

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# Structure and Dynamics of Macromolecular Assemblies from Electron Microscopy Maps

---

Xiongwu Wu and Bernard R. Brooks

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/62085>

---

## Abstract

Electron microscopy (EM) and electron tomography (ET) have found extensive application to probe structural and dynamic properties of macromolecular assemblies. As an important complementary to X-ray and NMR in atomic structural determination, EM has reached a milestone resolution of 2.2Å, enough for understanding atomic interactions, interpreting mechanism of functions, and for structure-based drug design. This work describes approaches to derive structural information from EM/ET images and methods to study dynamics of macromolecular systems using EM/ET images as starting or ending targets. For low-resolution EM/ET maps, X-ray or NMR atomic structures of molecular components are needed to reduce the number of degrees of uncertainty. Depending on the resolution of EM/ET maps and the conformational differences from the X-ray or NMR structures, either rigid fitting or flexible fitting is used to obtain atomic structures. To illustrate the procedures of the atomic structure derivation, this work describes the core-weighted grid-threading Monte Carlo (CW-GTMC) rigid fitting and the map-restrained self-guided Langevin dynamics (MapSGLD) flexible fitting methods. Their applications are highlighted with four examples: architecture of an icosahedral pyruvate dehydrogenase complex, dynamics of a group II chaperonin, high-resolution structure of the cell-permeant inhibitor phenylethyl  $\beta$ -D-thiogalactopyranoside, and the mechanism of kinesin walking on microtubule.

**Keywords:** Electron microscopy, electron tomography, rigid fitting, flexible fitting, molecular simulation, self-guided Langevin dynamics, map restraint

---

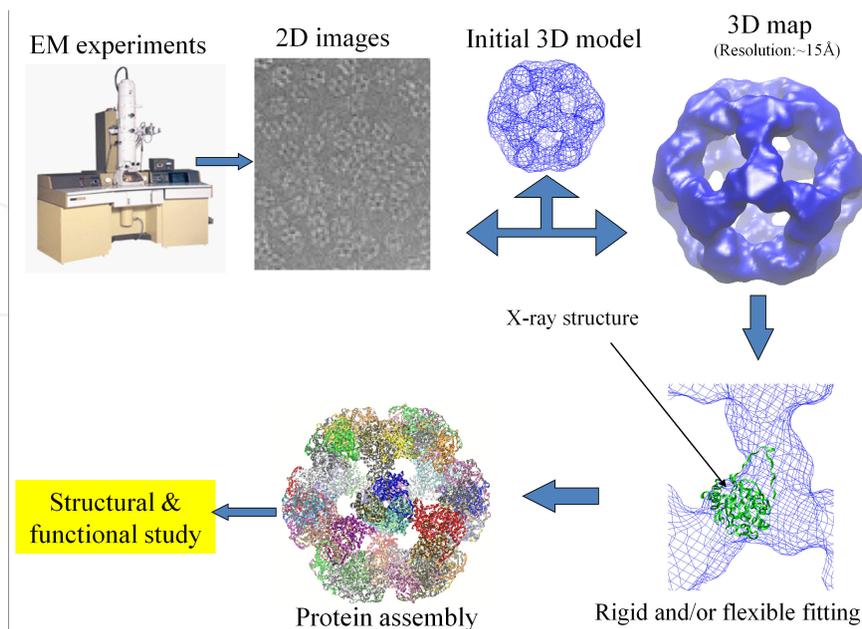
## 1. Introduction

The advance of cryo-electron microscopy (cryo-EM) and electron tomography (ET) opens a new window to the analysis of large biomolecular assemblies under biologically relevant conditions. Figure 1 illustrates the basic procedure from EM/ET experiment to the under-

standing of molecular complexes. 2D images of samples are the raw data from EM/ET experiment. A 3D reconstruction is needed to build volume maps, which often needs an initial 3D model to start with. In most cases, the resolution of EM/ET maps are not high enough to determine atom positions; therefore, people rely on known high-resolution structures from X-ray, NMR, or homology modelling to derive structures of molecular assemblies through either rigid fitting or flexible fitting. Because molecular environments are different in X-ray crystal, in NMR solution, and on cryo-EM grids, molecular structures in these samples are not identical, especially in the side chains, in loops, and around hinge regions. When the resolution of EM/ET maps is not high enough to show structure difference from X-ray or NMR structures, rigid fitting is often used to produce complex structures [1-8].

When EM/ET maps have high enough resolution or for systems with flexible components, rigid fitting cannot handle structural variation information in EM/ET maps. In different states, proteins and protein assemblies often adopt different conformations, such as in bound and unbound states. Typically, protein side chains have certain conformational flexibilities and can adapt to different environmental conditions. To accommodate conformational variations, a process called flexible fitting is used to change structures from X-ray or NMR experiments to match EM/ET maps. There are a series of methods available to perform flexible fitting [9-14].

Whether to choose rigid fitting or flexible fitting depends on the resolution of EM/ET maps. There is no absolute resolution boundary for choosing rigid fitting from flexible fitting. The rule of thumb is that when the resolution cannot distinguish structure difference of EM/ET maps from fitting structures, then rigid fitting should be used. Flexible fitting becomes necessary when the resolution is high enough to show significant structural variation between the fitting X-ray/NMR structures and the EM/ET maps. For proteins undergoing domain hinge motion, a map with a resolution as low as 50 Å could show the difference, while for side chain rotations, a map needs a resolution of 10 Å or higher to identify an alternate orientation.



**Figure 1.** Basic procedure from electron microscopy experiment to macromolecular structures.

To illustrate the procedure of atomic structure derivation from EM maps, we describe here the core-weighted grid-threading Monte Carlo (CW-GTMC) rigid fitting method [2] and the map-restrained self-guided Langevin dynamics (MapSGLD) flexible fitting method [14]. Using four examples, we highlight their applications in structural, dynamics, and mechanism studies. The rigid fitting and flexible fitting methods have been implemented in both AMBER [15] and CHARMM [16]. In AMBER, this method has been implemented in Sander, which is part of AmberTools and is freely available at: <http://ambermd.org/#AmberTools>, as well as in pmemd, which is the main AMBER simulation engine (<http://ambermd.org/#Amber14>). In CHARMM, it is part of the EMAP module, where a rigid fitting is done through EMAP DOCK command, and a flexible fitting is carried out through DYNA simulation command. Users can refer to the emap test cases in both AMBER and CHARMM as examples of rigid and flexible fitting.

