

## Structure 16

### Supplemental Data

# Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics

Leonardo G. Trabuco, Elizabeth Villa, Kakoli Mitra, Joachim Frank, and Klaus Schulten

## S1. Molecular Dynamics Simulations

Molecular dynamics (MD) simulation is a powerful computational method that describes in atomic detail the behavior of a macromolecular system over time solving the equations of classical mechanics (Frenkel and Smit, 2002). Given a set of initial coordinates for all atoms in a system, often obtained from an X-ray crystallography or nuclear magnetic resonance (NMR) experiment, the positions  $\mathbf{r}_i$  and velocities  $\mathbf{v}_i$  of all atoms  $i$  are calculated by integrating a discrete form of Newton's equations of motion, with each atom represented by a point mass  $m_i$ .

The interaction among the atoms in the system is described by a potential energy function  $U_{MD}(\mathbf{R})$ , where  $\mathbf{R} = \mathbf{r}_1, \mathbf{r}_2 \dots \mathbf{r}_N$  is the collection of all atomic coordinates. The potential energy function is composed of contributions that can be classified into bonded and nonbonded interactions. The bonded terms describe interactions between covalently-bonded atoms, and include  $U_{bond}$  which represents high-frequency vibrations along covalent bonds;  $U_{angle}$  which represents bending motions between two adjacent bonds;  $U_{dihedral}$  which represents torsional motions around a bond; and  $U_{improper}$  which describes the planar orientation of one atom relative to three others. The nonbonded terms describe interactions between atoms that are not covalently bonded or that are separated by three or more covalent bonds, and include  $U_{vdW}$ , the pairwise van der Waals energy, modeled as a 6-12 Lennard-Jones potential, and  $U_{elec}$ , the pairwise Coulomb energy. The form of the empirical potential energy function as used, *e.g.*, in the CHARMM force field (MacKerell et al., 1998; Foloppe and MacKerell Jr., 2000; MacKerell Jr. and Banavali, 2000; MacKerell et al., 2004), is

$$\begin{aligned}
U_{MD}(\mathbf{R}) = & \underbrace{\sum_{bonds} k_{\alpha} (r_{\alpha} - r_{\alpha}^0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_{\beta} (\theta_{\beta} - \theta_{\beta}^0)^2}_{U_{angle}} \\
& + \underbrace{\sum_{dihedrals} k_{\gamma} ([1 + \cos(n_{\gamma} \psi_{\gamma} + \delta_{\gamma})])}_{U_{dihedral}} + \underbrace{\sum_{impropers} k_{\delta} (\phi_{\delta} - \phi_{\delta}^0)^2}_{U_{improper}} \\
& + \underbrace{\sum_i \sum_{i \neq j} 4 \varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]}_{U_{vdW}} + \underbrace{\sum_i \sum_{i \neq j} \frac{q_i q_j}{4 \pi \varepsilon_0 \varepsilon r_{ij}}}_{U_{elec}}
\end{aligned} \tag{S.1}$$

where  $r_{\alpha}$  is a bond length,  $r_{\alpha}^0$  the associated equilibrium bond length,  $k_{\alpha}$  the respective bond spring constant;  $\theta_{\beta}$  is the angle between two bonds,  $\theta_{\beta}^0$  the associated equilibrium angle,  $k_{\beta}$  the respective angular spring constant;  $\psi_{\gamma}$  is the angle of rotation around a bond,  $k_{\gamma}$  the rotational spring constant,  $n_{\gamma}$  the number of maxima or minima in one full rotation, and  $\delta_{\gamma}$  the angular offset;  $\phi_{\delta}$  is the so-called improper torsion angle with associated equilibrium angle  $\phi_{\delta}^0$  and spring constant  $k_{\delta}$ ;  $\varepsilon_{ij}$  is the depth of the energy minimum for  $U_{vdW}$ , the minimum located at  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j| = 2^{1/6} \sigma_{ij}$ ;  $q_i$  and  $q_j$  are the atomic partial charges separated by a distance  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ ,  $\varepsilon_0$  is the permittivity of free space, and  $\varepsilon$  is the dielectric constant of the medium in which the charges are placed. The sums in  $U_{vdW}$  and  $U_{elec}$  are performed over all atom pairs ( $i, j$ ).

The force applied to an atom resulting from this potential energy function in a simulation time step is defined by

$$\mathbf{f}_i = -\frac{\partial}{\partial \mathbf{x}_i} U_{MD}(\mathbf{R}) + \mathbf{f}_i^{ext}, \tag{S.2}$$

where  $\mathbf{f}_i^{ext}$  can be any external force applied to the atom in the simulation.

The parameters for the force field defined through Eq. (S.1), *e.g.*, equilibrium values for bonds and angles, are obtained through a calibration of experimental results and quantum mechanical calculations of small molecules. The energy parameters used in the examples throughout this paper are those of the CHARMM27 force field. The MD software used is NAMD (Phillips et al., 2005).

