Review Article

Tools for the cryo-EM gold rush: going from the cryo-EM map to the atomistic model

Doo Nam Kim[1] and Karissa Y. Sanbonmatsu[1,2]

[1]Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, U.S.A.; [2]New Mexico Consortium, Los Alamos, U.S.A.

Correspondence: Karissa Y. Sanbonmatsu (kys@lanl.gov)

As cryo-electron microscopy (cryo-EM) enters mainstream structural biology, the demand for fitting methods is high. Here, we review existing flexible fitting methods for cryo-EM. We discuss their importance, potential concerns and assessment strategies. We aim to give readers concrete descriptions of cryo-EM flexible fitting methods with corresponding examples.

Introduction

In the past two years alone, more than 300 cryo-electron microscopy (cryo-EM) reconstructions with resolutions <4 Å were deposited in the Electron Microscopy Data Bank (EMDB) [1], enabling atomistic structures for the *Bacillus anthracis* pore complex, many types of viruses, membrane proteins, transcription complexes, and other macromolecular complexes. cryo-EM structures of large complexes now rival X-ray crystallography in resolution. One example of this is the 1.8 Å resolution structure of glutamate dehydrogenase by cryo-EM, whereas the highest X-ray structure resolution is 2.7 Å [1,2].

Among the four methods of cryo-EM (cryo-tomography, single particle analysis, electron crystallography, and micro-electron diffraction), single particle analysis (SPA) and cryo-tomography do not require crystallization [3]. SPA is one of the major cryo-EM approaches (78% among all cryo-EM methods) [4]. It can elucidate compositional/conformational heterogeneity and capture more native-like states of macromolecular complexes than X-ray crystallography [5]. For example, isocitrate dehydrogenase was believed as an asymmetric dimer as seen by X-ray crystallography. However, cryo-EM had shown that the apparent asymmetry is likely due to crystal lattice formation rather than intrinsic properties. Indeed, inspection of lattice packing confirmed this [6]. Other examples include medium resolution cryo-EM structures of the AMPA glutamate receptors that showed how stabilizing mutations for crystallization perturbed the native structure [7,8]. In addition, cryo-EM structures have been solved for helical polymers that cannot be crystallized in their native form because of symmetry mismatch with any crystal space group. Many proteins in cells and viruses are helical polymers [7,9]. Another example is the 2.3 Å resolution cryo-EM structure of human p97, an AAA ATPase for which crystal structures of the full-length complex have been difficult to obtain. Crystal structures of individual domains are not useful for analysis of inhibitors that bind to interfaces between these domains [7,10]. Overall, there are limitations of side chain rotamers in high resolution of X-ray crystallography [1].

General aspects of cryo-EM

In cryo-EM SPA, larger complexes are typically easier to solve than smaller systems since the larger particles are more easily discernible with a higher signal-to-noise ratio [11]. Recently, however, cryo-EM SPA structures have been solved for proteins as small as 64 kDa [5,12] and 93 kDa (794 amino acids, PDB accession code: 5K10) [6]. A new optical device, the volta-phase plate, has been used for smaller molecules by improving contrast by an order of magnitude [13]. Since SPA can determine complexes as big as 11,200 kDa, the method is applicable to a wide distribution of sizes. Additionally, it is expected that cryo-EM SPA will be eventually applicable to determine biological structures *in vivo* [14]. cryo-EM already has

been used to determine structures *in situ* either by SPA [15,16] or tomography [17,18]. To meet these growing needs, many cryo-EM facilities are being established around the world including United States, Europe [19], and China [20] with the latest microscopes including the *Krios* and *Arctica* [21].

However, a well-known challenge exists: the interaction with the air–water interface can damage particles when the sample becomes too thin after blotting [22]. To easily predict this status, it was recently found that streptavidin monolayer crystals can be used as an internal standard to assess single-particle cryo-EM image quality even when Thon ring qualities are the same [23]. Additionally, the spraying–plunging method using microfluidics is poised to be more readily adopted to produce highly consistent thicknesses, replacing pipetting–blotting–plunging method [24]. These endeavors for better image quality are expected to further fuel the rapid growth of cryo-EM along with newly built cryo-EM facilities.