## 2. Methods

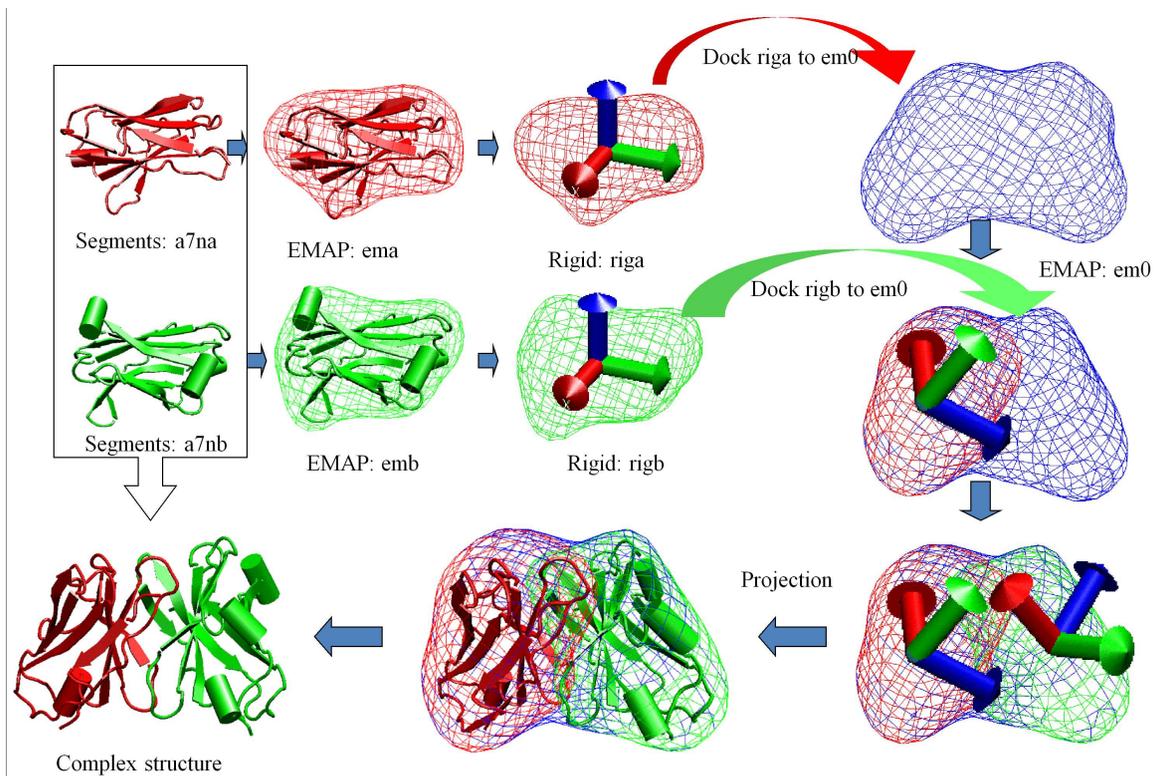
### 2.1. Rigid fitting

Rigid fitting of atomic structures to electron microscopy images is a common approach to interpret low-resolution maps and to obtain complex structural information. Figure 2 illustrates the procedure of a rigid fitting to obtain a complex structure. Assume EM/ET experiment produced the complex map em0 and the structures of components, a7na and a7nb are available. By rigid fitting a7na and a7nb into em0, one can obtain the complex structure of a7na+a7nb. A typical rigid fitting starts with blurring the molecular structures, a7na and a7nb, to create maps, ema and emb, which is often done by distributing atomic masses into grid points in the space. The component maps, ema and emb, are translated and rotated to fit the EM/ET map em0. After all component maps are fit, component molecules are positioned accordingly to produce the complex structure.

A map represents a distribution over a spatial region. Typically, a map is described by quantities, e.g., density, on lattice grid points:

$$M(N_x, N_y, N_z) = \bigcup_{i,j,k}^{N_x, N_y, N_z} \rho(x_i, y_j, z_k) \quad (1)$$

Here, the map grids are defined by grid point positions,  $(x_i, y_j, z_k)$ . The numbers of grids in the  $x$ ,  $y$ , and  $z$  directions are  $N_x$ ,  $N_y$ , and  $N_z$ , respectively. The union symbol  $\bigcup_{i,j,k}^{N_x, N_y, N_z}$  indicates that all entities with  $i \in N_x$ ,  $j \in N_y$ , and  $k \in N_z$  are included in the map. The  $x$ ,  $y$ , and  $z$  axes can form any angle, but for convenience, in this work, we assume them to be orthogonal to each other. The distribution property  $\rho(x_i, y_j, z_k)$  can be any special properties such as charge density, electric potential, solvent accessibility, etc. It is electron density for EM/ET maps.



**Figure 2.** Rigid fitting steps to derive atomic structures of complexes.

A major application of EM/ET experiment is to determine structures of molecular assemblies. For low resolution of EM/ET maps, electron densities of structure components spread to a certain range beyond atomic boundaries, which causes significant density overlap between neighboring structures. We define the “core” region of a structure as the part whose density distribution is unlikely to overlap with the density distribution of the other maps and to be altered by the presence of adjacent components. The “surface” region is defined as the part that is accessible or can interact with other components. Therefore, the core region is always enclosed by the surface region.

It is then possible to define a core index, which describes the depth of a grid point located within this core as follows:

$$f_{ijk} = \begin{cases} 0 & \rho_{ijk} \leq \rho_c \text{ and } \min[f_{i\pm 1jk}, f_{ij\pm 1k}, f_{ijk\pm 1}] = 0 \\ 0 & \nabla^2 \rho_{ijk} > 0 \text{ and } \min[f_{i\pm 1jk}, f_{ij\pm 1k}, f_{ijk\pm 1}] = 0 \\ \min[f_{i\pm 1jk}, f_{ij\pm 1k}, f_{ijk\pm 1}] + 1 & \text{otherwise} \end{cases} \quad (2)$$

where,  $f_{ijk}$  is the core index of grid point  $(i, j, k)$ ,  $\rho_c$  is a cut-off density to define the surface position. The core index is zero for grid points outside the surface region and increases progressively as a grid point goes deeper into the core. The core index is calculated with

$\min[f_{i\pm 1jk}, f_{ij\pm 1k}, f_{ijk\pm 1}] + 1$ , that is, the index of a grid point  $(i, j, k)$  is the minimum core index of the neighboring grid points around this grid point plus 1. The core index has the following features when matching between the map of an individual component and the map of a multicomponent assembly:

1. For a correct fit, a match in the core regions (high core index) corresponds to a high correlation in densities.
2. For a correct fit, a match in the surface regions (low core index) corresponds to a high correlation in densities.
3. For a correct fit, a match between the core region of components and the surface region of assemblies must have low correlation in densities.
4. For a correct fit, a match between a surface (low core index) region and a core region may have low correlation in densities.

For scenarios 1–3, a normal correlation function works fine to distinguish the correct fit from wrong fits. But for scenario 4, the distribution property is altered by the overlap density from neighboring components in a complex map and a regular correlation function is likely to fail. To account for the effect of overlap, and properly recognize the correct fit from misfit, we need to minimize the contribution to correlation function from scenario 4. This can be achieved by “down-weighting” the match between a region with low core index in the map of individual components and a region with high core index in the complex map. We choose the following weighting function to implement this idea:

$$w_{mn} = \frac{f_m^a}{f_m^a + bf_n^a + c} \quad (3)$$

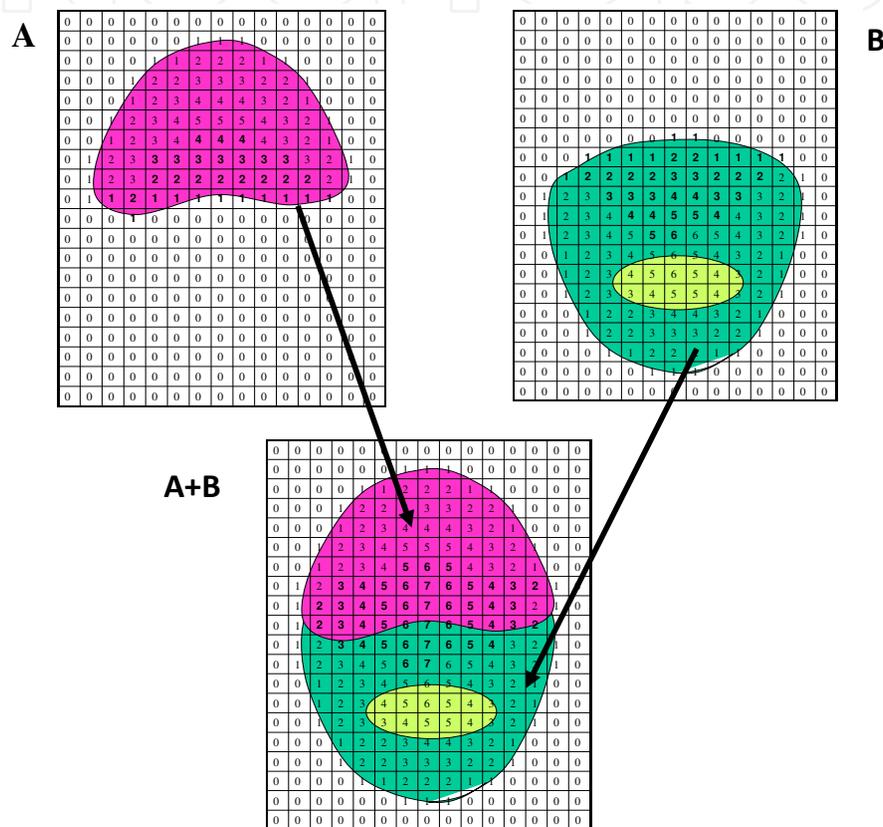
where  $w_{mn}$  is the core-weighting function for the individual component,  $m$ , to the complex,  $n$ . Three parameters,  $a$ ,  $b$ , and  $c$ , control the dependence of the function on the core indices. Typically, we chose  $a = 2$  and  $b = 1$ , and set  $c$  to be a very small constant (e.g.,  $10^{-6}$ ) to ensure  $w_{mn} = 0$  when  $f_m = 0$  and  $f_n = 0$ . We call eq. (3) the core-weighting function because it is based on the core index. Using this core-weighting function we calculate the core-weighted correlation function as shown below:

