Accurate Flexible Fitting of High-Resolution Protein Structures into Cryo-Electron Microscopy Maps Using Coarse-Grained Pseudo-Energy Minimization

Wenjun Zheng*
Physics Department, University at Buffalo, Buffalo, New York

ABSTRACT Cryo-electron microscopy (cryo-EM) has been widely used to explore conformational states of large biomolecular assemblies. The detailed interpretation of cryo-EM data requires the flexible fitting of a known high-resolution protein structure into a low-resolution cryo-EM map. To this end, we have developed what we believe is a new method based on a two-bead-per-residue protein representation, and a modified form of the elastic network model that allows large-scale conformational changes while maintaining pseudobonds and secondary structures. Our method minimizes a pseudo-energy which linearly combines various terms of the modified elastic network model energy with a cryo-EM-fitting score and a collision energy that penalizes steric collisions. Unlike previous flexible fitting efforts using the lowest few normal modes, our method effectively utilizes all normal modes so that both global and local structural changes can be fully modeled. We have validated our method for a diverse set of 10 pairs of protein structures using simulated cryo-EM maps with a range of resolutions and in the absence/presence of random noise. We have shown that our method is both accurate and efficient compared with alternative techniques, and its performance is robust to the addition of random noise. Our method is also shown to be useful for the flexible fitting of three experimental cryo-EM maps.

INTRODUCTION

The biological functions of many biomolecules involve dynamic transitions between different conformational states. Given the difficulty of capturing all conformational states of large biomolecules by x-ray crystallography, medium- to low-resolution structure-determining techniques are increasingly used. In particular, cryo-electron microscopy (cryo-EM) has played a central role in probing conformational states of large biomolecular assemblies (1).

Because cryo-EM only obtains electron density maps with medium to low resolution, its interpretation requires the fitting of a known high-resolution structure of the same biomolecule into the density map. A variety of computational algorithms have been developed to perform automated fitting of cryo-EM maps (for a review, see (2)). The early approaches only considered rigid-body motions of biomolecules (3–9). More recently, thanks to improving resolution of cryo-EM data, cryo-EM-fitting techniques that incorporate biomolecular flexibility have been introduced. Partial account of flexibility was achieved by treating a biomolecule as a set of domains that were fitted independently as rigid bodies (10–14). However, such methods depend on a subjective and error-prone partition of a biomolecule into rigid domains and ignore coupled motions between domains which may be functionally important. A finer treatment of flexibility was achieved by partitioning a biomolecule into secondary structure elements which were treated as rigid bodies (15,16), or identifying rigid elements of a biomolecule using graph theory (17–19). The flexibility of protein domains was also modeled based on the principal component analysis of a superfamily of protein structures (20,21), or via comparative modeling (22).

The flexibility of biomolecules can be described by reduced or coarse-grained models, for example, using vector quantization (3,23). A variety of coarse-grained models (24) have been developed to simulate protein conformational dynamics with high efficiency. The elastic network model (ENM) (25–27) represents a protein structure as a network of C_α atoms with neighboring ones connected by springs with a uniform force constant (28). The ENM has been successfully used to assist the fitting of cryo-EM and x-ray data (29–31). In particular, ENM-based normal mode analysis has been widely utilized to flexibly fit high-resolution structures into low-resolution structural data (32–40), or satisfy a few pairwise distance constraints (41,42). Despite great success, the ENM-based flexible fitting methods are limited in accuracy because they usually only use a few low-frequency normal modes solved from ENM, which are less accurate for describing small local conformational changes (like rearrangement of helices inside a densely packed region) than large global ones (like domain motions) (27).

Molecular dynamics (MD) simulation is, in principle, able to describe the dynamics of biomolecules at both global and local scales, which makes it a method of choice for cryo-EM fitting (43). In earlier studies, all-atom MD simulation was used for real-space refinement, which assumed certain parts of a biomolecule are rigid and simultaneously optimized the fitting to cryo-EM data and the stereochemical properties (2,44,45). Recently, several MD-based methods have been introduced for cryo-EM fitting with full flexibility. The

Submitted September 3, 2010, and accepted for publication December 2, 2010.

*Correspondence: wjzheng@buffalo.edu

Editor: Gerhard Hummer.

0006-3495/11/01/0478/11 $2.00

doi: 10.1016/j.bpj.2010.12.3680

common strategy of these methods is to bias the MD simulation toward a conformation that optimally fits the cryo-EM data by using a biasing potential function (46–50). The application of these methods, however, has been limited by the high computational cost of running MD simulations for large biomolecular systems. To reduce computing cost, Grubisic et al. (51) developed an MD-based optimization method using a reduced protein representation and a simplified Gō potential function.

To achieve both accuracy and efficiency in the flexible fitting of cryo-EM maps, we propose a coarse-grained method with the following novel, to our knowledge, features:

First, it uses a two-bead-per-residue protein representation (one bead at the C_α atom, the other at the center of mass of side-chain atoms, see (52)), which enables the fitting of all-atom cryo-EM density maps with higher accuracy than the C_α-only representation.

Second, it is based on a modification of the ENM energy that combines the harmonic interactions for maintaining pseudo-bonds and secondary structures, and the anharmonic interactions between nonbonded beads that allow nonbonded residues to move apart readily. As a result, large global structural changes can be sampled without distorting local structures. Similarly, in another study, the ENM energy was modified using multiple force constants to allow both locally stiff and globally flexible structures (29).

Third, it minimizes a pseudo-energy which linearly combines various terms of the modified ENM energy with a cryo-EM-fitting score and a collision energy that penalizes steric collisions (see Methods). Our minimization-based approach has two advantages:

1. Unlike previous flexible fitting efforts using the lowest few normal modes (see (33,34)), our method effectively utilizes all normal modes so that both global and local structural changes can be fully modeled while overfitting remains under control.
2. It is efficiently implemented using the Newton-Raphson algorithm based on a sparse linear equation solver (53).

We have validated our method for a diverse set of 10 pairs of protein structures using simulated cryo-EM maps with a range of resolutions and in the absence or presence of random noise. We have shown that our method is both accurate and efficient compared with alternative techniques, and its performance is robust to the addition of random noise. Our method is also shown to work well for the flexible fitting of three experimental cryo-EM maps.

METHODS

Modified elastic network model (mENM) with two-bead-per-residue representation

A two-bead-per-residue coarse-grained model is constructed from the atomic coordinates of a protein crystal structure—one bead is at the C_α atom and the other is at the center of mass (CM) of the nonhydrogen side-chain atoms.

The potential energy of the original ENM (28) is

$$E_{ENM} = \frac{1}{2}\sum_{i=2}^{N}\sum_{j=1}^{i-1} C_{ij}\theta\left(R_c - d_{ij,0}\right)\left(d_{ij} - d_{ij,0}\right)^2, \quad (1)$$

where d_{ij} is the distance between bead i and j, and $d_{ij,0}$ is the value of d_{ij} given by the crystal structure, N is the total number of beads, $\theta(x)$ is the Heaviside function, R_c is a cutoff distance chosen to be 7 Å (it is lower than the 10 Å previously used for C_α-only ENM because of the use of a finer two-bead-per-residue representation), and C_{ij} is the force constant of the spring between bead i and j. Following Tekpinar and Zheng (54), two force constant values are used here: C_{ij} is set to 10 for pseudobonded beads (i.e., if bead i and j correspond to the C_α atoms of two bonded residues, or the C_α atom and the side-chain CM of the same residue), and 1 otherwise (the unit of C_{ij} can be arbitrarily chosen without changing the modeling results).