Both map reconstruction and model building require significant compute resources. For example, it is estimated that more than half a million CPU hours of classification and high-resolution refinement were spent to generate 3.7 Å resolution structure of a spliceosomal complex [25,26]. As a result, many graphics processing unit (GPU) based applications have been and are in development [25,27–29]. It is also notable that *Nvidia*'s GPU (Tesla V100) achieves 15 teraflops for single-precision calculations [30] and the upcoming Intel memory, *Optane* is expected to reduce a realistic bottleneck of modern computers by boosting data input and output speed drastically [31]. In order to save compute time, cryo-EM modelers are advised to know that single precision calculations have shown similar accuracy as with double-precision for the widely used map reconstruction software package, *relion* [25].

A new map reconstruction software package, *cryoSPARC,* refines a 3D map almost nine times faster than the widely used GPU enabled *relion* by removing redundant computation [27]. This significant speed-up is very useful because many different approaches need to be explored to process different datasets [32]. Among many 3D classification methods, *cryoSPARC* and *relion* (starting in version 2.1) stand out, as these do not require a template for map reconstruction. Most other existing techniques for 3D classification in cryo-EM map reconstruction need a single initial reference from which discovery of heterogeneity proceeds. This unique advantage of *de novo* map reconstruction has already enabled *cryoSPARC* to find more unexpected conformations [27,33]. Finding more conformations of biomolecules opens a new horizon for flexible fitting, as it can sample more biologically relevant conformations which are distinguished by cryo-EM SPA [34]. Once time-resolved cryo-EM [24,35] matures, flexible fitting to study short-lived intermediates is expected to increase its importance. Finally, as reviewed previously [11], energy filters to decrease noise, gold films, and films made of graphene/graphene oxide to reduce motion are promising new directions.

Toward higher resolution cryo-EM maps

Until recently, cryo-EM has been limited to intermediate resolution (5–20 Å) [36]. At these resolutions, highly accurate atomic modeling directly from the data has not been feasible and fitting methods have to be used. Here, the electron density does not have sufficiently distinctive features for an unambiguous placement of a component. One of the earliest fitting methods for low resolution cryo-EM data was core-weighting and grid-threading Monte Carlo (GTMC) [37]. The core-weighting simplifies a prohibitively large exhaustive conformational search with many components into single component fitting procedure. Then each single component fitting is performed using a GTMC that identifies the global maximum state (best fit) among a series of local maximum states determined by short Monte Carlo searches originating at a variety of grid points. Cross-linking techniques to determine the spatial distribution of proteins were also used to supplement this low-resolution map [38]. However, fitting results into low-resolution maps often had to be corrected once higher resolution maps were provided [38].

Although the resolutions of the map were not high, useful models could be obtained by fitting into a map [39]. For example, a fitted model in a low resolution map was used to predict the change of binding energy [40]. Additionally, modeling low resolution cryo-EM maps was possible using individual components from other high-resolution structure determination methods, such as X-ray crystallography, NMR, or comparative modeling [34,39,41,42]. There are several excellent previous reviews with respect to various computational fitting methods for these low-resolution maps [34,39,43].

Recent breakthroughs such as emergence of the direct electron detector camera, better motion correction, and improved contrast transfer functions [28,44] for SPA have produced cryo-EM maps with significantly higher resolutions, such as 1.8 [6] and 2.2 Å [45]. This improvement in resolution allows more in-depth analysis. For example, with 3.6 Å resolution, it was possible to see how an inhibitor interacted specifically with the *Plasmodium falciparum* proteasome, not the human one [7,46]. With respect to placing small molecules accurately, 2.5–3.0 Å resolution is typically required for X-ray crystallography derived maps: a similar effective resolution is expected for cryo-EM structures.

One of the greatest examples that will prove this resolution significance is study of autophagy [22] which was discovered by transmission electron microscopy (TEM) by more than 50 years ago [47] (most modern cryo-EM uses TEM as well).