## S2. 3-D EM reconstructions

Single-particle cryo-EM uses electron micrographs of single (*i.e.*, unattached), usually large, multimeric molecules flash-frozen in vitreous ice. The method complements other structure-determination methods, such as X-ray crystallography and NMR spectroscopy. The former is the most prolific structural method and can determine structures from small molecules to large assemblies of hundreds of kilodaltons (Dutta and Berman, 2005). The limitations of X-ray crystallography stem largely from the necessity to crystallize the complex of interest. Thus, it is difficult to study complexes that (i) are highly dynamic and resist crystal formation, or (ii) cannot be purified in a homogeneous state in large quantities. Even when crystals can be obtained the functional state of the complex might be undefined. NMR can capture both structural and dynamical information of macromolecules, but is limited to sizes of tens of kilodaltons (Clare and Gronenborn, 1998).

Many cellular processes involve assemblies of macromolecules, resulting in sizes challenging for NMR and X-ray crystallography. As the interaction between the components of a complex changes over time, knowledge of different conformational states is crucial to understanding function. Single-particle cryo-EM is suitable for imaging large complexes, presenting various advantages: large assemblies can be studied; crystallization is not required; small amounts of material are sufficient to obtain data; and the particles imaged can be trapped in transient or functional states in their natural environment. Examples of macromolecular assemblies for which cryo-EM has revealed structures in different states are poliovirus (Belnap et al., 2000), ribosome (Frank and Spahn, 2006), and ATPase (Roseman et al., 1996). In principle, 3-D cryo-EM single-particle reconstructions could yield atomic resolution, but practical limitations, including electron-optical aberrations (Penczek et al., 2006), prevent achieving resolutions better than 3 Å (Chiu et al., 2005). Presently, single-particle cryo-EM maps can be routinely obtained at resolutions of 10-20 Å (Saibil, 2000).

The single-particle cryo-EM 3-D reconstruction technique uses 2-D images of a macromolecular complex obtained by a transmission electron microscope, where the sample has been blotted and rapidly plunge-frozen, generating a very thin specimen with molecules arranged in random orientations trapped in the same defined conformational state. Typically, tens to hundreds of thousands of images, or projections, are obtained and subsequently classified according to their orientation; these projections are combined to reconstruct a three-dimensional image, or density map, of the biomolecule. For details on single-particle three-dimensional reconstruction from cryo-EM data, the reader is referred to Frank (2006). In what follows, we discuss the features present in the final reconstructions that are relevant to the flexible-fitting methodology.

The process of image formation by a transmission electron microscope is well understood (Spence, 1988; Reimer, 1997). The electrons in the beam impinging on the sample interact with both the nuclei and the electrons in the atoms of the sample. Different types of interactions occur, resulting in elastic and inelastic scattering. The former provides the structural information used in 3-D reconstructions, while the latter causes radiation damage and background noise (Reimer, 1997). Image formation for very thin biological samples in bright-field transmission EM can be described by the weak-phase approximation, according to which the image contrast resulting from the interference of unscattered and elastically scattered electrons (Frank, 2006) is linearly related to the 2-D projection of the Coulomb potential of the atoms in the sample (Spence, 1988). When the sample is reconstructed in three dimensions, the data reported by the cryo-EM density map is a reconstruction of the three-dimensional Coulomb potential  $\Phi(\mathbf{r})$  generated by the nuclei and the electrons of the sample. The main contribution to this potential comes from the nuclei; the potential from a given nucleus is proportional to the atomic number  $Z$  of that atom, *i.e.*, the number of protons in its nucleus. For atoms contained in biomolecules, this number is roughly proportional to the atomic mass; thus, in principle, a mass density distribution of the macromolecular assembly can be obtained from cryo-EM data (Langmore and Smith, 1992).

The weak-phase approximation is widely used, but it is worth noting that it disregards other contributions to image formation, including those from inelastically scattered electrons, that can be eliminated during data collection (energy filtering), and from elastically scattered electrons that are lost before participating in image formation due to subsequent inelastic collisions or due to scattering outside the objective aperture (amplitude contrast). The latter has a strong dependency on the atomic species, and can be neglected for low atomic number nuclei as occur in biological specimens, but becomes important when comparing the EM map to simulated maps derived from X-ray structures of molecules composed of atoms with strongly different atomic numbers, such as RNA-protein complexes. In practice, the description of image formation has to take lens aberrations into account, which result in a distortion of the image intensity, described in Fourier space by the so-called contrast transfer function (CTF) (Spence, 1988). The images can be post-processed to correct for the CTF in a systematic way (Frank, 2006).

The Coulomb potential is reported in the reconstruction on a discrete 3-D grid. It is important to note that the magnitude of the values reported varies considerably from map to map, which should be considered when defining an external potential for the MD simulations from the EM data, as described in the text. A source of such variability arises during the reconstruction process, when a scaling factor is chosen by the reporting authors by which all grid point values,

or voxel values, are multiplied for convenience in visualization and data storage. Another source of discrepancy between different EM maps concerns negative values reported. Generally, a zero value is assigned to the average intensity using optical-density scaling (Frank, 2006), with all values below that threshold becoming negative; however, some maps also report data arising from negative staining experiments, and in these, negative values represent regions which are dominated by accumulation of stain.