To allow nonbonded beads to move apart readily while maintaining pseudo-bonds and secondary structures, we have modified the ENM energy in Eq. 1 to the following form (named mENM energy):

$$\begin{aligned} E_{mENM} &= E_b + E_{SS} + E_{nb},\\ E_b &= \tfrac{1}{2}\sum_{\langle ij\rangle\in P_b} C_b\left(d_{ij} - d_{ij,0}\right)^2,\\ E_{SS} &= \tfrac{1}{2}\sum_{\langle ij\rangle\in P_{SS}} C_{SS}\left(d_{ij} - d_{ij,0}\right)^2,\\ E_{nb} &= \tfrac{1}{2}\sum_{\langle ij\rangle\notin P_b \,\&\, \langle ij\rangle\notin P_{SS}} C_{nb}\theta\left(R_c - d_{ij,0}\right)\times\frac{d_{ij,0}^2}{e^2}\left(1-\frac{d_{ij,0}^e}{d_{ij}^e}\right)^2, \end{aligned} \quad (2)$$

where E_b is the pseudobonded energy (P_b is the set of pseudobonded bead pairs, the bonded force constant $C_b = 10$); E_{SS} is the nonbonded energy which maintains the local structure of α-helices and β-strands (P_{SS} is the set of C_α atom pairs which are either in a helix with a sequential offset ≤ 4, or in a β-strand with a sequential offset ≤ 3, or forming hydrogen bonds between two adjacent parallel/antiparallel β-strands, with the associated force constant $C_{SS} = 1$); and E_{nb} is the remaining nonbonded energy with a new parameter $e = 6$, corresponding to the Lennard-Jones potential—it has a minimum at $d_{ij,0}$, saturates as d_{ij} goes to infinity, and diverges as d_{ij} approaches zero (the nonbonded force constant $C_{nb} = 1$). Therefore, unlike the harmonic potential in Eq. 1, the mENM energy allows two nonbonded beads to move apart at a finite energy cost.

The mENM energy in Eq. 2 can be expanded near a given conformation X_* to the second order,

$$E_{mENM}(X) \approx E_{mENM}(X_*) + \delta X^{\mathrm{T}} G + \frac{1}{2}\delta X^{\mathrm{T}} H \delta X, \quad (3)$$

where $\delta X = X - X_*$, $G = \nabla E_{mENM}|_{X=X_*}$ is the gradient of E_{mENM} at $X = X_*$, and H is the $3N\times 3N$ Hessian matrix comprised of the following 3×3 blocks,

$$H_{ij} = \begin{bmatrix} \left.\frac{\partial^2 E_{mENM}}{\partial x_i \partial x_j}\right|_{X=X_*} & \left.\frac{\partial^2 E_{mENM}}{\partial x_i \partial y_j}\right|_{X=X_*} & \left.\frac{\partial^2 E_{mENM}}{\partial x_i \partial z_j}\right|_{X=X_*} \\ \left.\frac{\partial^2 E_{mENM}}{\partial y_i \partial x_j}\right|_{X=X_*} & \left.\frac{\partial^2 E_{mENM}}{\partial y_i \partial y_j}\right|_{X=X_*} & \left.\frac{\partial^2 E_{mENM}}{\partial y_i \partial z_j}\right|_{X=X_*} \\ \left.\frac{\partial^2 E_{mENM}}{\partial z_i \partial x_j}\right|_{X=X_*} & \left.\frac{\partial^2 E_{mENM}}{\partial z_i \partial y_j}\right|_{X=X_*} & \left.\frac{\partial^2 E_{mENM}}{\partial z_i \partial z_j}\right|_{X=X_*} \end{bmatrix}, \quad (4)$$

where x_i, y_i, z_i (x_j, y_j, z_j) is the x, y, z component of the coordinate of bead i (j). The calculated gradient and Hessian matrix will be used in the flexible cryo-EM-fitting protocol based on the Newton-Raphson algorithm (see below).

Simulated cryo-EM map

Given a set of atomic coordinates, we calculate simulated cryo-EM maps by locating three-dimensional Gaussian functions on each atom and integrating these functions for each voxel to get the density function (see (33,34))

$$\rho(i,j,k) = \frac{1}{\left(\sqrt{2\pi/3}\sigma u\right)^3}\sum_{n=1}^{N}\int_{V_{ijk}} \exp\left\{-\frac{3}{2\sigma^2}\left[(x-x_n)^2 + (y-y_n)^2+(z-z_n)^2\right]\right\}dxdydz, \quad (5)$$

where σ is one-half the map resolution (23), u is the grid spacing (the edge-length of a cubic voxel), N is the number of atoms, and $V_{ijk} = u^3$ denotes the volume of a cubic voxel centered at $(x_i, y_j, z_k) = (i, j, k) \times u$, (x_n, y_n, z_n) is the coordinate of atom n.

Equation 5 can be further reduced to

$$\rho(i,j,k) = \sum_{n=1}^{N} f(x_i - x_n) f(y_j - y_n) f(z_k - z_n), \quad (6)$$

where

$$f(w) = \frac{1}{2u}\left[erf\left(\frac{w+u/2}{\sqrt{2\sigma^2/3}}\right) - erf\left(\frac{w-u/2}{\sqrt{2\sigma^2/3}}\right)\right],$$

and

$$\mathrm{erf}(w) = \frac{2}{\sqrt{\pi}}\int_0^w e^{-t^2}dt.$$

We note that the resolution parameter (2σ) used here does not exactly correspond to the resolution reported in cryo-EM experiments using a Fourier filter. However, previous studies with coarse-grained models have shown that the details of simulated cryo-EM maps do not affect the fitting performance (33,34).

Simulated cryo-EM maps are generated at four resolutions (5 Å, 10 Å, 15 Å, and 20 Å), and used as target maps in the flexible fitting. Among these resolution values, 10 Å represents a typical resolution for state-of-the-art cryo-EM maps, so it is used as the default resolution to demonstrate our method in comparison with alternative ones.

In previous studies, the grid spacing was usually set to 1 Å or 2 Å regardless of resolution (see (20,47)). In one study, the grid spacing was set to 3 Å and 4 Å for the cryo-EM maps with 10 Å and 20 Å resolution, respectively (33). In general, a finer grid allows a more accurate representation of cryo-EM maps at the expense of higher computing cost. To determine the grid spacing value with a proper balance between accuracy and efficiency, we have explored a range of grid spacing values from 2 Å to 5 Å for 10 Å resolution maps (see Table S1 in the Supporting Material). We have found that the quality of final fitted models depends only weakly on grid spacing, while the computing time is reduced significantly (by >5 times) as the grid spacing increases from 2 Å to 5 Å (see Table S1). Our finding agrees with a previous study which showed that optimal fitting can be obtained by choosing a grid spacing smaller than the length scale set by the map resolution (19). Therefore, we set the grid spacing to be one-half the map resolution, which allows significant reduction of computing time without losing model accuracy.

Simulated noise of cryo-EM map

To simulate the noise present in experimental cryo-EM maps, random Gaussian noise is added to the simulated maps. Following Jolley et al. (19), we assume that the noise level in a voxel is proportional to the squared root of the noise-free density value. An overall signal/noise ratio (SNR) is calculated after the noisy map is generated. Here we consider three noise levels with SNR set to 1, 2, and 5, respectively.

mENM-based flexible cryo-EM fitting

We start from two given inputs: the initial protein structure and the target cryo-EM map simulated from the target structure. Similar to Schroder et al. (30), we optimize a pseudo-energy which is a linear combination of the following components—the nonbonded mENM energy based on the initial structure (weighted by λ), the EM-fitting score for the target cryo-EM map (weighted by 1-λ), the mENM energy terms maintaining pseudo-bonds and secondary structures (weighted by 1), and the collision energy (weighted by 1, see below). The weight parameter $\lambda \in [0,1]$ controls the degree of data fitting (akin to the γ parameter in (30)): at $\lambda \sim 1$, the conformational search is restricted near the initial structure with poor fitting to the target cryo-EM map; at $\lambda \sim 0$, the influence of initial structure is too weak to maintain the structural integrity so the risk for overfitting is high. To find an optimal λ, we optimize the pseudo-energy progressively as λ decreases gradually from 1 to 0 until the EM-fitting score reaches a threshold value determined by the inaccuracy of our coarse-grained model or the level of noise (see below).