$$CWDC_{mn} = \frac{\overline{(\rho_m \rho_n)_w} - \overline{(\rho_m)_w} \overline{(\rho_n)_w}}{\delta_w(\rho_m) \delta_w(\rho_n)}, \quad (4)$$

where  $\overline{(X)_w} = \frac{\sum_{i,j,k} w_{mn}(i, j, k) X(i, j, k)}{\sum_{i,j,k} w_{mn}(i, j, k)}$  is a core-weighted average and  $\delta_w(X) = \sqrt{\overline{(X^2)_w} - (\overline{X}_w)^2}$  is

a core-weighted standard deviation of property  $X$ .

Figure 3 illustrates the definition of core indices for the maps of two individual proteins, A and B, and their complex. In each map, the core index is zero outside the density regions, 1 at the outer edge and becomes larger for the grid points that locate more deeply in the core region. Please note that the core region does not necessarily correspond to the region with high density. For internal cavities that are buried well below the surface of a structure (e.g., the cavity in protein B), it is possible that the core index can have high values while densities are low. When proteins A and B form complex, the core indices of their interaction surfaces dramatically increase in regions where the surfaces become deeply buried in the AB complex.

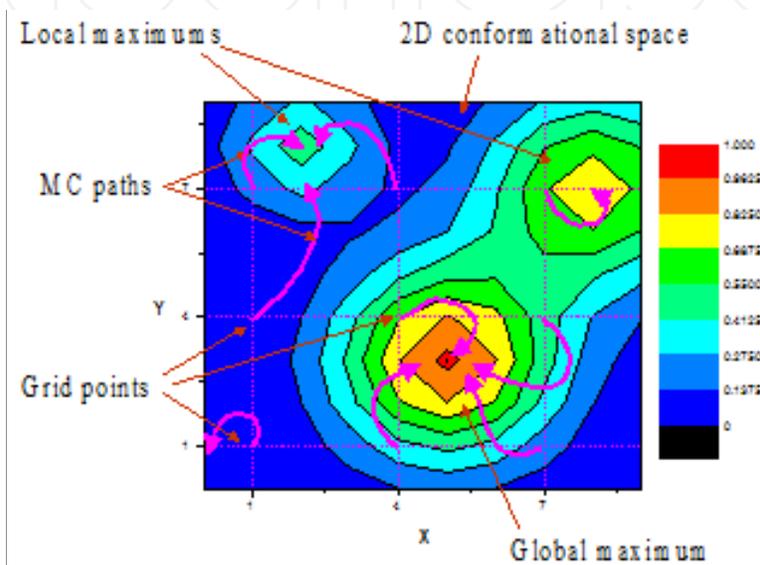


**Figure 3.** The core index distributions in schematic 2D maps of proteins A and B and their complex. Regions of protein density are colored red and green, respectively, and a region of protein B containing an inaccessible cavity as shown in light green. Regions outside of the protein are colored white. The numerical values of the core index for each grid point are labelled in each map. Bold numbers indicate the core indices of proteins A and B that change upon the formation of the AB complex.

These core-weighted correlation functions are designed to down-weight the regions overlapping with other components, while emphasizing the regions with no overlap. As explained above, the regions with significant overlap have small  $f_m$  and large  $f_n$ , and thus a small weighting function. By down-weighting the overlapping regions, the core-weighted correlation functions can minimize the overlap effect in predicting the correct fit.

To perform rigid fitting, one needs to search a conformational space with three translational and three rotational degrees of freedom. The grid-threading Monte Carlo (GTMC) search is a

combination of a grid search and Monte Carlo sampling [17]. The conformational space is split into searching grid points, which is not those defined for maps. Short Monte Carlo searches are performed starting from each of the searching grid points to identify local maxima close to the grid points. After searching out the local maxima, we can identify the global maximum among them. Figure 4 illustrates this procedure for a search in a 2D conformational space. The 2D conformational space is divided into a  $3 \times 3$  grid. Among all the local maxima, the local maximum at (5.3) has the highest correlation and is identified as the global maximum. More details of the core-weighted fitting method can be found elsewhere [2].



**Figure 4.** Illustration of the grid-threading Monte Carlo (GTMC) search in a 2D conformational space. The conformational space is split into a  $3 \times 3$  grid. From each of the 9 grid points, a short MC search (shown as purple curves) is performed to locate a nearby local maximum. Among these local maxima, the global maximum is identified. Only conformations along the 9 Monte Carlo paths are explored during a GTMC search.

## 2.2. Flexible fitting

When EM/ET maps have enough resolution to show structural difference from X-ray or NMR structures, flexible fitting should be applied. Flexible fitting allows conformational changes to adapt to the EM/ET maps. Low-resolution EM/ET maps often do not have enough structural information to uniquely define thousands of atomic positions of molecular systems. To make up the informational gap, an atomic force field is used to determine most of the degrees of freedom. An atomic force field defines covalent bonded terms, nonbonded interactions, as well as solvation-free energies. Theoretically, an ideal atomic force field aided with sufficient molecular simulation can determine molecular structures without the need of experimental information. However, current force fields are not accurate enough to determine molecular structures from scratch. In addition, biological systems are often too complicated to be accurately described by overly simplified simulation systems. A middle approach is to utilize an atomic force field to supplement the missing information in low-resolution maps in determining the structure of molecular systems. In other words, low-resolution maps can

provide structural information to compensate the inaccuracy in atomic force fields and X-ray or NMR structures provide reasonable initial conformations to simplify the conformational search in molecular simulation.

For a group of  $N$  atoms, at the target conformation, these atoms should reproduce the target map at given experimental condition and these atoms must all sit at high-density positions. Because atomic masses are closely correlated to their numbers of electrons, we can define a map-restraint potential as the total products between the atomic mass,  $m_a$  and the normalized density at the atom position,  $\hat{\rho}(x_a, y_a, z_a)$ :

$$E_{\text{map}} = -c_{\text{map}} \sum_a^N m_a \hat{\rho}(x_a, y_a, z_a) \quad (5)$$

The restraint constant,  $c_{\text{map}}$ , controls the strength of the map restraint. The units of variables,  $m_a$  and  $c_{\text{map}}$  are g/mol and kcal/g, respectively. A normalized map,  $\hat{\rho}(x_a, y_a, z_a)$ , has  $\langle \hat{\rho} \rangle = 0$  and  $\delta \hat{\rho} = 1$ . Equation (5) produces an energy landscape in the shape of the density distribution  $-c_{\text{map}} \hat{\rho}(x_a, y_a, z_a)$ . This energy landscape induces atoms to move to positions of lower energies, or of higher densities. Obviously, eq. (5) mimics a simplified correlation between atom masses and the map density distribution.

It should be noted that this map-restraint potential only captures the low-resolution characteristics of molecular systems and is not designed to reproduce atomic structures by itself. Instead, when aided with an all-atom force field, which contains bonded interactions (bond lengths, bond angles, dihedral angles) and nonbonded interactions (van der Waals, electrostatic interactions, solvation), the map-restraint potential can help to stabilize conformations that agree with the restraint map. It is the combination of a force field and the map restraint that drives an atomic system to the target conformation. This map-restraint potential has an order of  $O(N)$  and is very efficient to calculate as compared to other pair-wise nonbonded interactions, especially for large systems.