### S3. Fitting atomic structures into simulated maps

**Table S1.** Effect of resolution and grid spacing on flexible fitting into noise-free simulated maps using MDFF. Further test cases are presented; for each system, maps were created computationally from a given crystal structure of conformation I and a crystal structure available in an alternative conformation, II, was fitted into the computed map. The initial backbone RMSD and cross-correlation coefficients correspond to rigid-body docked structures into computed maps with grid spacing of 2.0 Å using Situs with default options. The final mean backbone RMSD values were calculated from the last 200 ps of 500-ps trajectories. The final mean cross-correlation coefficients were calculated using computed maps generated from the average of the trajectories. Corresponding local cross-correlation coefficients that consider only the molecular envelope are also shown in parenthesis (a threshold of  $0.2\sigma$  was used in these examples).

System (PDB codes)	Resolution (Å)	Initial RMSD (Å)	Initial (local) correlation	Grid spacing (Å)	Final mean RMSD (Å)	Final mean (local) correlation
hammerhead ribozyme (1HMH)	5.0	3.76	0.855 (0.630)	1.0	0.60	0.992 (0.962)
				2.0	0.81	0.986 (0.938)
	10.0	3.54	0.949 (0.854)	1.0	0.90	0.998 (0.993)
				2.0	0.96	0.996 (0.987)
	15.0	3.73	0.972 (0.926)	1.0	1.38	0.997 (0.992)
				2.0	1.49	0.996 (0.987)
mRNA-capping enzyme (1CKN)	5.0	3.83	0.897 (0.623)	1.0	0.73	0.987 (0.934)
				2.0	0.89	0.985 (0.918)
	10.0	3.71	0.950 (0.817)	1.0	1.19	0.994 (0.973)
				2.0	1.26	0.994 (0.970)
	15.0	3.58	0.970 (0.901)	1.0	1.80	0.994 (0.981)
				2.0	2.04	0.993 (0.978)
biotin carboxylase (1DV2, 1BNC)	5.0	4.70	0.890 (0.648)	1.0	0.66	0.981 (0.904)
				2.0	0.91	0.980 (0.883)
	10.0	4.59	0.933 (0.786)	1.0	1.41	0.993 (0.969)
				2.0	1.62	0.992 (0.960)
	15.0	4.10	0.961 (0.876)	1.0	2.06	0.995 (0.984)
				2.0	2.06	0.994 (0.976)
dipeptide binding protein (1DPP, 1DPE)	5.0	12.88	0.718 (0.371)	1.0	0.70	0.997 (0.990)
				2.0	1.07	0.985 (0.901)
	10.0	12.22	0.796 (0.530)	1.0	1.14	0.996 (0.980)
				2.0	1.22	0.995 (0.975)
	15.0	6.97	0.947 (0.827)	1.0	1.54	0.997 (0.990)
				2.0	1.65	0.995 (0.983)
<i>Table S1 – continued.</i>						

System (PDB codes)	Resolution (Å)	Initial RMSD (Å)	Initial (local) correlation	Grid spacing (Å)	Final mean RMSD (Å)	Final mean (local) correlation
PvuII endonuclease (1PVI, 1PVU)	5.0	3.95	0.862 (0.635)	1.0	0.86	0.983 (0.927)
				2.0	1.02	0.984 (0.923)
	10.0	3.03	0.955 (0.876)	1.0	1.33	0.995 (0.880)
				2.0	1.57	0.994 (0.980)
	15.0	3.06	0.977 (0.942)	1.0	1.98	0.994 (0.985)
				2.0	1.92	0.994 (0.984)
Ap4A hydrolase (1F3Y, 1JKN)	5.0	4.05	0.869 (0.633)	1.0	1.49	0.975 (0.893)
				2.0	1.75	0.973 (0.866)
	10.0	3.82	0.944 (0.869)	1.0	2.28	0.990 (0.965)
				2.0	2.53	0.990 (0.966)
	15.0	3.96	0.978 (0.945)	1.0	2.70	0.992 (0.978)
				2.0	2.80	0.991 (0.976)
catabolite activator protein (1G6N)	5.0	3.30	0.914 (0.727)	1.0	0.63	0.989 (0.952)
				2.0	0.85	0.986 (0.937)
	10.0	2.79	0.974 (0.915)	1.0	1.08	0.994 (0.978)
				2.0	1.17	0.993 (0.969)
	15.0	2.76	0.987 (0.960)	1.0	1.90	0.993 (0.981)
				2.0	1.86	0.993 (0.979)
adenylate kinase (1AKE, 4AKE)	5.0	8.22	0.685 (0.374)	1.0	0.72	0.986 (0.936)
				2.0	0.93	0.984 (0.924)
	10.0	8.19	0.780 (0.517)	1.0	1.24	0.995 (0.981)
				2.0	1.24	0.995 (0.978)
	15.0	7.64	0.848 (0.640)	1.0	1.71	0.996 (0.988)
				2.0	1.79	0.995 (0.986)

**Table S2.** Accuracy of MDDF using the benchmark set suggested by Topf et al. (2008). The

benchmark set (initial structures and density maps) was downloaded from the website <http://salilab.org/Flex-EM/>. For each system, the initial structure was rigid-body docked with Situs, and MDFF was applied with the same settings as the other test cases described in this paper. The final mean  $C_{\alpha}$  RMSDs with respect to the target structures were calculated from the last 300 ps of 500-ps trajectories, unless otherwise stated.