The total pseudo-energy function is defined as

$$E_{total} = \lambda E_{nb} + (1-\lambda)E_{map} + E_b + E_{SS} + E_{col}, \quad (7)$$

where E_{nb} is the nonbonded energy based on the initial structure (see Eq. 2), E_b is the pseudobonded energy (see Eq. 2), E_{SS} is the nonbonded energy which maintains the local structure of α-helices and β-strands (see Eq. 2), E_{col} is the collision energy between two nonbonded beads defined as

$$E_{col} = \frac{1}{2}\sum_{\langle ij\rangle \notin P_b \& \langle ij\rangle \notin P_{SS}} C_{col}\theta\left(R_{col} - d_{ij}\right)\left(d_{ij} - R_{col}\right)^2, \quad (8)$$

where the collision force constant $C_{col} = 10$, and R_{col} is the minimal distance between nonbonded beads in the initial structure (those bead pairs considered in E_b and E_{SS} are excluded from the summation) (following (54)). The addition of E_{col} penalizes steric collisions between residues whose C_α atoms or CM beads are within a distance of R_{col}. A similar soft-sphere energy term was used in another flexible fitting method (29).

E_{map} is the EM-fitting score defined as

$$E_{map} = 100 \times \frac{\sum_{i,j,k}[\rho_m(i,j,k) - \rho_t(i,j,k)]^2}{\sum_{i,j,k}[\rho_t(i,j,k)]^2}, \quad (9)$$

where ρ_m is the model density function calculated from the bead coordinates of a coarse-grained model (using Eq. 5, where each backbone atom is assigned the coordinate of the corresponding C_α atom, and each side-chain atom is assigned the coordinate of the corresponding CM bead), ρ_t is the target density function, and (i, j, k) is the voxel index. The denominator in Eq. 9 normalizes E_{map} to a value that only weakly depends on protein size (number of atoms) and map size (number of voxels). The weight 100 is chosen to provide proper driving force for fast fitting without causing overfitting.

For a series of gradually decreasing λ, the cryo-EM-fitted model is found by minimizing E_{total} in Eq. 7. We use the Newton-Raphson algorithm to solve $\nabla E_{total}(\lambda, X_{min}) = 0$ by using the following iterative procedure:

1. Initialization: set $n = 0$, $\lambda_0 = 1$, $X_{min,0} = X_1$, which represents the bead coordinates of the initial structure.
2. Multiply λ_n by $10^{-0.2}$.
3. For the conformation $X_{min,n}$, calculate the pseudo-energy $E_{total,n}$ using Eq. 7.
4. If $E_{total,n}$ fails to be lowered after 10 iterations, stop minimization and go to step 7.
5. Displace $X_{min,n}$ by the incremental displacement

$$\delta X_{min,n} = -\left(\lambda_n H_{nb} + (1-\lambda_n)H_{map} + H_b + H_{SS} + H_{col}\right)^{-1} \left(\lambda_n \nabla E_{nb} + (1-\lambda_n)\nabla E_{map} + \nabla E_b + \nabla E_{SS} + \nabla E_{col}\right), \quad (10)$$

where H_{nb}, H_{map}, H_b, H_{SS}, and H_{col} are the Hessian matrices calculated from E_{nb}, E_{map}, E_b, E_{SS}, and E_{col}, respectively (see Eq. 4).

6. Go to Step 3.
7. Calculate the EM-fitting score $E_{map,n}$ using Eq. 9.
8. Stop if $E_{map,n}$ has reached its threshold value (see below), otherwise set $n \leftarrow n+1$, then go to Step 2.

To reduce accumulation of structural distortions, we limit the magnitude of each incremental displacement (see Eq. 10) within an root mean-square deviation (RMSD) of 0.2 Å. This is attained by adding εI, where I is an identity matrix and ε is an adjustable parameter, to the sum of Hessian matrices such that the linear-equation solution in Eq. 10 satisfies this condition. This idea is akin to the trusted region method used in a recent study of flexible fitting to low-dimensional structural data (40).

To find the optimally fitted model without overfitting, we minimize E_{total} for λ gradually decreasing from 1 toward 0, until E_{map} is below the threshold value $E_{map,th}$ determined by either the inaccuracy of our coarse-grained model (in the absence of noise) or the noise level (in the presence of noise). In the former case, $E_{map,th}$ is calculated using Eq. 9 where the model density ρ_m is generated from the coarse-grained model based on the target structure. In the latter case, $E_{map,th}$ is given by $100/SNR + 1$.

Flexible fitting of an experimental cryo-EM map

An experimental cryo-EM map is preprocessed as follows:

1. The initial structure is docked into the experimental cryo-EM map as a rigid body (or multiple rigid domains) using the QDock command of the SITUS program (3).
2. A single-molecule map is obtained by keeping the voxels within a distance of σ (one-half the map resolution) from the docked initial structure.
3. The average $\langle \rho_{\exp} \rangle$ and standard deviation

$$\delta_{\exp} = \sqrt{\left\langle \left(\rho_{\exp} - \langle \rho_{\exp} \rangle\right)^2 \right\rangle}$$

of the experimental densities within the single-molecule map are computed.

4. The average $\langle \rho_m \rangle$ and standard deviation

$$\delta_m = \sqrt{\left\langle \left(\rho_m - \langle \rho_m \rangle\right)^2 \right\rangle}$$

of the modeled densities based on the initial structure are computed.

5. The experimental densities are shifted and rescaled:

$$\rho_{\exp} \Rightarrow \left(\rho_{\exp} - \langle \rho_{\exp} \rangle\right) \cdot \frac{\delta_m}{\delta_{\exp}} + \langle \rho_m \rangle.$$

Then we use the mENM method to perform flexible fitting to the preprocessed experimental cryo-EM map.

RESULTS

Test cases and model evaluation

From two previous studies of cryo-EM flexible fitting (49,51), we have selected 10 pairs of protein structures with the first structure used as the initial structure and the second structure used to simulate the target cryo-EM map. These cases sample a wide range of conformational changes from 1.6 Å to 14.4 Å in RMSD of C_α atoms (see Table 1).

Simulated cryo-EM maps are generated with resolutions of 5 Å, 10 Å, 15 Å, and 20 Å, respectively (see Methods). Simulated maps are used here for method validation, because the target structures are known and can be used to rigorously evaluate the quality of final fitted models. The effects of random noise that mimics the noise present in experimental maps are also considered (see below). The applicability of our method to experimental cryo-EM maps is also tested (see below).

We first perform a rigid-body fitting of an initial structure to a target map, then we iteratively displace the initial structure to minimize a pseudo-energy that linearly combines several energy terms that maintain the initial tertiary structure, pseudo-bonds, secondary structures, and penalize steric collisions, together with an EM-fitting score (see Methods). We define the EM-fitting score as the squared deviation between a model map and a target map (see Methods, also see (35)). An alternative EM-fitting score, the real-space cross-correlation coefficient (CC), was also used in previous studies (see (33,34)). We do not use CC here because its derivatives are mathematically more cumbersome to calculate. Neither score alone is sufficient because their overoptimization may lead to overfitting. To unambiguously assess the quality of a final fitted model, we compute the C_α-based RMSD between the model and the target structure following a superposition that removes the net translation and rotation between them (denoted as RMSD). It is then compared with the C_α-based RMSD between the initial structure and the target structure following the rigid-body fitting (denoted as $RMSD_0$). A significant decrease from $RMSD_0$ to RMSD indicates a good performance for the flexible fitting.