The map restraints drive atoms in a cooperative way so that atoms interacting through the force field contribute together to match the target map density distribution. The map restraints have the following convenient characteristics for a targeted conformational search:

1. The map-restraint energy surface is soft, which makes large-scale conformational transition feasible.
2. The map-restraint is atom-identity-blind, so the restraint energy calculation is of  $O(N)$ .
3. The map-restraint energy tolerates noises in target maps.
4. The map-restraint can be extended to maps of other properties such as partial charges, desolvation energies, and van der Waals interactions [18].

### 2.3. Targeted conformational search

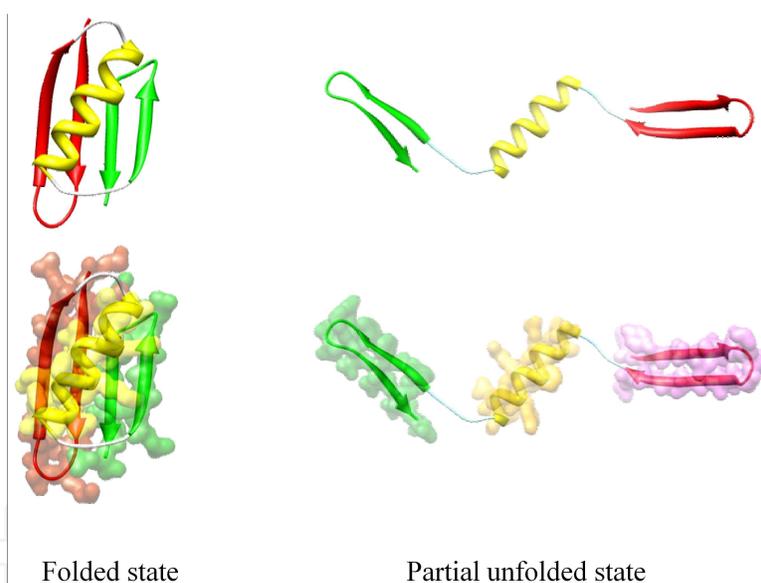
In some cases, there are large-scale conformational changes between X-ray/NMR structures and the EM/ET experimental systems. The self-guided Langevin dynamics (SGLD) simulation

method [19-22] is a method designed for an efficient search of conformational space. The equation of SGLD motion with map restraints has the following general form:

$$\dot{\mathbf{p}}_a = \mathbf{f}_a^{(\text{ff})} + \mathbf{f}_a^{(\text{map})} + \mathbf{g}_a - \gamma_a \mathbf{p}_a + \mathbf{R}_a \quad (6)$$

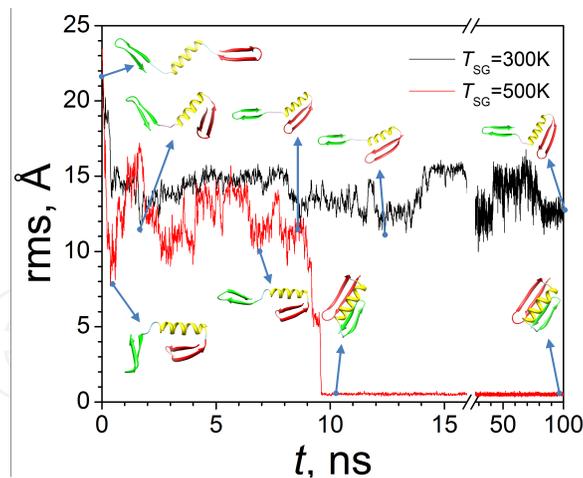
where  $\dot{\mathbf{p}}_a$  is the time derivative of momentum;  $\mathbf{f}_a^{(\text{ff})}$  is the atomic force due to the force field;  $\mathbf{f}_a^{(\text{map})}$  is the map-restraint force from the restraint map;  $\mathbf{R}_a$  is a random force, which is related to the mass,  $m_a$ , the collision frequency  $\gamma_a$ , and the simulation temperature  $T$ . There is a guiding force,  $\mathbf{g}_a$ , which is calculated based on the local average of momentum,  $\mathbf{p}_a$ , to accelerate conformational search through SGLD. Detailed description of the SGLD method and application can be found elsewhere [19, 21-23].

The method combining the map potential and the SGLD method is called MapSGLD [14]. This is a general method for targeted conformational search. Flexible fitting is a direct application of this method to efficiently identify conformations that match EM/ET maps.



**Figure 5.** Use restraint maps to maintain structures of GB1 substructures during conformational search.

Targeted conformational search is a convenient way to study dynamic properties of macromolecular systems. In protein-folding studies, it has been observed that secondary structure elements fold first, followed by their arrangement to form tertiary structures. To study how the secondary structure elements fold into the tertiary structures, it requires these secondary structures to be maintained during simulations. Similarly, for protein assemblies, it is desired to simulate how individual proteins assemble to form the complex structure while these individual proteins remain folded. These are typical examples of targeted conformational search. Figure 5 shows a small protein, the B1 domain of streptococcal protein G, abbreviated here as GB1, in its folded and unfolded states under the map restraints to maintain its



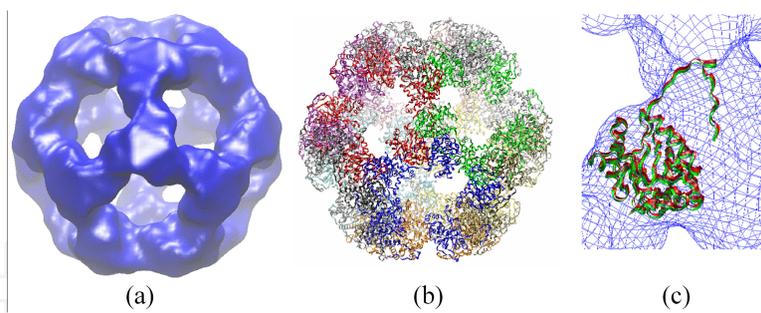
**Figure 6.** GB1 folding simulation by targeting the map of the folded state. The black line is for an LD simulation without enhanced sampling and the red line is for an SGLD simulation with an enhanced sampling.

secondary structures. GB1 has 56 residues with one  $\alpha$ -helix and one  $\beta$ -sheet. The  $\beta$ -sheet is made of two  $\beta$ -hairpins. Three restraint maps for the three secondary structure motifs were generated from the NMR structure: residues 1 to 19 for the N-terminal  $\beta$ -hairpin, residues 22 to 37 for the helix, and residues 42 to 56 for the C-terminal  $\beta$ -hairpin. Figure 6 shows two simulation results, one with  $T_{SG} = 300$  K and one with  $T_{SG} = 500$  K. When  $T_{SG} = 300$  K =  $T$ , it was a normal conformational search and the SGLD simulation reduced to a regular Langevin dynamics (LD) simulation, where the simulation failed to reach the folded conformation in up to 100 ns. While in the case of  $T_{SG} = 500$  K, conformational motion was accelerated, which prompted the protein to reach the folded state in 9.5 ns. In other words, the map restraint itself is not enough to bring the protein to the folded state in 100 ns, but SGLD can significantly accelerate the search to find a target conformation.

### 3. Applications

#### 3.1. Architecture of an icosahedral pyruvate dehydrogenase complex

The pyruvate dehydrogenase multienzyme complex couples the activity of three component enzymes (E1, E2, and E3) in the oxidative decarboxylation of pyruvate to generate acetyl-CoA, linking glycolysis and the tricarboxylic acid cycle. The *Bacillus stearothermophilus* PDH complex is assembled around a core of 60 dihydrolipoyl acetyltransferase (E2) chains arranged with icosahedral symmetry. Each E2 chain consists of three domains: (i) an N-terminal 9 kDa lipoyl domain, which visits the active sites of the pyruvate decarboxylase (E1) component and then those of E2 and dihydrolipoyl dehydrogenase (E3); (ii) a 4 kDa peripheral subunit-binding domain to which E1 and E3 bind tightly and mutually exclusively; and (iii) a C-terminal 28 kDa catalytic (acetyltransferase) domain, 60 copies of which assemble to form the icosahedral inner core. These domains are linked by stretches of extended, conformationally flexible polypeptide chain.