System (PDB code, range)	Resolution (Å)	Initial $C_{\alpha}$ RMSD (Å) <sup>a</sup>	Final mean $C_{\alpha}$ RMSD (Å) <sup>a</sup>
<i>Single-domain proteins</i>			
1akeA, 1-213	4.0	3.60 (4.5)	1.07 (1.07)
	6.0		1.12 (1.13)
	8.0		1.39 (1.40)
	10.0		1.76 (1.76)
	12.0		1.95 (1.98)
	14.0		2.14 (2.16)
1c1l, 4-147	4.0	4.90 (5.0)	1.52 (1.52)
	6.0		2.00 (2.01)
	8.0		1.95 (1.97)
	10.0		2.08 (2.08)
	12.0		2.61 (2.62)
	14.0		n/a <sup>b</sup>
1jxmA, 531-711	4.0	5.34 (5.4)	2.59 (2.61) <sup>c</sup>
	6.0		2.98 (3.03) <sup>c</sup>
	8.0		2.85 (2.88)
	10.0		2.81 (2.82)
	12.0		3.31 (3.33)
	14.0		3.28 (3.30)
1uwoA, 1-90	4.0	4.19 (4.7)	2.89 (2.89)
	6.0		2.35 (2.35)
	8.0		2.59 (2.59)
	10.0		2.56 (2.56)
	12.0		2.65 (2.65)
	14.0		2.95 (2.95)
1c1xA, 8-345	10.0	5.46 (6.6)	4.30 (4.40)
1cczA, 1-170	10.0	5.58 (5.2)	4.72 (4.73) <sup>d</sup>
1g5yD, 231-442	10.0	4.83 (5.4)	4.74 (4.93)
<i>Table S2 – continued.</i>			
System	Resolution (Å)	Initial $C_{\alpha}$ RMSD (Å) <sup>a</sup>	Final mean

(PDB code, range)			C <sub>α</sub> RMSD (Å) <sup>a</sup>
<i>Multi-domain proteins</i>			
1a45A, 1-172	10.0	12.34 (28.9)	1.95 (1.97) <sup>f</sup>
1ckmA, 60-319	10.0	8.32 (8.3)	6.27 (6.41)
1ffgAB, 2-226	10.0	9.06 (9.1)	3.24 (3.28) <sup>g</sup>
1hrdC, 22-446	10.0	5.96 (8.2)	4.13 (4.22)
1iknA, 19-273	10.0	7.73 (10.4)	5.85 (6.13)

<sup>a</sup> C<sub>α</sub> RMSDs were calculated after least-square fitting each structure to the target one, since the density maps are not aligned with the target structures. Notice that this is the only case in this paper where RMSD calculation was preceded by a least-squares fitting. Analysis programs implement different fitting algorithms; for easier comparison, we present in parenthesis RMSDs calculated with MODELLER (Sali and Blundell, 1993), with the initial values corresponding to data presented in (Topf et al., 2008). All other analyses presented in this paper were performed with VMD (Humphrey et al., 1996).

<sup>b</sup> We were not able to obtain a correct rigid-body docking using Situs.

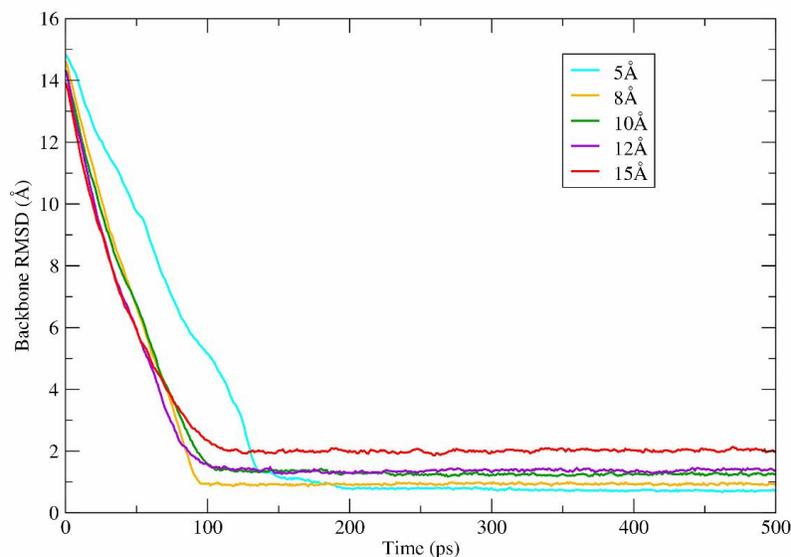
<sup>c</sup> Average calculated from the last 300 ps of a 1-ns simulation.