This resolution revolution has introduced different design considerations for flexible fitting. The maps tend to contain more localized noise features, making overfitting a greater challenge. On the other hand, modelers can better focus on missing maps or highly dynamic regions: simulation of cryo-EM maps with missing densities is more difficult to interpret as dynamics of biomolecules will be affected by removing domains during the fitting [48]. As resolution becomes better, molecular dynamics (MD) simulation can more effectively search conformational space, as root mean squared fluctuation (RMSF) values and local resolution are highly correlated ($r^2 = 0.64$–0.97) [49]. Additionally, a better a consensus among different flexible fitting methods is achieved with higher resolution maps as most fitting methods are more accurate at higher resolution [39,48,50]. Even with these high resolution cryo-EM maps, using individual components from other high resolution structure determination methods is still a dominant approach [51-53].

What is flexible fitting?

Because cryo-EM experiments can be prepared more readily and often contain many distinct conformations within a single experiment than X-ray crystallography, large numbers of cryo-EM maps are being produced by single cryo-EM labs. A bottleneck exists in going from the cryo-EM maps to the final product, namely the PDB structure.

All structural models derived from EM maps introduce information beyond the nominal resolution [43]. Therefore, structural modeling should be performed cautiously with deep understanding of map reconstruction. Consequently, there are many backlogs of EM models compared with the number of reconstructed EM maps [34]. In order to solve this imperative need, many modeling methods for cryo-EM maps have been developed, such as rigid domain fitting (RDF), multibody fitting, flexible fitting, homology modeling, high-resolution modeling, and *de novo* modeling [54,55]: pathwalker, one of the powerful methods for *de novo* modeling, automatically traces a protein backbone from a near-atomic resolution (3–6 Å) density map and has many features such as iterative secondary structure identification, termini detection, and modeling multiple subunits without prior segmentation [56]. Among these different modeling approaches, the flexible fitting can derive a new conformational model that conforms to cryo-EM map by changing a starting structure [57]. Flexible fitting has the advantage over rigid fitting for higher resolution maps, because the rigid fitting cannot handle structural variation in electron microscopy and electron tomography maps [58]. Therefore, more than 18 different flexible fitting methods have been developed [34,43,58]. We show several examples in Figure 1. Some of the first techniques invented were simulated annealing optimization (*YUP.SCX*) [59] and normal mode flexible fitting (*NMFF*) [60,61]. One common method of flexible fitting for cryo-EM map uses molecular dynamics simulation [62,63]. Molecular dynamics flexible fitting (*MDFF*) uses prepared structures [49,64], input parameters [65], or interactive input by users [62,66]. Iterative modular fitting requires manual intervention as well, e.g. segmentation [39]. A variant on the MD theme includes all-atom structure-based molecular dynamic flexible fitting (*MDfit*) [63,67-69]. Another method (*Direx*) uses a geometry-based conformational sampling under low resolution restraints [70,71], and has been used to flexibly fit linker DNA bridging the nucleosome core into the density [72]. Below, we summarize recommended procedures before and after this flexible fitting for any method of flexible fitting.

Data processing before flexible fitting

For more accurate flexible fitting, it is recommended to sharpen the cryo-EM map before flexible fitting [49,75] using either *Phenix* [76] or *cryoSparc* [27]. Of course, oversharpening should be avoided to prevent a partial noisy map which becomes ultimately uninterpretable [77]. After sharpening, one can manually place a starting molecule rigidly into a map using *UCSF Chimera* [78], because most flexible fitting methods require a decent level of initial cross-correlation between the model and the density [79]. In one example, the radius of convergence was reported as 25 Å in root mean square deviation (RMSD) for MD simulation based flexible fitting using progressively higher resolutions [49]. On the other hand, a direct MD simulation based flexible fitting and *Rosetta* [80] have substantially better radii of convergence. For more automated rigid-body fitting, one can use the *Fit in Map* routine in *UCSF Chimera* or *Powerfit* [79].