To separate the errors of initial rigid-body fitting and subsequent flexible fitting, we first assume the initial structure is perfectly superimposed with the target structure without net translation and rotation between them. Next, we will explore the effects of inaccurate initial rigid-body fitting, obtained by the SITUS program (3), on the results of flexible fitting.

A case study of the closed-to-open transition of adenylate kinase

To illustrate our method, we use the closed-to-open transition of adenylate kinase as an example, which was also explored in previous flexible fitting studies (see (19,29)). X-ray structures of adenylate kinase have been solved in both a closed form (PDB code: 1ake) and an open form (PDB code: 4ake). The former is used as the initial structure and the latter is used to simulate a target cryo-EM map with 10 Å resolution. After 15 rounds of minimization as the

TABLE 1 Results of flexible fitting for perfectly superimposed initial structures and cryo-EM maps with various resolutions and noise levels

				Final (minimal) RMSD at various noise levels (Å)			
Target PDB	Initial PDB	Resolution (Å)	Initial $RMSD_0$ (Å)	No noise	SNR = 5	SNR = 2	SNR = 1
1ama	9aat	5	1.66	0.52(0.49)	0.77(0.63)	0.72(0.70)	0.74(0.74)
		10		0.59(0.59)	0.95(0.89)	1.07(0.98)	1.33(1.19)
		15		0.76(0.72)	1.87(1.62)	1.91(1.63)	1.96(1.64)
		20		0.95(0.94)	1.90(1.66)	2.01(1.66)	2.14(1.66)
1anf	1omp	5	3.77	0.47(0.45)	0.72(0.54)	0.79(0.62)	0.72(0.67)
		10		0.52(0.51)	1.08(0.88)	1.28(1.12)	1.51(1.38)
		15		0.60(0.59)	2.77(1.60)	3.57(1.86)	3.71(2.31)
		20		0.66(0.64)	3.74(2.36)	3.76(3.26)	3.78(3.77)
1aonA	1aonH	5	12.25	0.90(0.85)	5.36(5.10)	5.94(5.89)	11.65(11.47)
		10		1.02(1.00)	1.28(1.26)	1.35(1.34)	1.48(1.46)
		15		1.08(1.07)	7.84(7.69)	8.22(7.94)	7.41(7.23)
		20		1.13(1.12)	8.72(5.49)	9.36(5.57)	9.45(4.85)
1lfh	1lfg	5	6.44	0.50(0.49)	0.71(0.57)	0.93(0.91)	1.07(1.06)
		10		0.64(0.63)	1.06(0.90)	2.24(1.75)	1.53(1.26)
		15		0.61(0.59)	1.75(1.54)	1.35(1.07)	3.39(2.76)
		20		0.80(0.78)	1.92(1.70)	2.34(2.33)	4.34(3.56)
1n0u	1n0v	5	14.44	0.68(0.65)	1.04(0.85)	1.07(1.07)	1.27(1.23)
		10		0.89(0.86)	1.41(1.22)	1.90(1.63)	2.69(2.03)
		15		1.01(1.00)	2.50(2.41)	3.30(3.01)	3.75(3.33)
		20		1.16(1.16)	2.34(2.31)	5.07(3.97)	7.52(5.02)
1oaoD	1oaoC	5	7.05	0.43(0.43)	0.73(0.56)	0.69(0.62)	0.73(0.68)
		10		0.47(0.45)	0.80(0.77)	0.88(0.88)	1.08(0.96)
		15		0.51(0.51)	1.26(1.05)	3.02(2.67)	2.29(2.17)
		20		0.57(0.54)	2.50(2.02)	2.24(1.52)	4.16(3.86)
1vfv	1i5s	5	1.81	0.72(0.70)	1.00(0.94)	1.04(1.01)	1.16(1.15)
		10		0.73(0.71)	1.64(1.42)	1.79(1.67)	1.81(1.81)
		15		0.92(0.91)	1.81(1.78)	1.81(1.80)	1.81(1.80)
		20		1.09(1.06)	1.81(1.81)	1.82(1.81)	1.81(1.81)
2lao	1lst	5	4.70	0.40(0.36)	0.74(0.48)	0.59(0.51)	0.61(0.54)
		10		0.51(0.47)	0.71(0.67)	0.74(0.72)	1.02(0.94)
		15		0.58(0.52)	1.46(1.43)	2.32(2.30)	3.82(3.08)
		20		0.56(0.55)	1.93(1.74)	4.54(2.03)	4.69(2.61)
4ake	1ake	5	7.13	0.60(0.59)	0.92(0.78)	0.89(0.89)	1.05(1.04)
		10		0.74(0.73)	1.25(1.13)	1.74(1.50)	2.26(2.00)
		15		0.86(0.85)	2.27(2.08)	2.41(2.32)	3.00(2.70)
		20		0.99(0.91)	2.85(2.81)	4.45(3.21)	4.66(3.46)
4tgl	3tgl	5	1.59	0.79(0.72)	1.32(1.16)	1.31(1.29)	1.36(1.34)
		10		1.02(1.02)	1.40(1.21)	1.57(1.50)	1.58(1.58)
		15		0.70(0.67)	1.59(1.35)	1.59(1.53)	1.59(1.57)
		20		0.68(0.68)	1.59(1.59)	1.59(1.59)	1.59(1.59)

weight parameter λ gradually decreases (see Methods), the flexible fitting is terminated at $\lambda = 0.001$ when the threshold value of EM-fitting score (determined by the inaccuracy of coarse-grained modeling, see Methods) is reached. During the fitting process, the EM-fitting score is reduced dramatically from 37.3 to 0.111, and the RMSD (relative to the target structure) is lowered from 7.13 Å to 0.74 Å (see Fig. 1 *a*). The final model agrees very well with the target structure without local structural distortion (see Fig. 1 *b*). As for computing time, the flexible fitting of adenylate kinase takes 180 s on a 2.5-GHz Intel dual-quad-core Xeon node. For comparison, we have tried the normal mode flexible fitting (NMFF) program (33,34) using the lowest 26 modes solved from an all-atom ENM and a C_α-only ENM, which finished with a final RMSD of 1.6 Å and 3.4 Å, respectively.