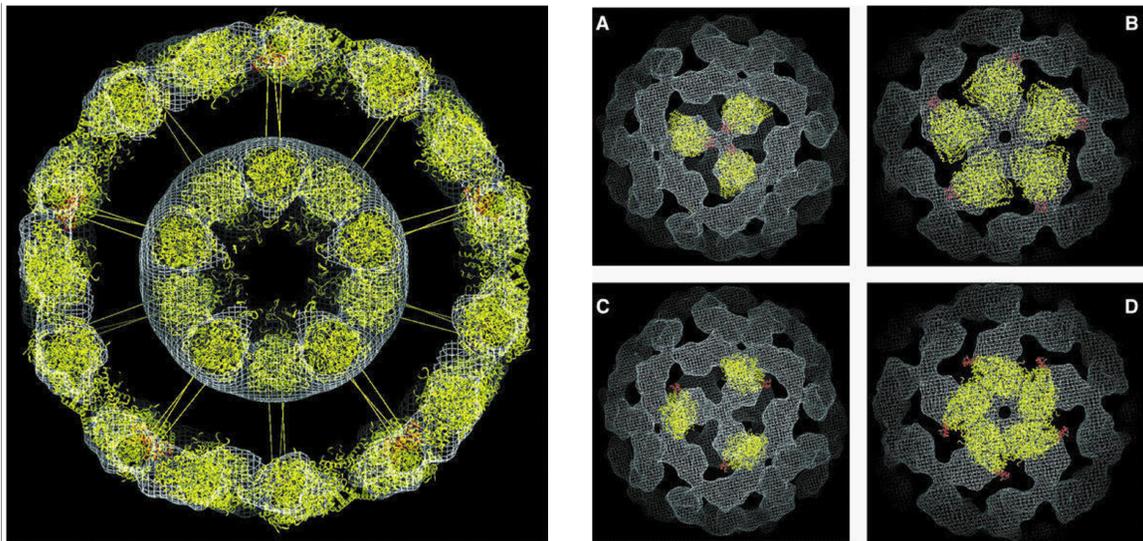


**Figure 7.** (a) The experimental map (at 14 Å resolution) of the icosahedral complex formed from 60 copies of the E2 catalytic domain of the pyruvate dehydrogenase. (b) The X-ray structure (PDB code: 1B5S) of the same complex. (c) Comparison of the location of the E2 catalytic domain obtained using a CW-GTMC search (green) with that from the X-ray structure (red) (rms = 2.13Å). The experimental EM map to fit is shown as blue wires.

Using the CW-GTMC method, we fit 60 copies of E2 into a 14 Å electron microscopic map of the icosahedral core of pyruvate dehydrogenase (Figure 7a), a 1.8 MDa complex comprised of 60 copies of the E2 catalytic domain, whose structure (Figure 7b) has been determined using X-ray crystallographic methods (PDB code: 1B5S). The CW-GTMC search with 5000 MC steps for each of 729 ( $3 \times 3 \times 3 \times 3 \times 3 \times 3$ ) grids identified all 60 global maximum positions [2]. Figure 7c compares the CW-GTMC fit and the corresponding X-ray structure position. The atomic structure of the E2 core from our rigid fit agrees well with the X-ray structure, validating the core-weighted GTMC method.

Since atomic coordinates of the *B. stearothermophilus* E1 enzyme are not available, the coordinates of the closely related E1  $\alpha_2\beta_2$  tetramer from *Pseudomonas putida* (PDB entry 1QS0) were used. From sequence alignment of the two enzymes, we found the presence of additional amino acid segments 42–51, 178–182, and 375–380 in the *P. putida* E1  $\alpha$ -subunit and 187–192 in its  $\beta$ -subunit, which are not present in the *B. stearothermophilus* E1 enzyme. By omitting these residues from the *P. putida* structure, we obtained a model for the *B. stearothermophilus* E1 enzyme. Using the core-weighted grid-threading Monte Carlo (CW-GTMC) method, we fit the coordinates to a 28 Å electron cryo-microscopy map [1]. The transformation space was divided into  $5 \times 5 \times 5$  grids and orientation space was divided into  $3 \times 3 \times 3$  grids. At each grid point, a 1000-step Monte Carlo search was performed with an initial transformational step size of 15 Å and initial rotational step size of 30°.

The automatic rigid fitting procedure identified two types of best-fit conformations into which all starting positions converged. The spatial relationship between these positions is indicated in the plot shown in Figure 8. In the second-best fit identified by the automated procedure, the E1 tetramer is pushed toward the fivefold vertex, covering some regions of the density not sampled by the best fit. In this orientation, the long axis of E1 is still along the surface of the icosahedron, but it is translated toward the fivefold vertex relative to the best fit. Analysis shows that the path from the best to the second-best fit involves a swivelling motion of E1 around an axis perpendicular to the icosahedral surface, with very little movement of the location and orientation of the E1 active site relative to the inner core. It is interesting that, while neither the best nor the second-best fit appears to span the outer density fully, the fits in Figure 8 show that almost all regions of the outer density are included in one or other of the



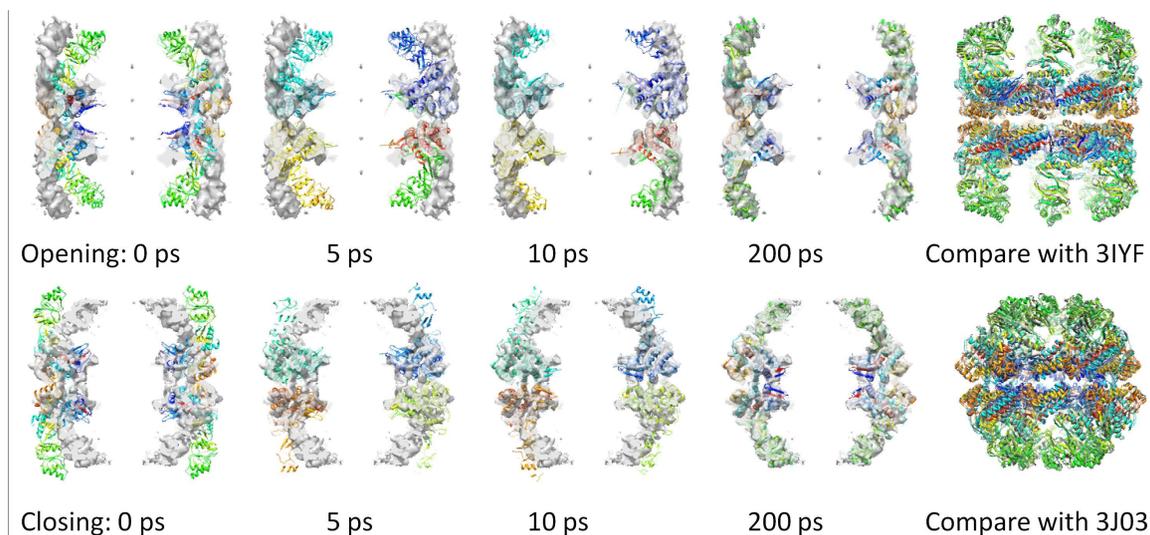
**Figure 8.** The E1E2 cryo-EM map at 28 Å resolution and the atomic structure obtained through CW-GTMC fitting. The left panel shows sectioned view of the density for the E1E2 complex with 60 docked copies of the atomic coordinates of the E1  $\alpha\beta\gamma\delta$  tetramer, the E2 catalytic domain, and the E2 peripheral subunit-binding domain. To illustrate the probable location of the linker domain connecting the peripheral subunit-binding and catalytic domains of the E2 chain, 60 radial spokes have been included. Right panels show the views of the docked E1 molecules (in yellow) and E2 peripheral subunit-binding domains (in red) from threefold (A and C) and fivefold (B and D) vertices. The views in (A) and (B) are for the best fit obtained by the CW-GTMC, while the views in (C) and (D) are for the second-best fit.

two orientations of E1. Based on these findings our working hypothesis is that the two orientations reflect stable positions for the E1 tetramer in the E1E2 complex. Since each E1 molecule in the complex may sample both orientations over time, the smear of the outer density in our 3D reconstruction probably reflects the averaged contribution to the structure from these two populations.