Refinement after flexible fitting

Once it is confirmed that the cross-correlation (CC) value has been increased through the flexible fitting, one needs to refine modeled structures for better geometric qualities such as correct rotamer placement and protein backbone dihedral angles in accordance with Ramachandran plot, just before depositing to protein data bank (PDB) [81]. For this
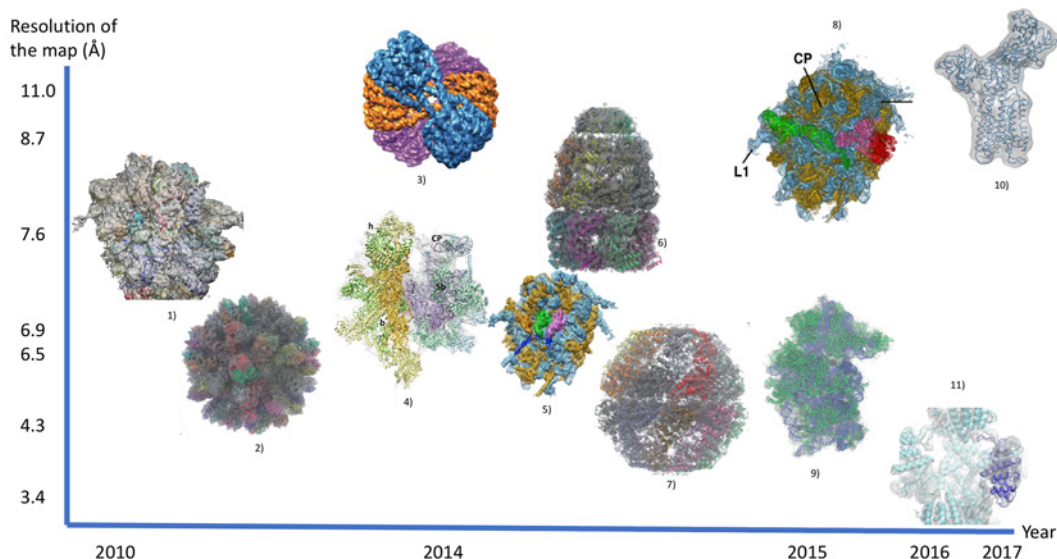
Figure 1. Reconstruction of biomolecules from cryo-EM maps using flexible fitting (scales of axes are arbitrary)
1) Head swivel motion on the ribosome was discovered [67]. 2) Rabbit hemorrhagic disease virus capsid (PDB 3J1P) [73]. 3) 30-nm chromatin fibers reconstituted on 12 × 187 bp DNA [72]. 4) Small subunit of the mammalian mitochondrial ribosome [64]. 5) Subunit rolling of a eukaryotic-specific ribosome was discovered [63]. 6) GroEL–GroES complex in the ATP-bound state (PDB 2C7D) [73]. 7) Mm-cpn in the closed state (PDB 3LOS) [73]. 8) Ribosomal 80S complex with termination factors [69]. 9) Entire small subunit of human ribosome fitted by cryoFIT [74]. The resolution was estimated based on article context. 10) Open structure of Ca^{2+} ATPase [57]. The resolution was estimated based on article context. 11) Transient receptor potential cation channel subfamily V member 1 (TRPV1) [49].

refinement, either *phenix.real_space_refine* [82] or *Rosetta* density-guided iterative local refinement [80] is highly recommended. Both *phenix.real_space_refine* and *Rosetta* refinement produced nearly same quality of geometries, but *phenix.real_space_refine* refined almost 20,000 times faster than *Rosetta* refinement with one test case of TRPV1 [83]. However, *Rosetta* could perturb the structure more than *Phenix* by larger backbone corrections with fragment insertion.

