38,39,40,41]. MDFF simulations can be performed with explicit solvent [24,25]; however, for large macromolecular complexes, one may wish to omit the solvent and use an overall dielectric constant for Coulombic forces to decrease the computational cost [21,22,23]. A natural extension, which is under development, is the use of an implicit solvent model. MDFF simulations commonly make use of a multiple time-stepping integration scheme, with bonded interactions calculated every 1 fs and nonbonded interactions calculated every 2 fs. Other simulation parameters generally used within MDFF are a cut-off distance of 10 Å for the

nonbonded interactions and a temperature of 300K maintained using a Langevin thermostat [42] coupled to all heavy atoms with a damping coefficient of 5 ps^{-1} .

2.4.2. Adjustable MDFF parameters—There are four main parameters in MDFF that can be adjusted in order to obtain a better fit. First, in some cases it has proved effective to perform an initial MDFF simulation with strong *secondary structure restraints* (see Section 2.3.2), or even with equivalent restraints applied to all residues, followed by a subsequent MDFF simulation with weaker restraints. The strong restraints render the structure relatively rigid, preventing structural distortions during the initial phase of the fitting when the movements of the atomic model may be large. The weak restraints in the subsequent step allow for a better refinement of the structure (see Section 2.4.3 for a concrete example). Second, increasing the *force scaling* (ξ in Eq. (2)) will generally improve the fit, although stronger forces may lead to structural distortions (overfitting); furthermore, very strong forces may render the simulation unstable. A typical value of ξ is 0.3 kcal/mol, but a value around 1.0 kcal/mol is appropriate for fitting low-occupancy ligands. Third, the *input map* may also be manipulated in different ways, such as by low-pass filtering it in the initial phase in order to smooth the potential energy function, allowing the structure to undergo large conformational changes while preventing the system from becoming trapped in a local minimum. The user may also delete a density region, which is useful when fitting ribonucleoprotein complexes, as described in the next section. Finally, *temperature* may be temporarily increased in order to overcome energy barriers. Simulated annealing was successfully used in the context of a number of flexible fitting methods [43,15,17,16] and is another natural extension to MDFF currently under development.

2.4.3. MDFF multistep protocol applied to the ribosome—A macromolecular complex such as the ribosome imposes several significant challenges to flexible fitting methods, namely:

1. two thirds of its mass is composed of RNA, and modeling techniques designed for RNA are relatively immature compared to those for proteins;
2. the cryo-EM density corresponding to RNA is significantly higher than for proteins due to their difference in molecular weight;
3. ligands such as tRNAs and translation factors have typically occupancy lower than 100%.

In order to overcome these challenges, we designed secondary structure restraints for RNA (Section 2.3.2), as well as a multistep flexible fitting protocol, which was applied to cryo-EM maps of the ribosome (see Fig. 4). Each MDFF step is extended until convergence is reached (see Section 3 and Fig. 5). The initial atomic model used is described in detail elsewhere [21].

Before starting MDFF (*step 0*), the ribosome is rigid-body docked into the cryo-EM map using an FFT-accelerated 6-D exhaustive search, as implemented in the *colores* program from the Situs package [32]. *MDFF step 1* aims to obtain a good fit for the RNA components, while the protein components are kept relatively rigid by imposing strong restraints. In particular, in the first MDFF step, the ribosome is flexibly fitted using $\xi = 0.3 \text{ kcal/mol}$, $k_{\mu,\text{RNA}} = 200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$ or $200 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for angles and distances, respectively, and $k_{\mu,\text{protein}} = 400 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. Additionally, ϕ and ψ angles of the amino acid residues not included in secondary structure restraints are restrained with $k_{\mu,\text{protein}} = 200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$. The restraints preserve the secondary structure and the overall shape of the proteins, while the RNA, with higher molecular weight, fills its corresponding density. In *MDFF step 2*, the cryo-EM density corresponding to RNA is deleted. The new map Φ' is given by

$$\Phi'(\mathbf{r}) = \left(1 - \frac{\Phi_{\text{sim}}(\mathbf{r})}{\max(\Phi_{\text{sim}}(\mathbf{r}))}\right) \Phi(\mathbf{r}), \quad (5)$$

where $\Phi_{\text{sim}}(\mathbf{r})$ is the simulated map generated from the fitted RNA structure with the same resolution as the experimental cryo-EM map $\Phi(\mathbf{r})$; the generation of simulated maps is described in the next section. During the fitting in the second MDFF step, the RNA atomic structure is restrained to its current Cartesian coordinates with a relatively large force constant of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, while proteins are restrained identically to *MDFF step 1*. In *MDFF step 3*, the structure of the proteins is refined by reducing the restraints to $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$, restraining only residues in helices and β sheets. Finally, in *MDFF step 4*, the atomic structures of the remaining ligands (tRNAs and translation factors) are added by rigid-body docking and subsequently refined with MDFF. For this final step, the original potential derived from the experimental cryo-EM map is used (before RNA density deletion), ribosomal proteins are positionally restrained, and the positional restraints on the RNA structure are removed. The secondary structure restraints described above are applied to all proteins, RNA, and newly introduced ligands.