Improvement of flexible fitting by the use of two-bead-per-residue representation

In previous cryo-EM flexible fitting studies, a C_α-only coarse-grained representation was often used (see (33,34,51)). Here we adopt a finer representation where two beads are used to represent an amino-acid residue (one at the C_α atom, the other at the center of mass of side-chain atoms if present, see (52)). We have run flexible fitting using both representations and compared the results. In all but one case, we have obtained a better final model

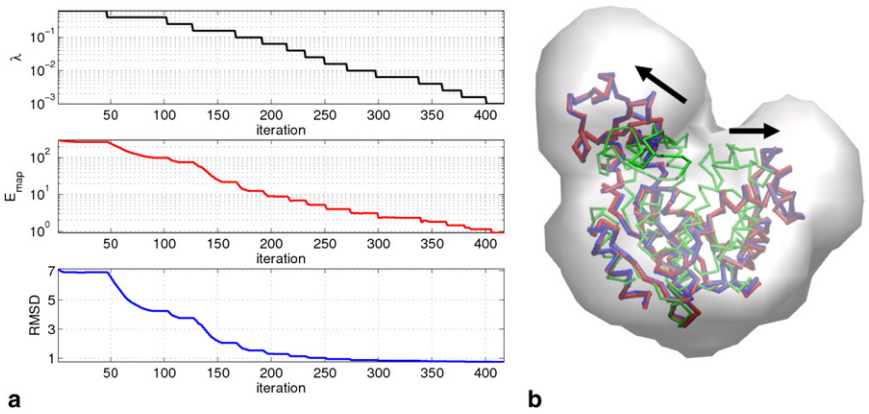

FIGURE 1 Result of flexible fitting of a closed-form structure of adenylate kinase to a 10 Å resolution cryo-EM map simulated from an open-form structure of adenylate kinase: (*a*) weight parameter λ, EM-fitting score E_{map}, and RMSD (relative to the target structure) as a function of iteration step are shown. (*b*) The initial structure (*thin trace*), target structure (*thick trace*), and final model (*thick trace*) are shown; the target cryo-EM map is shown (*solid surface*).

using the two-bead-per-residue representation than the C_α-only one (the final RMSD is on average lowered by ~0.2 Å, see Table S2). Such improvement, although small, may be significant when a high-accuracy cryo-EM-fitted model is needed.

Improvement of flexible fitting by modifying the ENM energy

The original ENM energy has a parabolic form (see Eq. 1), which does not allow large conformational changes that move contacting residues far apart. Consequently, ENM-based normal modes are known to perform worse in describing closed-to-open structural transitions than open-to-closed ones (27). To enhance the sampling of large conformational changes, we have modified the original ENM energy with a new form termed the mENM energy (see Eq. 2), which allows two nonbonded beads to move far apart at a finite energy cost. We have run flexible fitting using both energy functions and compared the results. In all cases, we have obtained a better final model using the mENM energy than the ENM energy (the final RMSD is on average lowered by ~0.6 Å, see Table S2). For example, in the case of adenylate kinase, the use of the mENM energy lowers the final RMSD significantly from 2.09 Å to 0.74 Å.

Improvement of flexible fitting by effectively using all normal modes

Rather than using the lowest few normal modes (see (33,34)), our calculation of the incremental structural displacement implicitly uses all modes of the linearly combined Hessian matrix (see Eq. 10). To show the improvement by using all modes, we have run the NMFF program (33,34) using the lowest 26 modes (including six translation and rotation modes) solved from an all-atom ENM and a C_α-only ENM, respectively. In all but one case, our method has obtained a better final model than NMFF (the final RMSD is on average lowered by ~0.6 Å and 1 Å, respectively, see Table 2), while the computing time remains comparable (see Table S3). For example, we consider the transition between two Kif1A structures—the initial structure is bound with ADP (PDB code: 1i5s), and the target structure is bound with an ATP analog (PDB code: 1vfv). To ensure the same number of residues from both structures is included, the residues only present in 1vfv (residues 353–362) and 1i5s (residues 206–212, 254–268) are deleted. This transition involves a translation and rotation of switch II helix (see Fig. 2 *b*), which is poorly described by the lowest normal modes (55,56). Indeed, NMFF only lowers the RMSD slightly from 1.81 Å to 1.37 Å (using all-atom ENM) and 1.51 Å (using C_α-only ENM). By effectively using all modes, our method has lowered the RMSD significantly from 1.81 Å to 0.73 Å (see Fig. 2, *a* and *b*).

Gradual tuning of the pseudo-energy avoids trapping in local minima

We minimize the pseudo-energy with the weight parameter λ gradually decreasing from 1 to 0 rather than being fixed (see Methods). The idea is to remove λ-dependent local minima of the pseudo-energy—corresponding to the solutions of $\nabla E_{total}(\lambda) = 0$, that may trap conformational sampling. To show the effectiveness of this minimization scheme, we have rerun the flexible fitting using a fixed $\lambda = 0.001$ (for noise-free cryo-EM maps with resolution of 10 Å). In all but two cases, the final RMSD is higher than if λ is gradually tuned (see Table S2). For example, in four cases with large conformational change (1aonH→1aonA, 1lfg→1lfh, 1n0v→1n0u, and 1ake→4ake), the final model has a much higher RMSD than if λ is gradually tuned (see Table S2), which indicates trapping in local minima far from the target structure. Therefore, the tuning of λ indeed helps to avoid trapping in local minima which may occur if λ is fixed.

TABLE 2 Comparison of various flexible fitting methods for noise-free cryo-EM maps

Target PDB	Initial PDB	Initial $RMSD_0$ (Å)	Final RMSD for various methods (resolution) (Å) mENM (10 Å)	NMFF* (10 Å)	NMFF† (10 Å)	Biased MD (all-atom)‡ (8 Å)	Biased MD (Gō-model)§ (10 Å)
1ama	9aat	1.66	0.59	0.65	0.89	1.59	—
1anf	1omp	3.77	0.52	0.97	1.18	1.22	—
1aonA	1aonH	12.25	1.02	1.48	1.67	—	—
1lfh	1lfg	6.44	0.61	0.90	1.08	—	1.3
1n0u	1n0v	14.44	0.89	2.65	2.90	—	1.2
1oaoD	1oaoC	7.05	0.47	1.16	1.99	—	1.1
1vfv	1i5s	1.81	0.73	1.37	1.51	—	—
2lao	1lst	4.70	0.51	1.18	1.76	1.16	—
4ake	1ake	7.13	0.74	1.60	3.40	—	1.8
4tgl	3tgl	1.59	1.02	0.92	1.07	1.39	—
	Average		0.71	1.29	1.75	—	—

*See Tama et al. (34) (using 26 lowest modes solved from an all-atom ENM with cutoff distance of 5.5 Å).
†See Tama et al. (34) (using 26 lowest modes solved from a C_α-only ENM with cutoff distance of 8 Å).
‡See Orzechowski and Tama (49).
§See Grubisic et al. (51).

Dependence of flexible fitting performance on map resolution

We have performed flexible fitting using noise-free cryo-EM maps with resolutions of 5 Å, 10 Å, 15 Å, and 20 Å. For all except one case, the final RMSD decreases as the resolution parameter decreases (see Table 1), which supports the general notion that higher-resolution cryo-EM maps enable more accurate structural modeling. For all cases, our method has consistently obtained accurately fitted models with RMSD $\leq$ 1.2 Å for noise-free cryo-EM maps (see Table 1) using the same model parameters. Therefore, our method is applicable to a wide range of protein conformational changes and map resolutions. The computing time decreases as the resolution parameter increases (see Table S4), thanks to the use of coarser grids and fewer voxels for lower-resolution maps, which lowers the cost of calculating EM-fitting score and its derivatives.

Dependence of flexible fitting performance on noise level

Most cryo-EM-fitting studies (except (19,30)) focused on noise-free simulated cryo-EM maps, whereas experimental maps contain noise caused by imaging process and image processing and reconstruction. To evaluate the effect of random noise at a controllable level, Gaussian noise with a given SNR is added to the simulated maps (following (19)). We have systematically explored SNR of 1, 2, and 5, which corresponds to

$$CC \sim \sqrt{SNR/(1+SNR)}$$

of 0.7, 0.8, and 0.9, respectively. In recent cryo-EM-fitting studies, experimental maps were typically fitted with a CC of 0.8 ~ 0.9 (29,34,35,47), which corresponds to SNR $\geq$ 2.