Advantages of flexible fitting

It is generally regarded that flexible fitting is best suited for cryo-EM maps with resolutions between ∼3 and 9 Å [43]. If the resolution is poorer than 9 Å, the required conformational sampling can be prohibitive. If the resolution is better than 4 Å, the maps begin to show individual β-strands and side chains. In this case, *de novo* modeling becomes possible [54]. This technique can produce conformational changes that cannot be adequately modeled as movements of rigid domains alone [39,84]. In flexible fitting via MD, the MD potential is biased by a weighting factor proportional to the correlation between a cryo-EM electron density map and a simulated map based on the simulation model [63]. This method excels at deforming static X-ray crystal structures to fit the low-resolution EM data [57] and is especially useful for refining highly dynamic regions such as flexible exterior regions [62], fitting with significantly large conformational change such as closed-to-open transition [48], and loop modeling (in practice, loop regions are often deleted if they cannot be fit into cryo-EM maps, but most MD based flexible fitting can fit loop regions as well). One of the most popular methods is MDFF [49]. A second MD-based method is *MDfit*, which uses a native contact based all-atom Go model. This method has the advantage of preserving stereochemistry in the original model while simultaneously producing large conformational changes highly consistent with cryo-EM maps. In addition, the method uses a simplified model, which can fit ribosomes on a single desktop computer. *MDfit* was used to model the head–swivel transition, a key step during ribosome translocation [67]. The method also modeled a new conformational change specific to eukaryotic ribosomes (subunit rolling) [63]. In summary, when the resolution is high enough to show significant structural variation between the fitting high-resolution structure (from X-ray crystallography or NMR) and the EM maps, flexible fitting should be used [58].

For this reason, some benchmark studies have shown that flexible fitting fits better to crystal structures than RDF does [48] by giving better correlations between cryo-EM maps and modeled structures [85]. For example, it was reported that a domain in a model that has been deposited in PDB was oriented incorrectly (100° compared with an X-ray structure) by RDF [50]. For high-resolution data, flexible fitting has important advantages over modular fitting by picking up conformational changes that cannot be modeled by movements of rigid domains [39]. Flexible fitting has advantages over *de novo* methods as well. For example, when the β-galactosidase structure was modeled with different methods, flexible fitting methods generated structures with a clash score of zero, while *de novo* methods modeled a structure with a clash score of 54 [49]. Flexible fitting methods though have produced more Ramachandran outliers (1.6–3.0 %) than *de novo* methods (0.2 % Ramachandran outliers). We note that the references for Ramachandran distributions come from merely 500 files [86]. Overall, *MolProbity* scores from flexible fitting were better with ∼1.1–1.4 compared with 3.1 from *de novo* methods.

These advantages of flexible fitting are understandable, given that proteins are intrinsically flexible [87] and go through significant conformational changes. For example, because there was substantial flexibility in the structure, cryo-EM analysis had to use just 10% of the particles identified in the images [1,88]. Of course, structural changes are also induced by ligand binding or pH change as observed by solution NMR, cryo-EM single particle analysis, and room temperature X-ray crystallography [33,89,90].

## Overcoming challenges in flexible fitting

There are concerns over flexible fitting since this methodology may lead to overfitting and misinterpretation of noisy density features [39]. Restraining with secondary structure information often reduces overfitting [91]. Additionally, clustering analysis is known to be effective to avoid the overfitting problem, necessitating a large number of fitting trials to generate ensembles [57]. This approach (sampling ensemble of conformations and clustering later) has been used for *Rosetta* protein structure prediction as well [92], and makes sense given the inherent protein dynamics and its corresponding cryo-EM map space as shown with RMSF values [49].

Additionally, there is a concern over stereochemistry errors from MD simulation based flexible fitting. For example, because force fields in MD simulation do not have any energy term to enforce given chirality or *cis/trans* peptide bond configuration, an external force to fit to cryo-EM map may result in artificial chirality or peptide bond configuration [62]. However, the *MDfit* and hybrid MD technique (cryoFIT) [74] help to remedy this problem since its potential is based on initial native contact and dihedral values from the starting structure.

Finally, one caveat when applying flexible fitting to high-resolution maps is that steep wells in which the structure can be trapped are commonly found. To overcome these wells, map-restrained self-guided Langevin Dynamics (*SGLD*) was applied [93]. The SGLD selectively enhances and suppresses molecular motions based on their frequency to accelerate conformational searching without modifying energy surfaces or raising temperature. The restraint maps are included in simulations as movable objects (rigid domains) that interact with simulation systems and move with their restraining atoms. Another method of surpassing these wells was applying low-pass or Gaussian filter to the cryo-EM density map smoothing the resulting potential energy function [57,62,67].