3. Analysis of MDFF results

Convergence of an MDFF simulation (or an MDFF step in a multistep protocol) can be assessed by monitoring over time the root mean square deviation (RMSD) with respect to the original structure, or the cross-correlation coefficient (CCC) between the atomic structure and the original cryo-EM density map. The RMSD is given by $\langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle)^2 \rangle^{1/2}$, where $\mathbf{r}_i$ is the coordinate vector of atom i and the averages are taken over a set of atoms in the atomic structure, such as all backbone atoms. Fig. 5A presents an example of an RMSD plot over a multistep MDFF trajectory. Notice how the RMSD always reaches a plateau before the next MDFF step is initiated.

In order to calculate a CCC between an atomic structure and a cryo-EM map, first a simulated map has to be generated from the atomic structure. This can be accomplished by interpolating the atomic numbers of all atoms onto a 3-D grid, and subsequently applying a low-pass (Gaussian) filter. The target resolution of the simulated map is determined by the width (2σ) of the Gaussian filter. Once a simulated map has been created with resolution matching the one from the cryo-EM reconstruction, the CCC can be calculated by

$$\rho = \frac{\langle (S - \langle S \rangle)(E - \langle E \rangle) \rangle}{\sigma_S \sigma_E}, \quad (6)$$

where $\langle S \rangle$ and $\langle E \rangle$ are the average voxel values from the simulated and original experimental maps, respectively, σ_S and σ_E are the corresponding standard deviations, and $\rho \in [-1, 1]$. The averages in Eq. (6) are taken over the intersection of the simulated and experimental 3-D grids.

The CCC as defined in Eq. (6) is a global measure of the similarity between an atomic structure and a cryo-EM map. Even though this measure is widely used in the literature, it suffers from a serious problem: a considerable fraction of the statistical sample corresponds to the solvent, not to the macromolecule. In fact, the value of the CCC can be made arbitrarily close to the optimal value by defining a large enough 3-D grid, i.e., by padding the space with zero density. To overcome this problem, it is appropriate to define a local CCC in which only the volume encompassing the macromolecule is considered. A simulated map is created from the atomic

structure, from which a volume can be defined by all voxels with value above a chosen threshold. The intersection between this volume and the experimental cryo-EM map defines the sample space for the local CCC calculation. To contrast with the local CCC, we refer to the CCC which considers the entire map as the global CCC. Fig. 5B presents a plot of both the global and the local CCC over a multistep MDFF trajectory. As for the RMSD, the CCC always levels off before the next MDFF step is initiated.

The VMD plugin `mdff` allows the user to monitor over time the RMSD and the global or local CCC. Notice that neither the RMSD nor the CCC are used to drive an MDFF simulation; they are simply measures used to assess the progress and quality of the fit. Not using a global measure to drive the fit allows MDFF to be used for large systems, and avoids pitfalls stemming from the variability in resolution in different regions of the EM map.

After convergence is reached in the last step of MDFF, a representative set of structures, all conforming to the EM data, is obtained. While any of these structures represents a plausible structure in the ensemble imaged by cryo-EM, interpretation of the EM map beyond the nominal resolution can only be achieved by considering different conformations sampled by this representative set. Structures obtained with MDFF can greatly aid in the interpretation of the 3-D reconstructions; however, all the limitations should be kept in mind, in particular the experimental resolution and the quality of the starting models. It is therefore advisable to always inspect the density together with the atomic models when interpreting the data.

4. Concluding remarks

MDFF is a flexible fitting method particularly well-suited for the study of macromolecular assemblies. Since MDFF is implemented in NAMD [29], a highly scalable MD software with a memory-efficient implementation of its `gridForces` feature, it can be used to fit large systems in their entirety in the same simulation [21,22,23,24,25]. MDFF has also been successfully applied to RNA systems [21,22,23]. The setup and analysis of MDFF simulations is achieved with the VMD plugin `mdff`, while the simulations are run using NAMD. Alternatively, MDFF can be combined with interactive molecular dynamics [44,45], allowing a user to apply extra forces using, e.g., a haptic device.

Future extensions of MDFF will include use of implicit solvent models, simulated annealing, enhanced sampling techniques such as normal mode-accelerated dynamics, more sophisticated secondary structure restraints, and others. Explicit-solvent MDFF applied to the ribosome suggests that secondary structure restraints may not be required in this case, though results are still preliminary.