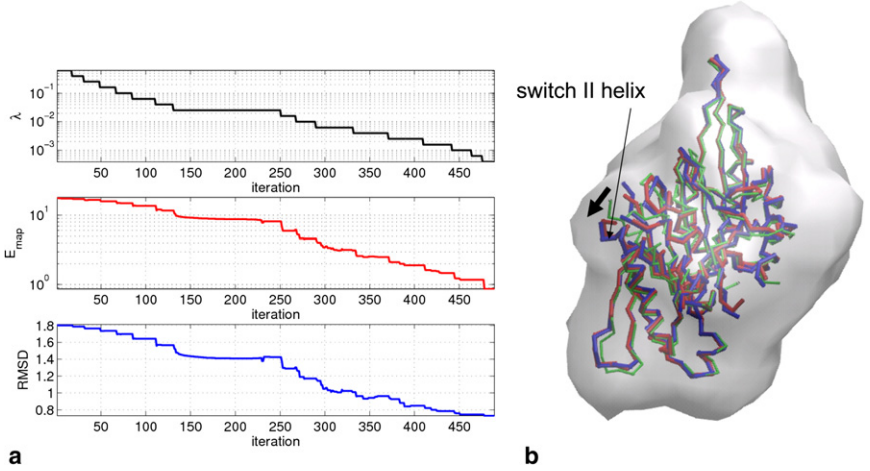

FIGURE 2 Result of flexible fitting of an ADP-bound Kif1A structure to a 10 Å resolution cryo-EM map simulated from an ATP-analog-bound Kif1A structure: (*a*) weight parameter λ, EM-fitting score E_{map}, and RMSD (relative to the target structure) as a function of iteration step are shown. (*b*) The initial structure (*thin trace*), target structure (*thick trace*), and final model (*thick trace*) are shown; the target cryo-EM map is shown (*solid surface*).

As expected, the flexible fitting performance gets worse as the noise level increases, although the degree of worsening varies between proteins (see Table 1). For example, in case of adenylate kinase (with map resolution of 10 Å), the final RMSD increases from 0.74 Å to 1.25 Å, 1.74 Å and 2.26 Å, all of which remain much lower than the initial $RMSD_0$ of 7.13 Å, as SNR decreases from infinity to 5, 2, and 1. However, in case of Kif1A (with map resolution of 10 Å), the final RMSD quickly increases from 0.73 Å to 1.64 Å, 1.79 Å, and 1.81 Å, which are near the initial $RMSD_0$ of 1.81 Å, as SNR decreases from infinity to 5, 2, and 1. So the flexible fitting of adenylate kinase is more robust to noise than Kif1A.

Generally, the flexible fitting of lower-resolution maps is more susceptible to noise. For example, in case of adenylate kinase (with map resolution of 20 Å), the final RMSD increases from 0.99 Å to 2.85 Å, 4.45 Å, and 4.66 Å as SNR decreases from infinity to 5, 2, and 1.

For most cases, our method has generated accurately fitted models with RMSD < 2 Å for noisy cryo-EM maps with resolution of 10 Å and SNR ≥ 2 (see Table 1). Therefore, our method is highly promising for fitting medium/high-resolution experimental cryo-EM maps.

Flexible fitting after inaccurate initial rigid-body fitting

The above test of our flexible fitting method assumes that the initial structure is perfectly superimposed on the target structure with no net rotation or translation. In practice, a given initial structure has to be fitted to a target map manually or using a rigid-body cryo-EM-fitting program, which results in an inaccurately oriented initial model. We have performed rigid-body fitting using the QDock command of the SITUS program (3) with 10 codebook vectors (following (34)). The top five initial models are ranked by CC. To test the effect of inaccurate rigid-body fitting, we have performed flexible fitting starting from each of the five initial models.

In most cases, not all initial models can be refined by flexible fitting to a final model with a low RMSD (see Table S5). For example, in case of adenylate kinase, only the first two initial models allow successful flexible fitting whereas the last three do not. Compared with the two successful initial models, the three failed ones all have very large $RMSD_0$, suggesting that they have a wrong orientation. Therefore, a roughly correct orientation of the initial model is required for our flexible fitting method to succeed.

How to select a good initial model from the top five models generated by SITUS: For all cases, we have found the model(s) with the highest CC to be good enough for flexible fitting (see Table S5), which result in a final model with a low RMSD similar to that obtained for a perfectly oriented initial structure (see Table 1 and Table S5). Additionally, as long as the initial orientation is roughly correct, the flexible fitting converges well, as judged from the similar final RMSD values reached starting from different initial models (see Table S5), although the convergence is worse as the noise level increases.

Comparison with alternative cryo-EM flexible fitting methods

Our method offers a useful addition to a pool of cryo-EM flexible fitting techniques developed previously (19,29,30,33,34,49,51,57). Here we compare our method with a number of closely related methods.

The NMFF program (33,34) used the lowest few normal modes of ENM to iteratively optimize the fitting to a target cryo-EM map. For the same set of test cases, our method has obtained better final models than NMFF (see Table 2). The computing cost of our method is comparable to NMFF (see Table S3). The improvement in accuracy is attributed to the effective use of all modes, which allows a full description of protein flexibility including both global and local structural changes. The biased MD simulation based on an all-atom force field (49) or a coarse-grained Gō-model (51) was used to perform flexible fitting of cryo-EM maps. We have compared our results with other studies (49,51) for the common test cases. In all cases, we have obtained better final models (see Table 2).

Flexible fitting of experimental cryo-EM maps

To demonstrate the applicability of our method to experimental cryo-EM data, we have applied it to the myosin motor protein which contains ~1100 residues. Myosin generates force and movement by binding with the actin filaments in an ATP-dependent fashion (58). The cryo-EM reconstruction at 14 Å resolution of a nucleotide-free myosin bound to the actin filaments revealed a conformation different from the nucleotide-bound x-ray structure of myosin (59). A satisfactory fitting of the cryo-EM data was attained after the Upper-50 kDa (U50) subdomain was rotated toward the Lower-50 kDa (L50) subdomain (59) (see Fig. 3 *b*). We have performed flexible fitting of the preprocessed single-molecule myosin map (see Methods) using the x-ray structure of an ATP-analog-bound myosin (following (59)). The initial rigid-body fitting is performed using SITUS, which obtains a CC of 0.73 (calculated for the inside of the single-molecule envelope). The flexible fitting is terminated when the RMSD relative to the initial structure starts to saturate near ~3 Å. The final model's CC is improved to 0.84. In agreement with Holmes et al. (59), the flexible fitting reveals a rotation of the U50 subdomain that closes the cleft between the U50 and L50 subdomains (see Fig. 3 *b*). We have further tested our method using two more cases of experimental cryo-EM data (see the Supporting Material).