Quality assessment of flexible fitting

It is believed that general consensus from various flexible fitting methods often is quite close to experimentally determined structures. This way of assessment is often referred as a battery approach. This argument brought a concern since all flexible fitting methods may have some systematic errors [39]. Using multiple flexible fitting offers no assurance about which model is optimal [43] as evidenced with a benchmark of 12 different proteins [50]. However, the battery approach is a well-established method of statistical assessment with respect to experiments [94-96] and prediction modeling [97]. One of the most representative examples of a battery approach not related to cryo-EM flexible fitting is a random forest which is a collection of decision trees whose outputs are summed to derive a final prediction by majority voting [98]. Indeed, when modelers validated this consensus approach with many X-ray crystallography derived structures, they found a very high correlation between consensus level and model accuracy [48,50]. For example, a correlation (r^2) between the mean RMSD among the different models (a measure of consensus) and the mean RMSD between the models and the target structure for 12 different proteins was 0.99. This high correlation between consensus level and model accuracy was observed not only in global regions of the protein, but also in local regions. On top of that, at least *NMFF*, *MDfit*, *YUP.SCX*, and *DireX* are diverse enough not to have consensus on wrong solutions and yet to have consensus on correct solutions. On the other hand, the CC between the density map and the fitted model shows a very low correlation ($r^2 = 0.31$) with RMSD from targets [50]. Therefore, the battery
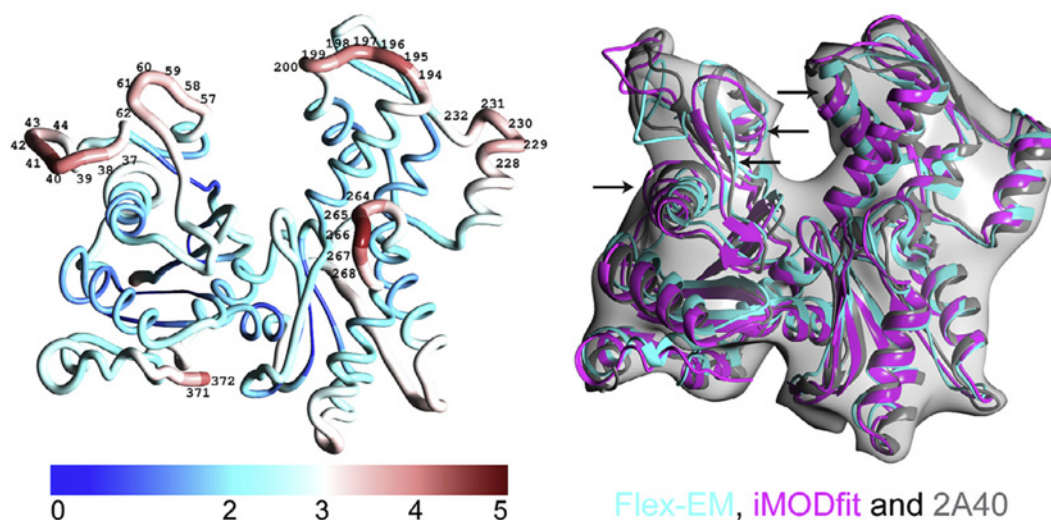
Figure 2. Modeling errors for the case of flexible fitting of an actin subunit homology model into simulated map.
Left: Local residue error (QMEAN, which is calculated with geometrical structural descriptors in Å). Error values above 3.5 Å that are considered unreliable are labeled. Right: Comparison of flexible fits obtained using *Flex-EM*, *iMODfit*, and the target fit (PDB ID 2A40) [36].

approach is useful to confer modeling confidence to identify local regions that are likely to fit well. A similar conclusion was made with other flexible methods as well: *flexEM* and *iMODfit* [36] (Figure 2). Not only did they derive the same conclusion, but also they improved the fit by using both of these flexible fitting methods (first they identified poorly fitted region by comparing results from two flexible fitting methods, then they improved that region only while keeping the rest of the structure fixed).