### 3.2. Dynamics of a group II chaperonin

The advance in electron microscopy experiment technology results in high-resolution maps, to show dynamic details of macromolecular systems, such as domain motion, loop rearrangements, and side chain reorientations. Here, we describe flexible fitting of EM maps through the MapSGLD method to address the dynamics of macromolecular systems [14]. One advantage of using SGLD for the targeted conformational search is to promote large-scale conformational changes necessary for protein functions. We chose the open–close transitions of a group II chaperonin to demonstrate the application of this method. By flexible fitting a group II chaperonin, mn-cpn, using maps from the EM databank: EMD-5138 (close state) and EMD-5140 (open state) [24], we obtained dynamics of the chaperonin-folding chamber opening and closing, as well as the structures in these states.

Zhang and colleagues have modelled the structures of the closed state (PDB:3J03) [25] and the open state (PDB:3IYF) [24]. We used these model structures as starting conformations and use the maps of the opposite states as restraints to perform MapSGLD simulations to examine the capability of MapSGLD to identify target conformations.



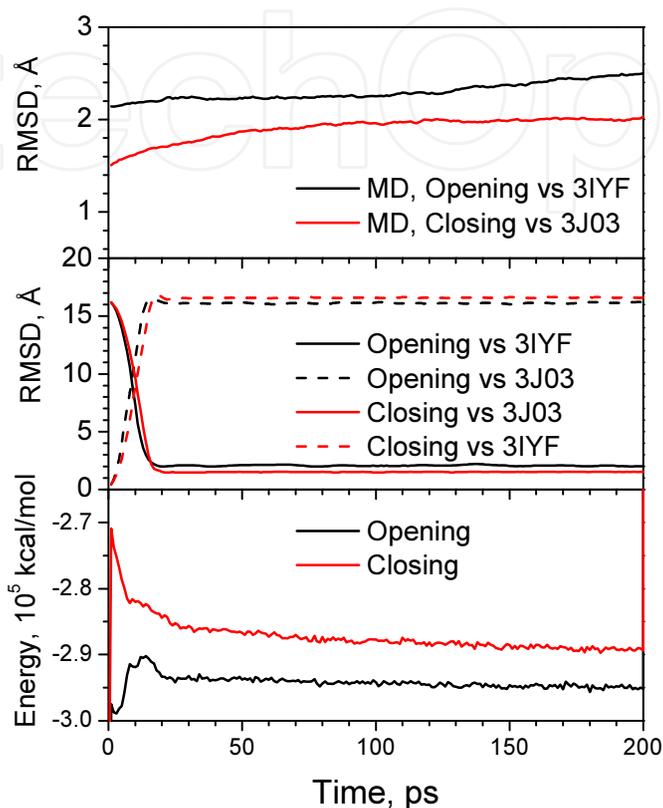
**Figure 9.** Opening (top) and closing (bottom) of the chaperonin-folding chamber during the MapSGLD simulations. The last column compares the final conformations with their corresponding PDB entries (in grey). The opening simulation started from the closed state (PDB: 3J03) and was restrained with the map at the open state: EMD-5140. The closing simulation started from the open state structure (PDB: 3IYF) and was restrained with the map at the closed state: EMD-5138.

Each monomer of the chaperonin assembly contains three domains, the apical domain (residues 205–334), the intermediate domain (residues 136–204, 335–371) and the equatorial domain (residues 1–135, 372–491). We used movable map restraints to maintain the structures of these domains. The movable map restraints were generated from the initial conformation with a resolution of 3 Å and a map-restraint constant of 0.1 kcal/g. The experimental EM maps were applied to the system as fixed map restraints with a restraint constant of 0.05 kcal/g. Figure 9 shows the conformations during the opening and closing of the folding chamber obtained from MapSGLD simulations. We can see the apparent difference in the starting conformation from the EM maps and the agreement of the final conformations with their restraint maps. The root-mean-square-deviation (rmsd) of the final conformation from the closing simulation was within 1.5 Å from the closed structure (PDB:3J03) and the rmsd of the final conformation from the opening simulation was within 2.0 Å from the opened structure (PDB:3IYF), demonstrating the reliability of the flexible fitting method. From the simulation trajectories we can extract atomic details of the folding chamber dynamics during opening and closing.

Figure 10 shows the energies and rmsd during these MapSGLD simulations. From the middle panel of Figure 10, we can see that the closing simulation reached within 1.51 Å from the closed structure in 20 ps and remained there with little change afterward. Also we can see that the opening simulation reached within 2.01 Å from the opened structure in 20 ps and remained there with little change afterward.

Examining the energy profiles shown in the bottom panel of Figure 10, we can see that in order for the closed assembly to open up, it first went through an energy barrier. It reached a peak

at about 15ps, and began an energy decrease throughout the rest of the simulation. Throughout the closing simulation, there was no energy barrier. It was the energy barrier overcoming ability of SGLD made it efficient to overcome of the energy barrier during the opening process.

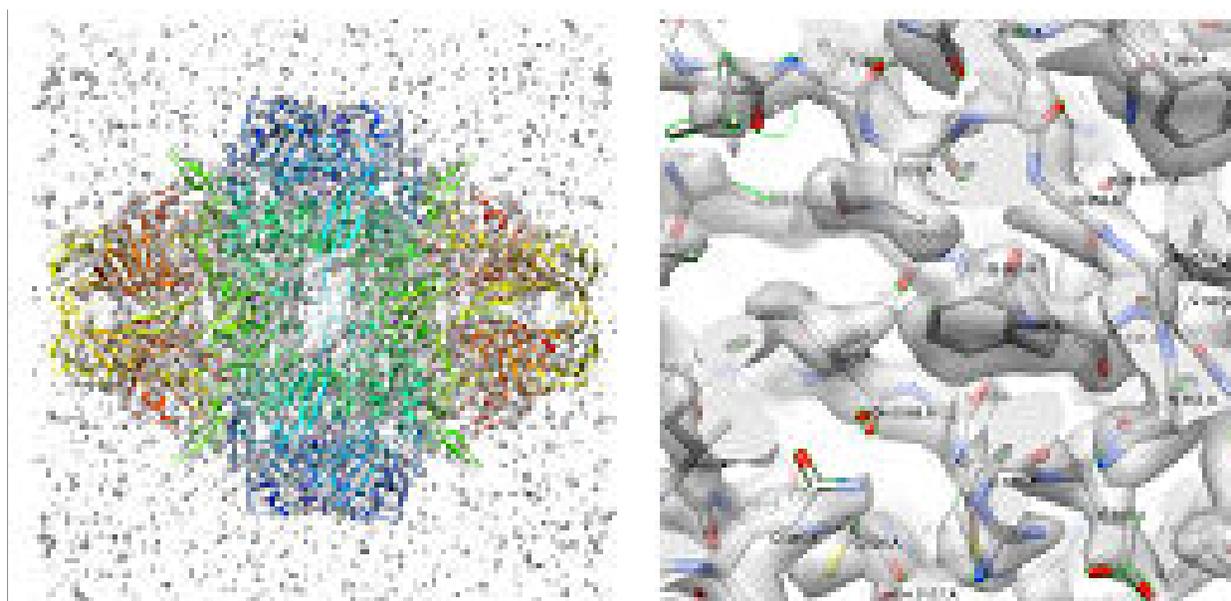


**Figure 10.** The opening and closing simulation profiles. Bottom panel: the molecular potential energy profiles during the MapSGLD simulations. Middle panel: the rmsd profiles during the MapSGLD simulations. Top panel: the rmsd profiles during the MD simulations continuing from the final conformations of the MapSGLD simulations.