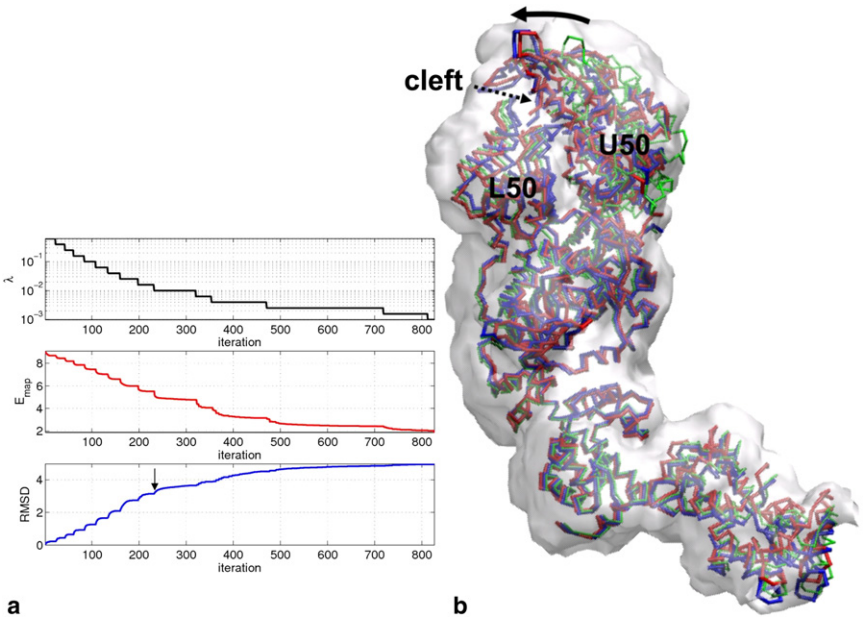

FIGURE 3 Result of flexible fitting of an ATP-analog-bound myosin structure to a 14 Å resolution cryo-EM map of a nucleotide-free myosin molecule: (*a*) weight parameter λ, EM-fitting score E_{map}, and RMSD (relative to the initial structure) as a function of iteration step are shown (the step of RMSD saturation is pointed to by an *arrow*). (*b*) The initial structure (*thin trace*), the final model (*thick trace*), and the fitted model obtained in Holmes et al. (59) (*thick trace*) are shown, and the target cryo-EM map is shown (*solid surface*). The rotation of the U50 subdomain toward the L50 subdomain is shown.

DISCUSSION AND CONCLUSION

We have developed what we believe is a new computational method to flexibly fit a given cryo-EM map with a coarse-grained model using a two-bead-per-residue representation, and a modified form of the elastic network model that allows large-scale conformational changes while maintaining pseudobonds and secondary structures. Our method effectively utilizes all normal modes so that both global and local structural changes can be fully modeled. Based on the cryo-EM-fitted models, it is straightforward to build all-atom models using existing side-chain modeling programs (available in MMTSB (60)). Compared with the C_α-only coarse-grained models, the addition of side-chain beads allows side-chain modeling to be conducted more accurately and efficiently.

In practice, our flexible fitting method should be combined with a rigid-body fitting program like SITUS (3) which is used to find an approximately correct orientation of an initial structure relative to a target map. Our method may not work if the initial orientation is wrong. In our test, the best model produced by the QDock command of the SITUS program (with the highest CC) gives a good initial model for subsequent flexible fitting. In practice, flexible fitting should be done starting from several initial models, and convergence toward a common conformation should be checked to ensure the reliability of modeling.

Our method compares favorably with alternative methods in accuracy, and its computing cost is very low. It is also very easy to use with the same parameters applicable to a variety of protein conformational changes and cryo-EM maps of different resolutions. Our method is also shown to be robust to random noise with SNR $\geq$ 2.

In principle, the flexible fitting of a high-resolution structure into a low-resolution cryo-EM map is prone to the overfitting problem, because the many degrees of freedom involved in fitting are generally insufficiently constrained by the given structural data. In our method, we take the following measures to reduce overfitting: first, the total number of degrees of freedom is reduced by using a coarse-grained model and maintaining the pseudobonds and the secondary structures (following (47)); second, the flexible fitting is timely terminated when a threshold EM-fitting score is reached, which depends on the noise level or the inaccuracy of coarse-grained modeling, or when the RMSD relative to the initial structure starts to saturate. Indeed, in our test of 10 cases with various map resolutions and noise levels, we have consistently obtained final models with low RMSD which is close to the minimal RMSD attained during fitting (see Table 1). We have also visually inspected the final models and did not find serious structural distortion indicative of overfitting.

A general concern about any minimization-based method is the local minima trapping problem. To avoid being trapped in local minima, we perform a gradual tuning of the weight parameter λ to eliminate local minima. Indeed, for all the cases with cryo-EM maps of 10 Å resolution, we do not find the fitting to be trapped in a conformation far from the target structure if it starts from a good initial model with roughly correct orientation (see Table S5). Alternative minimization protocols like simulated annealing (16,19,29) may also be used to assist the finding of global minimum at the expense of higher computing cost. Similarly, MD-based methods can be used to avoid trapping in local minima (see (49,51,61)).

The idea of adding an EM-fitting score to a physical potential to bias the conformational sampling was explored

in previous studies (see (49,51)). The weight of this score must be properly chosen to accelerate conformational sampling while maintaining protein structural integrity. It is not easy to determine the optimal weight before the simulation because it may vary between proteins. To address this issue, rather than using a fixed weight, our method gradually tunes the weight parameter λ until a threshold EM-fitting score is reached, which allows the fitting process to self-adjust to different cases.

The idea of minimizing a linear combination of various energy functions was previously used by us to model transition pathways (54), and will be used to fit other low-resolution structural data (40) in the future.

Despite its success, our method and other ENM-based cryo-EM-fitting methods have limitations—they are most suited for proteins undergoing en-block motions of domains or secondary structural elements. A more detailed simulation using MD and all-atom force-field (47) will be needed to model local changes involving folding/unfolding of secondary structures or restructuring of surface loops or side chains.

SUPPORTING MATERIAL

Additional results, six tables, and two figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(10)05186-6.

I thank the funding support from American Heart Association (grant No. 0835292N) and National Science Foundation (grant No. 0952736), the EM data of myosin-decorated actin from Dr. K. Holmes, and the EM data of *Escherichia coli* RNAP from Dr. Seth A. Darst. The method will be made available on a web server (enm.lobos.nih.gov).