# Comparison of some flexible fitting methods

Although there are at least 18 different flexible fitting methods [34,43,58], there is no system-wide level of comparison to our knowledge. However, there are several benchmark papers. Ahmed et al. [48] tested *NMFF*, *MDfit*, and *YUP.SCX* with five different proteins. Among these methods, *MDfit* and *YUP.SCX* fitted structures (elongation factor 2) as close as 1.4 Å RMSD even when the required conformational changes are large (on average 9 Å). The initial structures needed to undergo a closed-to-open transition in order to fit into the EM maps. With respect to convergence, *MDfit* was most robust among the three approaches. Ahmed et al. [50] also tested *NMFF*, *MDfit*, *YUP.SCX*, and *DireX* with 12 different proteins. Overall, *MDfit*, *YUP.SCX*, and *DireX* showed high agreement in conformation with each other and resulted in the lowest mean RMSD from the target structure.

When *MDFF*, *DireX*, *Rosetta*, *Gorgon*, and *Froda* were tested with 1–7 biological systems, *MDFF* and *Rosetta* produced models free from chirality errors and *cis* peptide [99]. *MDFF* prevented stereochemical errors using harmonic restraints and *Rosetta* prevented these errors using experimentally derived force field and rotamers with fragment insertion [100].

When comparing speed between replica cryo-EM map resolution exchange *MDFF* and simple cascade of resolution change *MDFF*, replica exchange method converged 8–14 times faster than simple cascade one with similar fitting accuracy [49].

With respect to model accuracy, *Rosetta* refinement generated more accurate models than the direct *MDFF* [80]. Additionally, the accuracy of *Rosetta* refined models was independent of the starting-model quality when using cryo-EM maps of 4.5 Å or better resolution. However, when *MDFF* used cascade resolution of maps (*cMDFF*, *ReMDFF*), both *MDFF* and *Rosetta* generated comparable results [49].

To specifically address higher resolution cryo-EM maps, a hybrid MD technique (cryoFIT) was developed to maintain stereochemistry of the initial starting structure while having the capability to form remodeled secondary structures and local folds [74] (Figure 3). This method combines a native contact potential with the *ab initio* potentials such as *Amber*. The method has shown promising results on the human ribosome.
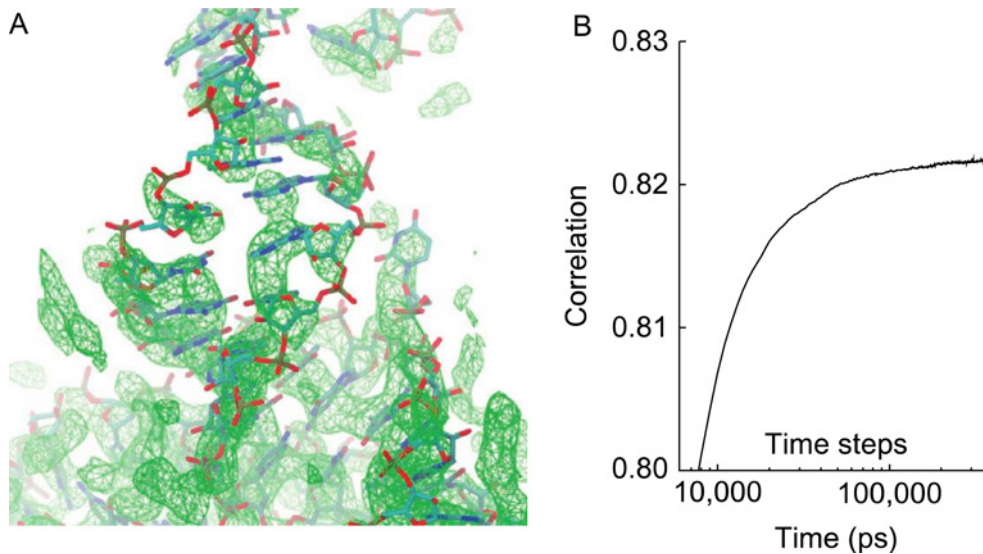
Figure 3. An example of cryoFIT used to fit high resolution cryo-EM map of the human ribosome [101].
(**A**) cryo-EM map of h40–h44 segment. (**B**) Time evolution of fit during simulations (courtesy Kirmizialtin et al., 2015) [74].