Due to its flexibility and scalability, NAMD is ideally suited to serve as the optimal platform to combine the most desirable features of the many flexible fitting techniques recently developed, with the aim to obtain the best possible atomic models from cryo-EM data. Furthermore, we foresee that the ongoing implementation of GPU acceleration technologies into NAMD will allow MDFF to be performed interactively in real time. It is worth noting that other sources of structural information can already be included in the MDFF modeling effort, such as restraints derived from nuclear magnetic resonance spectroscopy, Förster resonance energy transfer, cross-linking and immuno-labeling, small-angle X-ray scattering, among others.

Once an atomic model of a macromolecular complex in a functional state has been obtained by means of MDFF, further MD simulations can be carried out to test hypotheses on the function of the system under study [23], thus completing a structural dynamics description of the macromolecular complex.

Acknowledgments

The authors would like to acknowledge Joachim Frank for continued guidance in the development of MDFF, and the Theoretical and Computational Biophysics Group for support in the development, testing, and application of MDFF. This work is supported by grants National Institutes of Health grant P41-RR05969 and National Science Foundation grant PHY0822613. Computer time was provided through a National Science Foundation Large Resources Allocation Committee grant MCA93S028.

References

1. Mitra K, Frank J. *Annu Rev Biophys Biomol Struct* 2006;35:299–317. [PubMed: 16689638]
2. Wriggers W, Chacón P. *Structure* 2001;9:779–788. [PubMed: 11566128]
3. Mendelson R, Morris EP. *Proc Natl Acad Sci USA* 1997;94:8533–8538. [PubMed: 9238011]
4. Chapman MS. *Acta Cryst A* 1995;51:69–80.
5. Chen LF, Blanc E, Chapman MS, Taylor KA. *J Struct Biol* 2001;133:221–232. [PubMed: 11472093]
6. Chen JZ, Fürst J, Chapman MS, Grigorieff N. *J Struct Biol* 2003;144:144–151. [PubMed: 14643217]
7. Wriggers W, Agrawal RK, Drew DL, McCammon A, Frank J. *Biophys J* 2000;79:1670–1678. [PubMed: 10969026]
8. Wriggers W, Birmanns S. *J Struct Biol* 2001;133:193–202. [PubMed: 11472090]
9. Rusu M, Birmanns S, Wriggers W. *Bioinformatics* 2008;24:2460–2466. [PubMed: 18757874]
10. Tama F, Miyashita O, Brooks CL III. *J Struct Biol* 2004;147:315–326. [PubMed: 15450300]
11. Suhre K, Navaza J, Sanejouand YH. *Acta Cryst D* 2006;62:1098–1100. [PubMed: 16929111]
12. Topf M, Baker ML, Marc-Renom MA, Chiu W, Sali A. *J Mol Biol* 2006;357:1655–1668. [PubMed: 16490207]
13. Velazquez-Muriel JA, Valle M, Santamaría-Pang A, Kakadiaris IA, Carazo JM. *Structure* 2006;14:1115–1126. [PubMed: 16843893]
14. Li W, Frank J. *Proc Natl Acad Sci USA* 2007;104:16540–16545. [PubMed: 17925437]
15. Schröder GF, Brunger AT, Levitt M. *Structure* 2007;15:1630–1641. [PubMed: 18073112]
16. Tan RKZ, Devkota B, Harvey SC. *J Struct Biol* 2008;163:163–174. [PubMed: 18572416]
17. Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A. *Structure* 2008;16:295–307. [PubMed: 18275820]
18. Orzechowski M, Tama F. *Biophys J* 2008;95(12):5692–5705. [PubMed: 18849406]
19. Jolley CC, Wells SA, Fromme P, Thorpe MF. *Biophys J* 2008;94:1613–1621. [PubMed: 17993504]
20. Kovacs JA, Yeager M, Abagyan R. *Biophys J* 2008;95:3192–3207. [PubMed: 18586844]
21. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. *Structure* 2008;16:673–683. [PubMed: 18462672]
22. Villa E, Sengupta J, Trabuco LG, LeBarron J, Baxter WT, Shaikh TR, Grassucci RA, Nissen P, Ehrenberg M, Schulten K, Frank J. *Proc Natl Acad Sci USA* 2009;106:1063–1068. [PubMed: 19122150]
23. Gumbart J, Trabuco LG, Schreiner E, Villa E, Schulten K. Submitted
24. Sener MK, Hsin J, Trabuco LG, Villa E, Qian P, Hunter CN, Schulten K. *Chem Phys* 2009;357:188–197.
25. Hsin J, Gumbart J, Trabuco LG, Villa E, Quan P, Hunter CN, Bullough PA, Schulten K. Submitted
26. Frenkel, D.; Smit, B. *Understanding Molecular Simulation From Algorithms to Applications*. Academic Press; California: 2002.
27. Hashem Y, Auffinger P. doi:10.1016/j.ymeth.2008.09.020
28. Humphrey W, Dalke A, Schulten K. *J Mol Graphics* 1996;14:33–38.
29. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. *J Comp Chem* 2005;26:1781–1802. [PubMed: 16222654]
30. Frank, J. *Three-dimensional electron microscopy of macromolecular assemblies*. Oxford University Press; New York: 2006.
31. Wells D, Abramkina V, Aksimentiev A. *J Chem Phys* 2007;127:125101–125110. [PubMed: 17902937]