<sup>d</sup> Average calculated from the last nanosecond from a 2-ns simulation.

<sup>e</sup> Average calculated from the last nanosecond from a 2.5-ns simulation.

<sup>f</sup> For this system, since the two initial domains are disjoint in the benchmark set, MDFF cannot be directly applied; thus, the domains were first rigid-body docked separately, then fitted with MDFF considering each domain as a different chain for 500 ps, and finally the entire structure was fitted as a single chain for additional 500 ps. The reported accuracy is the average from the last 300 ps of this last step.

<sup>g</sup> Each chain was rigid-body docked separately before applying MDFF.



**Figure S1.** Backbone RMSD with respect to the target structure as a function of the resolution for the ACS/CODH test case presented in the paper. Simulated maps with grid spacing of 1 Å were used in this figure.

#### S4. Validation movies

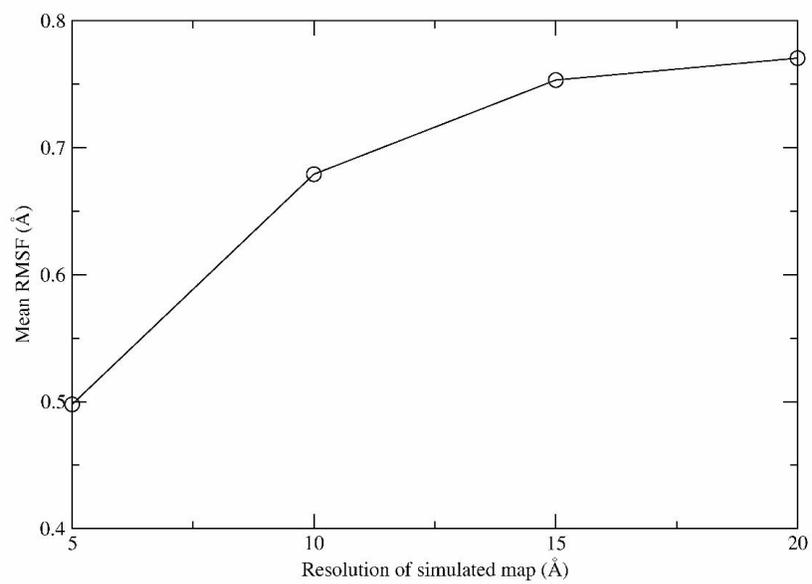
**Movie S1 and S2.** Movies corresponding to MDFF simulations presented in Fig. 3.

#### S5. Simulated maps generated from an ensemble of structures

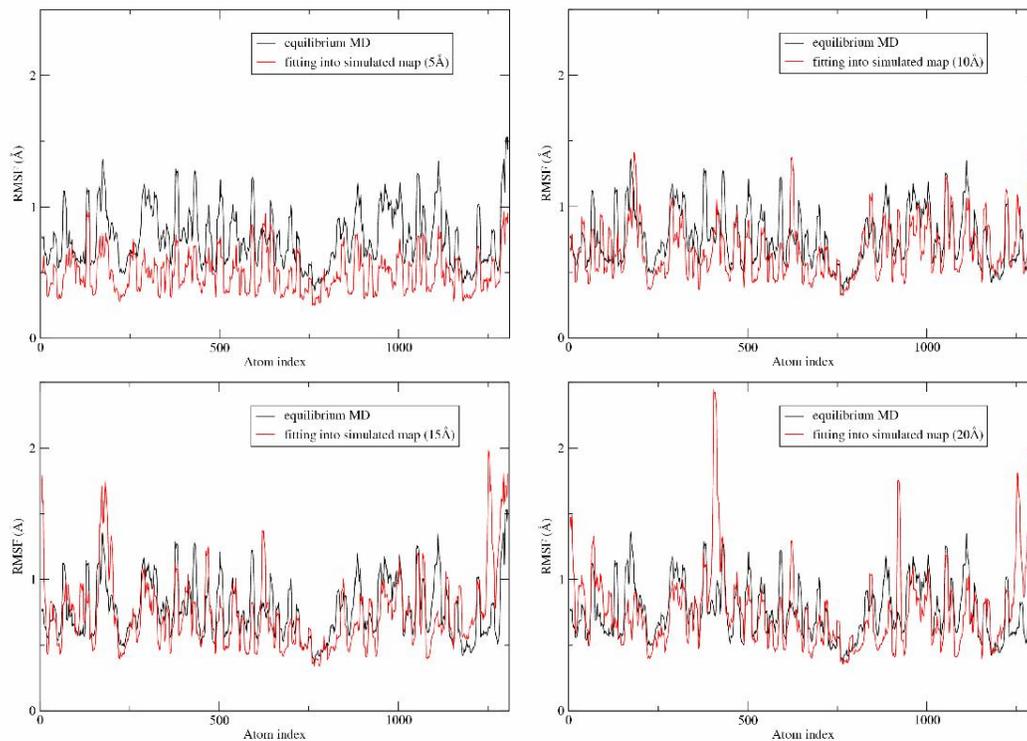
The conformational heterogeneity of biomolecular assemblies, arising from fluctuations of the structure in a given functional state, leads to variability in the EM data: regions of high flexibility in the structure are represented in the EM map by low-density voxel values. MDFF simulations allow the structure to visit all conformations representative of the EM map. In order to illustrate this, we use equilibrium MD trajectories to generate average simulated maps filtered to different resolutions, which contain information about the flexibility of the molecule, and the conformational states visited throughout the simulation. A structure sampled by the MD simulation is randomly chosen to provide the initial atomic coordinates. The equilibrium MD simulation used to generate a simulated map and the fitting simulations differ in three ways: (1) the fitting simulations were performed *in vacuo*, whereas the equilibrium MD simulation