Driven by the map-restraint potential, MapSGLD simulations search for conformations matching the EM maps. It is the map-restraint potential that makes the targeted conformation a global free-energy minimum state. Without the map-restraint potential, experimental structures may not be the global free-energy minimum states at this simplified simulation conditions. There are many factors that affect the global minimum, such as inaccuracies in the force field, overly simplified setup of the simulated system, or inadequate description of the effects of the solvent. To illustrate this point, we performed conventional molecular dynamics simulations without the map restraint from the final conformations of the MapSGLD simulations, and the rmsd profiles are shown in the top panel of Figure 10. We can see that in the MD simulations, the systems slowly drifted away from the corresponding states, which indicates that the simplification in simulation conditions can cause conformational deviations from experimental observations and the map restraint can help overcome the effect of the simplification in simulation conditions.

### 3.3. High-resolution structure determination from EM maps

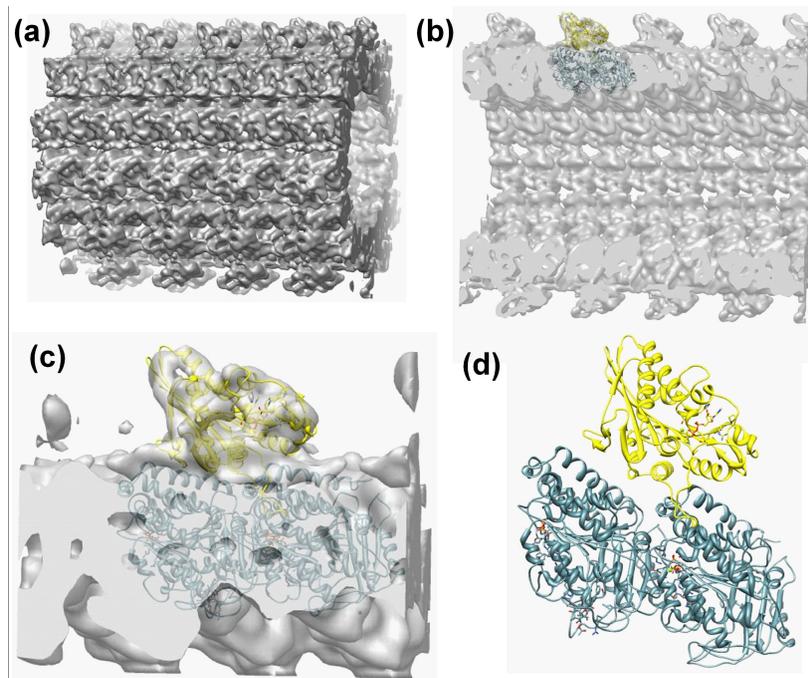
A milestone in high-resolution structure determination using cryo-electron microscopy (cryo-EM) has been reached recently [26] through the structure of a complex between *Escherichia coli*  $\beta$ -galactosidase and the cell-permeant inhibitor phenylethyl  $\beta$ -D-thiogalactopyranoside (PETG), determined by cryo-EM at an average resolution of  $\sim 2.2$  angstroms ( $\text{\AA}$ ). Besides the PETG ligand,  $\sim 800$  water molecules and for magnesium and sodium ions are identified in the map. To achieve resolutions close to  $2 \text{ \AA}$  using single-particle cryo-EM, preparation of specimens of adequate quality and intrinsic protein flexibility, rather than imaging or image-processing technologies, are the major bottlenecks. Figure 11 shows the EM map and the protein structure fit into the map. The right panel shows side chains and their map density. At this resolution, the EM density of heavy atoms is recognizable.



**Figure 11.** The  $2.2\text{\AA}$  resolution EM density map of a complex between *Escherichia coli*  $\beta$ -galactosidase and the cell-permeant inhibitor phenylethyl  $\beta$ -D-thiogalactopyranoside (PETG) and the flexible fit structure of the complex. The left panel shows the whole complex (as ribbon) and the right panel shows residues around residue A355 shown (sticks).

### 3.4. Kinesin walking on microtubule [27]

To understand kinesin walking mechanism, it is desirable first to see how kinesin complexes with microtubule. Due to the heterogeneity of the kinesin–microtubule complex, it is difficult to determine the complex structure through X-ray crystallography. Instead, the complex models are often built from low-resolution electron microscopy maps. The kinesin–microtubule complex system was built by fitting the kinesin motor domains (1BG2 or 2NCD) and tubulin (1JFF) into the complex EM map (EMD-5011). Figure 12 illustrates the structure obtained by the CW-GTMC rigid fitting. Rigid fitting provides a reasonable first guess of the complex structure at the limit of the map resolution. Their surface elements, such as loops and

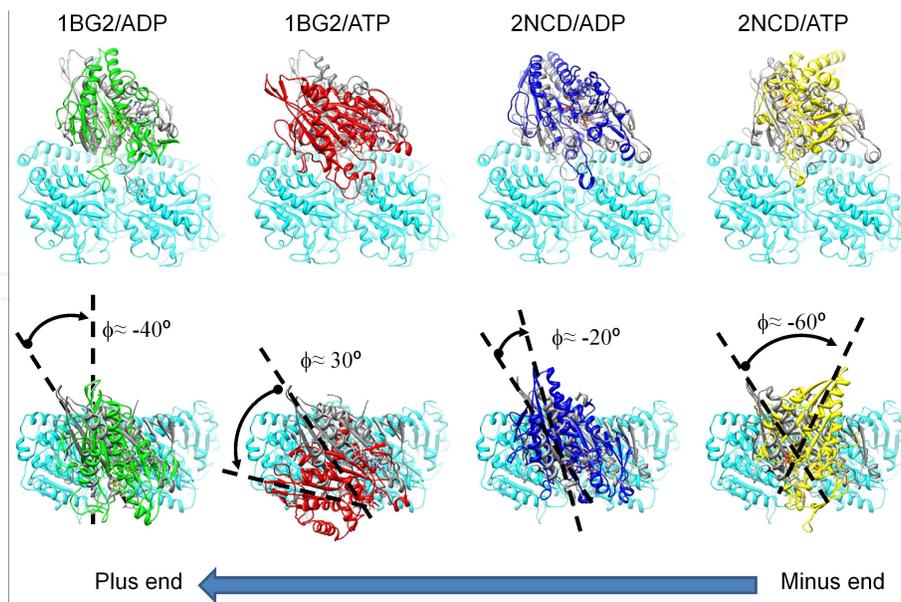


**Figure 12.** Fitting X-ray structures of kinesin (PDB: 1BG2 or 2NCD) (yellow) and microtubule (PDB: 1JFF) (cyan) into the 8 Å resolution EM map of a kinesin–microtubule complex to obtain initial conformations of the complex. (a) The EM map EMD-5011. (b) A cross-section view of the fitting result. (c) An enlarged view of the fitting. (d) The complex structure from the fitting.

side chains, must adjust to suit their environment. Energy minimizations were performed on these fitting results to derive energetically favorable structures.

Based on the EM fit complex structures, SGLD simulations were performed to study the dynamics of these complexes at different nucleotide states: 1BG2/ADP, 1BG2/ATP, 2NCD/ADP, and 2NCD/ATP. SGLD simulations showed significant conformational change and the final conformations are compared with the initial fit structures in Figure 13. As can be seen clearly, the motor domains rotated certain angles in related to their initial positions. After the SGLD simulations, 1BG2 in the ADP-binding state rotated about  $-40^\circ$  and in the ATP-binding state rotated about  $30^\circ$ . 2NCD in the ADP-binding state rotated approximately  $-20^\circ$  and in the ATP-binding state rotated approximately  $-60^\circ$ . In other words, 1BG2 rotated count clockwise for about  $70^\circ$ , while 2NCD rotated clockwise for about  $40^\circ$ , when changing from the ADP-binding state to the ATP-binding state. This directional difference in rotation clearly corresponds to the directional walking of the two types of kinesin. Because the coiled-coil linker to cargo extends near the C-terminal of the  $\beta_6$  strand, a counter clockwise rotation will result in a plus-end-directed movement of the linker, and a clockwise rotation will result in a minus-end-directed movement of the linker. This kinesin-specific directional rotation provides clues to understanding the mechanism of kinesin walking directions.