32. Wriggers W, Milligan RA, McCammon JA. *J Struct Biol* 1999;125:185–195. [PubMed: 10222274]
33. Eargle J, Black AA, Sethi A, Trabuco LG, Luthey-Schulten Z. *J Mol Biol* 2008;377:1382–1405. [PubMed: 18336835]
34. Stone JE, Phillips JC, Freddolino PL, Hardy DJ, Trabuco LG, Schulten K. *J Comp Chem* 2007;28:2618–2640. [PubMed: 17894371]
35. Zhang L, Hermans J. *Proteins Struct Func Gen* 1996;24:433–438.
36. Grubmüller H, Heymann B, Tavan P. *Science* 1996;271:997–999. [PubMed: 8584939]
37. Yang H, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E. *Nucl Acids Res* 2003;31:3450–3460. [PubMed: 12824344]
38. MacKerell A Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck J, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher IWE, Roux B, Schlenkrich M, Smith J, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. *J Phys Chem B* 1998;102:3586–3616.
39. Foloppe N, MacKerell AD Jr. *J Comp Chem* 2000;21:86–104.
40. MacKerell AD Jr, Banavali NK. *J Comp Chem* 2000;21(2):105–120.
41. MacKerell AD Jr, Feig M, Brooks CL III. *J Comp Chem* 2004;25:1400–1415. [PubMed: 15185334]
42. Brünger AT, Brooks CL III, Karplus M. *Chem Phys Lett* 1984;105(5):495–498.
43. Mears JA, Sharma MR, Gutell RR, McCook AS, Richardson PE, Caulfield TR, Agrawal RK, Harvey SC. *J Mol Biol* 2006;358:193–212. [PubMed: 16510155]
44. Stone, J.; Gullingsrud, J.; Grayson, P.; Schulten, K. A system for interactive molecular dynamics simulation. In: Hughes, JF.; Séquin, CH., editors. 2001 ACM Symposium on Interactive 3D Graphics; New York. 2001. p. 191-194.
45. Grayson P, Tajkhorshid E, Schulten K. *Biophys J* 2003;85:36–48. [PubMed: 12829462]
46. LeBarron J, Grassucci RA, Shaikh TR, Baxter W, Sengupta J, Frank J. *J Struct Biol* 2008;164:24–32. [PubMed: 18606549]
47. Leontis NB, Westhof E. *Quart Rev Biophys* 1998;31:399–455.

Figure 1.

Flexible fitting of a tRNA into a cryo-EM map. An atomic model of the ribosome was obtained by applying MDFF to a 6.7-Å cryo-EM map [21,22]. For illustration purposes, we show only a single tRNA and the corresponding density segmented from the cryo-EM map. The top panel contains the initial tRNA structure after rigid-body docking, whereas the bottom panel contains the fitted structure obtained with MDFF.

Figure 2. Potential derived from the cryo-EM map. (A) Surface representation of a 6.7-Å cryo-EM map of the ribosome [46,22]. (B) Section of a 2-D slice of the potential generated from the map shown in A, represented as a contour plot. Arrows represent forces calculated using Eq. (3) for each grid point, driving the atomic structure toward high-density regions. In an actual MDFF simulation, the force applied to each atom is calculated by interpolating between the on-lattice forces depicted in the figure.

Figure 3. Restraints applied to base-paired RNA residues. (A) RNA interaction edges for both pyrimidines (cytosine is shown on the left) and purines (guanine is shown on the right), according to [47]; (B) Dihedral angles and the two interatomic distances to which restraints are applied (see Section 2.3.2).

Figure 4. Multistep flexible fitting protocol applied to ribosome cryo-EM maps [21] (see Section 2.4.3 for details).

Figure 5.

Convergence of MDFF simulation steps. For each MDFF step, both the (A) RMSD (shown in blue) and (B) global and local CCC (shown in red and black, respectively) reach a plateau. In the figure, the local CCC was calculated with a threshold of 0.2σ above the mean (see Section 3 for details). The insets show the corresponding full-scale plots. The multistep MDFF simulation presented corresponds to flexible fitting applied to a 6.7-Å cryo-EM map of the ribosome [21].