was performed in water; (2) the fitting simulations include the external forces proportional to the gradient of the simulated map; (3) the fitting simulations include secondary-structure restraints. The local flexibility of the protein is assessed by the atomic root mean square fluctuation (RMSF), which is the standard deviation of the coordinates of atom  $i$  with respect to the average structure, given by  $\langle(\mathbf{r}_i - \langle\mathbf{r}_i\rangle)^2\rangle^{1/2}$ . This measure for atomic structures is equivalent to the variance per-voxel that can be obtained from EM data (Penczek et al., 2006); a simulated variance map from a fitting simulation can also be obtained to compare to variance EM maps.

Here we present as an example the bacteriophage T4 lysozyme, one of the structurally best characterized proteins, with the largest number of crystallographic structures in the Protein Data Bank (Berman et al., 2000). The equilibrium MD simulations were performed using a high-resolution X-ray crystallographic structure of the bacteriophage T4 lysozyme (PDB 2LZM (Weaver and Matthews, 1987)), that was solvated in a sphere of explicit water molecules with a diameter of 75 Å, resulting in approximately 20,000 atoms, with chloride ions added to neutralize the total charge of the system. The simulations were performed in the NVT ensemble at a temperature of 310K. The simulated average maps were obtained from 3 ns of simulation, filtered to resolutions of 5, 10, 15 and 20 Å. Figure S2 shows the average RMSF for the different fittings, calculated considering the last 2.5ns of 3.0-ns trajectories with each frame least-square fitted to the average structure. One can see that the fluctuations decrease as the resolution increases, due to the lower uncertainty of the structure at higher resolution. Figure S3 presents a comparison of the atomic RMSF; it can be seen that the atomic fluctuations reproduce the fluctuations in the target map reasonably well, especially for resolutions of 10 and 15 Å; the worse agreement occurs for 5 Å, where the fluctuations in MDFF are significantly lower than those in the target map. Cross-correlation coefficients calculated for the average trajectory with respect the 5, 10, 15 and 20-Å resolution maps are 0.995 (0.971), 0.996 (0.987), 0.995 (0.988), and 0.996 (0.989), respectively, showing that a representative set of conformations arising from MDFF closely reproduces the simulated map used to guide the fitting (the values in parenthesis correspond to local correlations calculated with a threshold of  $0.2\sigma$ ).



**Figure S2.** Mean atomic root mean square fluctuation (RMSF; see text for definition) of MDFF of lysozyme fitted into simulated maps of different resolutions created from an equilibrium MD simulation. The mean RMSF during the equilibrium MD simulation is 0.92Å.



**Figure S3.** Atomic root mean square fluctuations (RMSF; see text for definition) of an equilibrium MD simulation of lysozyme in a water sphere and MDFF applied to simulated maps with different resolutions created from the equilibrium MD. The RMSF was calculated from the last 2.5 ns of 3.0-ns trajectories. A running average with a window size of 10 was used in all plots.