REFERENCES

1. Saibil, H. R. 2000. Conformational changes studied by cryo-electron microscopy. *Nat. Struct. Biol.* 7:711–714.
2. Fabiola, F., and M. S. Chapman. 2005. Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure.* 13:389–400.
3. Wriggers, W., R. A. Milligan, and J. A. McCammon. 1999. SITUS: a package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.* 125:185–195.
4. Volkmann, N., and D. Hanein. 1999. Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J. Struct. Biol.* 125:176–184.
5. Rossmann, M. G. 2000. Fitting atomic models into electron-microscopy maps. *Acta Crystallogr. D Biol. Crystallogr.* 56:1341–1349.
6. Rossmann, M. G., R. Bernal, and S. V. Pletnev. 2001. Combining electron microscopic with x-ray crystallographic structures. *J. Struct. Biol.* 136:190–200.
7. Jiang, W., M. L. Baker, …, W. Chiu. 2001. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J. Mol. Biol.* 308:1033–1044.
8. Chacon, P., and W. Wriggers. 2002. Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.* 317:375–384.
9. Wu, X., J. L. Milne, …, B. R. Brooks. 2003. A core-weighted fitting method for docking atomic structures into low-resolution maps: application to cryo-electron microscopy. *J. Struct. Biol.* 141:63–76.
10. Volkmann, N., D. Hanein, …, S. Lowey. 2000. Evidence for cleft closure in actomyosin upon ADP release. *Nat. Struct. Biol.* 7:1147–1155.
11. Wendt, T., D. Taylor, …, K. Taylor. 2001. Three-dimensional image reconstruction of dephosphorylated smooth muscle heavy meromyosin reveals asymmetry in the interaction between myosin heads and placement of subfragment 2. *Proc. Natl. Acad. Sci. USA.* 98:4361–4366.
12. Rawat, U. B., A. V. Zavialov, …, J. Frank. 2003. A cryo-electron microscopic study of ribosome-bound termination factor RF2. *Nature.* 421:87–90.
13. Gao, H., J. Sengupta, …, J. Frank. 2003. Study of the structural dynamics of the *E. coli* 70S ribosome using real-space refinement. *Cell.* 113:789–801.
14. Gao, H., and J. Frank. 2005. Molding atomic structures into intermediate-resolution cryo-EM density maps of ribosomal complexes using real-space refinement. *Structure.* 13:401–406.
15. Mears, J. A., M. R. Sharma, …, S. C. Harvey. 2006. A structural model for the large subunit of the mammalian mitochondrial ribosome. *J. Mol. Biol.* 358:193–212.
16. Topf, M., K. Lasker, …, A. Sali. 2008. Protein structure fitting and refinement guided by cryo-EM density. *Structure.* 16:295–307.
17. Jacobs, D. J., and M. F. Thorpe. 1995. Generic rigidity percolation: the pebble game. *Phys. Rev. Lett.* 75:4051–4054.
18. Jacobs, D. J., A. J. Rader, …, M. F. Thorpe. 2001. Protein flexibility predictions using graph theory. *Proteins.* 44:150–165.
19. Jolley, C. C., S. A. Wells, …, M. F. Thorpe. 2008. Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations. *Biophys. J.* 94:1613–1621.
20. Velazquez-Muriel, J. A., M. Valle, …, J. M. Carazo. 2006. Flexible fitting in 3D-EM guided by the structural variability of protein superfamilies. *Structure.* 14:1115–1126.
21. Velazquez-Muriel, J. A., and J. M. Carazo. 2007. Flexible fitting in 3D-EM with incomplete data on superfamily variability. *J. Struct. Biol.* 158:165–181.
22. Topf, M., M. L. Baker, …, A. Sali. 2006. Refinement of protein structures by iterative comparative modeling and CryoEM density fitting. *J. Mol. Biol.* 357:1655–1668.
23. Wriggers, W., and S. Birmanns. 2001. Using SITUS for flexible and rigid-body fitting of multiresolution single-molecule data. *J. Struct. Biol.* 133:193–202.
24. Tozzini, V. 2005. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* 15:144–150.
25. Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins.* 33:417–429.
26. Atilgan, A. R., S. R. Durell, …, I. Bahar. 2001. Anisotropy of fluctuation dynamics of proteins with an elastic network model. *Biophys. J.* 80:505–515.
27. Tama, F., and Y. H. Sanejouand. 2001. Conformational change of proteins arising from normal mode calculations. *Protein Eng.* 14:1–6.
28. Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.
29. Tan, R. K., B. Devkota, and S. C. Harvey. 2008. YUP.SCX: coaxing atomic models into medium resolution electron density maps. *J. Struct. Biol.* 163:163–174.
30. Schroder, G. F., A. T. Brunger, and M. Levitt. 2007. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure.* 15:1630–1641.
31. Schroder, G. F., M. Levitt, and A. T. Brunger. 2010. Super-resolution biomolecular crystallography with low-resolution data. *Nature.* 464:1218–1222.
32. Delarue, M., and P. Dumas. 2004. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proc. Natl. Acad. Sci. USA.* 101:6957–6962.

33. Tama, F., O. Miyashita, and C. L. Brooks, 3rd. 2004. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J. Mol. Biol.* 337:985–999.

34. Tama, F., O. Miyashita, and C. L. Brooks, 3rd. 2004. NMFF: flexible high-resolution annotation of low-resolution experimental data from cryo-EM maps using normal mode analysis. *J. Struct. Biol.* 147: 315–326.

35. Hinsen, K., N. Reuter, …, J. J. Lacapere. 2005. Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys. J.* 88:818–827.

36. Suhre, K., J. Navaza, and Y. H. Sanejouand. 2006. NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallogr. D Biol. Crystallogr.* 62:1098–1100.

37. Mitra, K., C. Schaffitzel, …, J. Frank. 2005. Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature.* 438:318–324.

38. Falke, S., F. Tama, …, M. T. Fisher. 2005. The 13 Ångstroms structure of a chaperonin GroEL-protein substrate complex by cryo-electron microscopy. *J. Mol. Biol.* 348:219–230.

39. Tama, F., G. Ren, …, A. K. Mitra. 2006. Model of the toxic complex of anthrax: responsive conformational changes in both the lethal factor and the protective antigen heptamer. *Protein Sci.* 15:2190–2200.

40. Gorba, C., O. Miyashita, and F. Tama. 2008. Normal-mode flexible fitting of high-resolution structure of biological molecules toward one-dimensional low-resolution data. *Biophys. J.* 94:1589–1599.

41. Zheng, W., and B. R. Brooks. 2005. Normal-modes-based prediction of protein conformational changes guided by distance constraints. *Biophys. J.* 88:3109–3117.

42. Zheng, W., and B. R. Brooks. 2006. Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes. *Biophys. J.* 90:4327–4336.

43. Li, W., and J. Frank. 2007. Transfer RNA in the hybrid P/E state: correlating molecular dynamics simulations with cryo-EM data. *Proc. Natl. Acad. Sci. USA.* 104:16540–16545.

44. Chapman, M. S. 1995. Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function. *Acta Crystallogr. A.* 51:69–80.

45. Chen, J. Z., J. Furst, …, N. Grigorieff. 2003. Low-resolution structure refinement in electron microscopy. *J. Struct. Biol.* 144:144–151.

46. Noda, K., M. Nakamura, …, T. Yasunaga. 2006. Atomic model construction of protein complexes from electron micrographs and visualization of their 3D structure using a virtual reality system. *J. Plasma Phys.* 72:1037–1040.

47. Trabuco, L. G., E. Villa, …, K. Schulten. 2008. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure.* 16:673–683.

48. Trabuco, L. G., E. Villa, …, K. Schulten. 2009. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and x-ray crystallography. *Methods.* 49:174–180.

49. Orzechowski, M., and F. Tama. 2008. Flexible fitting of high-resolution x-ray structures into cryoelectron microscopy maps using biased molecular dynamics simulations. *Biophys. J.* 95:5692–5705.

50. Caulfield, T. R., and S. C. Harvey. 2007. Conformational fitting of atomic models to cryogenic-electron microscopy maps using Maxwell's demon molecular dynamics. Biophysical Society Meeting Abstracts. *Biophys. J.* (*Suppl.*):368a.

51. Grubisic, I., M. N. Shokhirev, …, F. Tama. 2010. Biased coarse-grained molecular dynamics simulation approach for flexible fitting of x-ray structure into cryo electron microscopy maps. *J. Struct. Biol.* 169:95–105.

52. Tehver, R., J. Chen, and D. Thirumalai. 2009. Allostery wiring diagrams in the transitions that drive the GroEL reaction cycle. *J. Mol. Biol.* 387:390–406.

53. Chen, Y., T. A. Davis, …, S. Rajamanickam. 2008. Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Update/ Downdate. *ACM Trans. Math. Softw.* 35:1–14.

54. Tekpinar, M., and W. Zheng. 2010. Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. *Proteins.* 78:2469–2481.

55. Zheng, W., and S. Doniach. 2003. A comparative study of motor-protein motions by using a simple elastic-network model. *Proc. Natl. Acad. Sci. USA.* 100:13253–13258.

56. Zheng, W., and B. R. Brooks. 2005. Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: myosin versus kinesin. *Biophys. J.* 89:167–178.

57. Lindahl, E., C. Azuara, …, M. Delarue. 2006. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.* 34:W52–W56.

58. Geeves, M. A., and K. C. Holmes. 1999. Structural mechanism of muscle contraction. *Annu. Rev. Biochem.* 68:687–728.

59. Holmes, K. C., I. Angert, …, R. R. Schroder. 2003. Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature.* 425:423–427.

60. Feig, M., J. Karanicolas, and C. L. Brooks, 3rd. 2004. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* 22:377–395.

61. Kovacs, J. A., M. Yeager, and R. Abagyan. 2008. Damped-dynamics flexible fitting. *Biophys. J.* 95:3192–3207.