These application examples illustrate the atomic structure derivation procedures through rigid fitting and flexible fitting. These derived atomic structures open a gateway for further understanding of dynamics and functions of molecular systems.



**Figure 13.** The kinesin–microtubule complexes before and after the SGLD simulations. The initial kinesin structures are shown in grey and microtubules are colored cyan. 1BG2/ADP, 1BG2/ATP, 2NCD/ADP, and 2NCD/ATP are colored green, red, blue, and yellow, respectively. The upper row shows their side views and the bottom row shows their top views. Rotation angles are calculated with the  $\beta 6$  strand in the kinesin motor domains against the initial conformation.

## Acknowledgements

This research was supported by the Intramural Research Programs of National Heart, Lung, and Blood Institute (Z01 HL001027-30). Katherine Wu of Massachusetts Institute of Technology performed the kinesin-microtubule simulation.

## Author details

Xiongwu Wu\* and Bernard R. Brooks

\*Address all correspondence to: wuxw@nhlbi.nih.gov

Laboratory of Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, USA

## References

- [1] Milne, J. L., Shi, D., Rosenthal, P. B., Sunshine, J. S., Domingo, G. J., Wu, X., Brooks, B. R., Perham, R. N., Henderson, R., and Subramaniam, S. Molecular architecture and

- mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine, *Embo J* (2002) 21, 5587-5598.
- [2] Wu, X., Milne, J. L., Borgnia, M. J., Rostapshov, A. V., Subramaniam, S., and Brooks, B. R. A core-weighted fitting method for docking atomic structures into low-resolution maps: application to cryo-electron microscopy, *J Struct Biol* (2003) 141, 63-76.
- [3] Roseman, A. M. Docking structures of domains into maps from cryo-electron microscopy using local correlation, *Acta Crystallogr D Biol Crystallogr* (2000) 56, 1332-1340.
- [4] Spahn, C. M., Penczek, P. A., Leith, A., and Frank, J. A method for differentiating proteins from nucleic acids in intermediate-resolution density maps: cryo-electron microscopy defines the quaternary structure of the Escherichia coli 70S ribosome, *Structure* (2000) 8, 937-948.
- [5] Antzutkin, O. N., Leapman, R. D., Balbach, J. J., and Tycko, R. Supramolecular structural constraints on Alzheimer's beta-amyloid fibrils from electron microscopy and solid-state nuclear magnetic resonance, *Biochemistry* (2002) 41, 15436-15450.
- [6] Wriggers, W., and Birmanns, S. Using situs for flexible and rigid-body fitting of multi-resolution single-molecule data, *J Struct Biol* (2001) 133, 193-202.
- [7] Wriggers, W., Milligan, R. A., and McCammon, J. A. Situs: A package for docking crystal structures into low-resolution maps from electron microscopy, *J Struct Biol* (1999) 125, 185-195.
- [8] Milne, J. L., Wu, X., Borgnia, M. J., Lengyel, J. S., Brooks, B. R., Shi, D., Perham, R. N., and Subramaniam, S. Molecular structure of a 9-MDa icosahedral pyruvate dehydrogenase subcomplex containing the E2 and E3 enzymes using cryoelectron microscopy, *J Biol Chem* (2006) 281, 4364-4370.
- [9] Tama, F., Miyashita, O., and Brooks, C. L., 3rd. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM, *J Struct Biol* (2004) 147, 315-326.
- [10] DiMaio, F., Tyka, M. D., Baker, M. L., Chiu, W., and Baker, D. Refinement of protein structures into low-resolution density maps using rosetta, *J Mol Biol* (2009) 392, 181-190.
- [11] Bradley, P., Misura, K. M. S., and Baker, D. Toward high-resolution de Novo structure prediction for small proteins, *Science* (2005) 309, 1868-1871.
- [12] Trabuco, L. G., Villa, E., Mitra, K., Frank, J., and Schulten, K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics, *Structure* (2008) 16, 673-683.
- [13] Trabuco, L. G., Villa, E., Schreiner, E., Harrison, C. B., and Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography, *Methods* (2009) 49, 174-180.

- [14] Wu, X., Subramaniam, S., Case, D. A., Wu, K. W., and Brooks, B. R. Targeted conformational search with map-restrained self-guided Langevin dynamics: Application to flexible fitting into electron microscopic density maps, *J Struct Biol* (2013) 183, 429-440.
- [15] Case, D. A., Babin, V., Berryman, J. T., Betz, R. M., Cai, Q., Cerutti, D. S., Cheatham Iii, T. E., Darden, T. A., Duke, R., Gohlke, H., Gusarov, S., Homeyer, N., Janowski, P., Kaus, J., Kolossvary, I., Kovalenko, A., Lee, T. S., LeGrand, S., Luchko, T., Luo, R., Madej, B., Merz, K. M., Paesani, F., Roe, D. R., Roitberg, A., Sagui, C., Salomon-Ferrer, R., Seabra, G., Simmerling, C. L., Smith, W., Swails, J., Walker, R. C., Wan, J., Wolf, R. M., Wu, X., and Kollman, P. A. AMBER 14, University of California, San Francisco (2014).
- [16] Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoseck, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. CHARMM: the biomolecular simulation program, *J Comput Chem* (2009) 30, 1545-1614.
- [17] Allen, M. P., and Tildesley, D. J. *Computer Simulations of Liquids*, Clarendon Press, Oxford (1987).
- [18] Wu, X., and Brooks, B. R. Modeling of Macromolecular assemblies with map objects., *Proceedings of the 2007 international conference on bioinformatics & computational biology*. (2007) II, 411-417.
- [19] Wu, X., and Brooks, B. R. Self-guided Langevin dynamics simulation method, *Chem Phys Lett* (2003) 381, 512-518.
- [20] Wu, X., and Brooks, B. R. Force-momentum-based self-guided Langevin dynamics: a rapid sampling method that approaches the canonical ensemble, *J Chem Phys* (2011) 135, 204101.
- [21] Wu, X., and Brooks, B. R. Toward canonical ensemble distribution from self-guided Langevin dynamics simulation, *J Chem Phys* (2011) 134, 134108.
- [22] Wu, X., Damjanovic, A., and Brooks, B. R. Efficient and Unbiased Sampling of Biomolecular Systems in the Canonical Ensemble: A Review of Self-Guided Langevin Dynamics, In *Advances in Chemical Physics* (Rice, S. A., and Dinner, A. R., Eds.), pp. 255-326, John Wiley & Sons, Inc., Hoboken (2012).
- [23] Wu, X., and Wang, S. Self-guided molecular dynamics simulation for efficient conformational search, *J Phys Chem B* (1998) 102, 7238-7250.

- [24] Zhang, J., Baker, M. L., Schroder, G. F., Douglas, N. R., Reissmann, S., Jakana, J., Dougherty, M., Fu, C. J., Levitt, M., Ludtke, S. J., Frydman, J., and Chiu, W. Mechanism of folding chamber closure in a group II chaperonin, *Nature* (2010) 463, 379-383.
- [25] Zhang, J., Ma, B., DiMaio, F., Douglas, Nicholai R., Joachimiak, Lukasz A., Baker, D., Frydman, J., Levitt, M., and Chiu, W. Cryo-EM structure of a Group II chaperonin in the prehydrolysis ATP-bound state leading to lid closure, *Structure* (2011) 19, 633-639.
- [26] Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J. L. S., and Subramaniam, S. 2.2 Å resolution cryo-EM structure of  $\beta$ -galactosidase in complex with a cell-permeant inhibitor, *Science* (2015) 348, 1147-1151.
- [27] Wu, K. W. The hydrolysis engine concept for motor proteins, In *Society for Science and The Public*, p. 20, Washington, DC (2013).