## S6. Atomic model of the *E. coli* ribosome

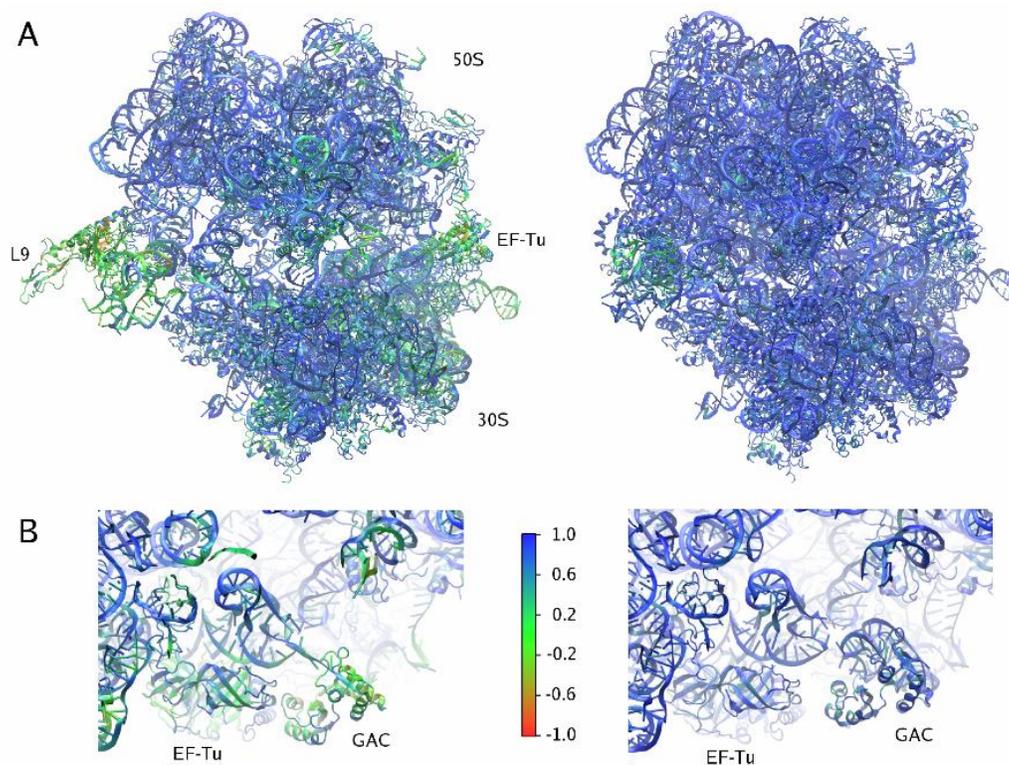
An X-ray structure of the *E. coli* ribosome at 3.22 Å resolution (PDB 2I2U/2I2V) (Berk et al., 2006) furnished the initial coordinates for applying MDFF to cryo-EM maps. Two missing regions in the 23S rRNA, the L1 protuberance and the A-site finger, were modeled. The model of the L1 protuberance was based on an X-ray structure of a *T. thermophilus* 23S rRNA fragment in complex with *S. acidocaldarius* L1 protein (PDB 1MZP) (Nikulin et al., 2003). A homology model of the ribosomal protein L1 was built using MODELLER (Sali and Blundell, 1993) with the structure of the *T. thermophilus* L1 in complex with mRNA (PDB 1ZHO) (Nevskaya et al., 2006) as a template. The A-site finger of the *T. thermophilus* 23S rRNA was resolved at the level of phosphorous atoms (Yusupov et al., 2001); Tung and Sanbonmatsu built an atomic homology model of this structure (PDB 1TWB) (Tung and Sanbonmatsu, 2004), which was used as a template. Models for P- and E-site tRNAs and mRNA were obtained from a 2.8 Å crystal structure of the *T. thermophilus* ribosome (Selmer et al., 2006). For the TC, an *E. coli* EF-Tu structure complexed with kirromycin, a GTP analog, and Phe-tRNA<sup>Phe</sup> obtained by P. Nissen and co-workers was used (PDB 1OB2; unpublished data).

## S7. Comparison between maps at different resolutions

**Table S3.** Convergence time for each step in the multi-step protocol applied to three ribosome cryo-EM maps (TC-bound at 6.7Å and 9.0Å resolution; and with an accommodated A-site tRNA at 9.0Å resolution). The protocol is described in the text. For most steps convergence times are close in value; step (iii), at which the fit of ribosomal proteins is improved, exhibits the largest variability, with significantly longer convergence times for lower resolutions. The cross-correlation coefficients are given in parenthesis, with the global correlation given first, followed by the local correlation calculated with a threshold of 0.2 s; EF-Tu and tRNAs were only included in the correlations in step (iv). The RB step corresponds to the initial rigid-body docking. Average fitted structures were obtained from the last 100 ps of the MDFF trajectories for the the 6.7 and 9.0-Å resolution TC-bound ribosome maps. The  $C_{\alpha}/P$  RMSDs between the ribosome structure after rigid-body docking into the 6.7-Å TC-bound ribosome map and the average fitted structures are 7.41 Å and 7.56 Å, for the 6.7 and 9.0-Å resolution maps, respectively, whereas the RMSD between the two average fitted structures is 3.10Å.

Step	6.7-Å TC-bound ribosome	9.0-Å TC-bound ribosome	9.0-Å Accomodated ribosome
(i)	250 ps (0.881)	250 ps (0.885)	250 ps (0.843)
(ii)	900 ps (0.883)	800 ps (0.888)	1600 ps (0.849)
(iii)	500 ps (0.884)	3000 ps (0.890)	10000 ps (0.854)
(iv)	100 ps (0.913)	325 ps (0.919)	100 ps (0.878)
Total	1750 ps	4375 ps	11950 ps

## S8. Local cross-correlation coefficient map



**Figure S4.** MDFF using a TC-bound ribosome cryo-EM map at 6.7-Å resolution. The structures are colored according to the local cross-correlation coefficient. The rigid-body docked structure is shown on the left, whereas the final structure obtained by the MDFF method is shown on the right. (A) View of the entire ribosome; (B) GTPase-associated center, in the same view as depicted in Fig. 4D. Correlation maps used to color the structures were calculated in the following way: for each grid point, the cross-correlation coefficient between the structure and the EM map was calculated considering only atoms in a sphere with radius of 5 Å centered on the grid point.

## S9. Movie showing fitting of the ribosome into a 6.7-Å map

**Movie S3.** The movie illustrates the multi-step protocol used when MDFF was applied to cryo-EM maps of the ribosome.

### Supplemental references

Belnap, D.M., Filman, D.J., Trus, B.L., Cheng, N., Booy, F.P., Conway, J.F., Curry, S., Hiremath, C.N., Tsang, S.K., Steven, A.C., et al. (2000). Molecular tectonic model of virus structural transitions: The putative cell entry states of poliovirus. *J. Virol.* *74*, 1342–1354.

Berk, V., Zhang, W., Pai, R.D., and Cate, J.H.D. (2006). Structural basis for mRNA and tRNA positioning on the ribosome. *Proc. Natl. Acad. Sci. USA* *103*, 15830–15834.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. *Nucl. Acids Res.* *28*, 235–242.

Chiu, W., Baker, M.L., Jiang, W., Dougherty, M., and Schmid, M.F. (2005). Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure* *13*, 363–372.

Clore, G.M., and Gronenborn, A.M. (1998). NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* *2*, 564–570.

Dutta, S., and Berman, H.M. (2005). Large macromolecular complexes in the protein data bank: A status report. *Structure* *13*, 381–388.

Foloppe, N., and MacKerrell Jr., A.D. (2000). All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.* *21*, 86–104.

Frank, J. (2006). *Three-dimensional electron microscopy of macromolecular assemblies* (New York: Oxford University Press).

Frank, J., and Spahn, C.M.T. (2006). The ribosome and the mechanism of protein synthesis. *Rep. Progr. Phys.* *69*, 1383–1417.

Frenkel, D., and Smit, B. (2002). *Understanding Molecular Simulation From Algorithms to Applications* (California: Academic Press).

- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD – Visual Molecular Dynamics. *J. Mol. Graphics* *14*, 33–38.
- Langmore, J., and Smith, M. (1992). Quantitative energy-filtered electron microscopy of biological molecules in ice. *Ultramicroscopy* *46*, 349–373.
- MacKerell, A., Jr., Bashford, D., Bellott, M., Dunbrack, R.L., Jr., Evanseck, J., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* *102*, 3586–3616.
- MacKerell Jr., A.D., and Banavali, N.K. (2000). All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comp. Chem.* *21*, 105–120.
- MacKerrel, A.D., Feig, M., and Brooks III, C.L. (2004). Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.* *25*, 1400–1415.
- Nevskaya, N., Tishchenko, S., Volchkov, S., Kljashtorny, V., Nikonova, E., Nikonov, O., Nikulin, A., Köhrer, C., Piendl, W., Zimmermann, R., et al. (2006). New insights into the interaction of ribosomal protein L1 with RNA. *J. Mol. Biol.* *355*, 747–759.
- Nikulin, A., Eliseikina, I., Tishchenko, S., Nevskaya, N., Davydova, N., Platonova, O., Piendl, W., Selmer, M., Liljas, A., Drygin, D., et al. (2003). Structure of the L1 protuberance in the ribosome. *Nat. Struct. Biol.* *10*, 104–108.
- Penczek, P.A., Yang, C., Frank, J., and Spahn, C.M.T. (2006). Estimation of variance in single-particle reconstruction using the bootstrap technique. *J. Struct. Biol.* *154*, 168–183.
- Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. *J. Comp. Chem.* *26*, 1781–1802.
- Reimer, L. (1997). *Transmission Electron Microscopy* (Berlin: Springer-Verlag).
- Roseman, A.M., Chen, S., White, H., Braig, K., and Saibil, H.R. (1996). The chaperonin ATPase cycle: Mechanism of allosteric switching and movements of substrate-binding domains in GroEL. *Cell* *87*, 241–251.
- Saibil, H.R. (2000). Macromolecular structure determination by cryo-electron microscopy. *Acta Cryst. D* *56*, 1215–1222.
- Sali, A., and Blundell, T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* *234*, 779.

Selmer, M., Dunham, C.M., Murphy IV, F.V., Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942.

Spence, J. (1988). *Experimental High-Resolution Electron Microscopy* (New York: Oxford University Press).

Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W., and Sali, A. (2008). Protein structure fitting and refinement guided by cryo-EM density. *Structure* 16, 295–307.

Tung, C.S., and Sanbonmatsu, K.Y. (2004). Atomic model of the *Thermus thermophilus* 70S ribosome developed in silico. *Biophys. J.* 87, 2714–2722.

Weaver, L., and Matthews, B. (1987). Structure of bacteriophage T4 lysozyme refined at 1.7 Angstroms resolution. *J. Mol. Biol.* 193, 189–199.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292, 883–896.