Outlook

It is expected that more MD based flexible fitting methods will adopt replica exchange or simulated annealing along with application of different Gaussian blurs of various half-widths to the experimental density as the hybrid MD simulation fitting method (cryoFIT) [74] and *MDFF* did [49], as the evidence of benefits using those enhanced sampling methods for flexible fitting has accumulated [57]. Whether the replica exchange is for either traditional temperature change or cryo-EM map density, these enhanced samplings are to overcome energetic local minima efficiently. Similarly, ensemble based sampling [57] such as RMSF [49] is also expected to be applied in order to better assess model quality. These methods also have complementary benefits to traditional CC based fit assessment.

When it comes to quantitative characterization of domain motions, minor residue mismatch often makes it challenging to compare two different structures. Moment of inertia tensors (after near-automated segmentation) may aid with this comparison [102]. This method, implemented in *VMD*, had shown its practical usefulness for ribosome with atomic structure derived maps. However, it is applicable to any molecular machines and its further development with cryo-EM derived maps is done as well.

To sample conformational space more effectively, additional constraints from evolutionary information from aligned homologous proteins are expected to be added to MD simulation based flexible fitting methods. In this way, one can find biologically more meaningful solutions by sampling more relevant conformations with reduced simulation time. The use of evolutionary conservation of residues is already being used to detect incorrectly fitted proteins [103,104] in cryo-EM maps, to predict protein structures [105,106] and druggable interfaces of protein–protein interactions [107].

To model symmetric molecules more effectively, symmetry restraints as realized in *MDFF* [108] can be considered to be applied to other flexible fitting methods as well. *MDFF* allows users to use *Amazon Web Service Cloud* platform that costs merely $0.8–1.7 for carbon monoxide dehydrogenase fitting [49,109]. Additionally, it presented GUI based applications that enable convenient model quality check plugins such as *cispeptide*, chirality, TorsionPlot, and Timeline [62].

One thing to note in this expanding field is that an increasing number of cryo-EM structures are being reported from dynamic protein assemblies [7] such as nuclear complexes that have been difficult to study conventionally [11]. Therefore, flexible fitting methods will be more important tools in this gold-rush era for cryo-EM.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

Abbreviations

CC, cross-correlation; cryo-EM, cryo-electron microscopy; GTMC, grid-threading Monte Carlo; MD, molecular dynamics; MDFF, molecular dynamics flexible fitting; NMFF, normal mode flexible fitting; RMSD, root mean square deviation; SGLD, self-guided Langevin Dynamics; SPA, single particle analysis.

References

1   Chiu, W. and Downing, K.H. (2017) Editorial overview: cryo electron microscopy: exciting advances in cryoEM Herald a new era in structural biology. *Curr. Opin. Struct. Biol.* **46**, 1–5

2   Earl, L.A., Falconieri, V., Milne, J.L.S. and Subramaniam, S. (2017) cryo-EM: beyond the microscope. *Curr. Opin. Struct. Biol.* **46**, 71–78

3   Nannenga, B.L. and Gonen, T. (2016) MicroED opens a new era for biological structure determination. *Curr. Opin. Struct. Biol.* **40**, 128–135

4   Lawson, C.L. et al. (2016) EMDataBank unified data resource for 3DEM. *Nucleic Acids Res.* **44**, D396–D403

5   Khoshouei, M., Radjainia, M., Baumeister, W. and Danev, R. (2017) cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate. *Nat. Commun.* **8**, 1–6

6   Merk, A. et al. (2016) Breaking cryo-EM resolution barriers to facilitate drug discovery. *Cell* **165**, 1698–1707

7   Subramaniam, S., Earl, L.A., Falconieri, V., Milne, J.L. and Egelman, E.H. (2016) Resolution advances in cryo-EM enable application to drug discovery. *Curr. Opin. Struct. Biol.* **41**, 194–202

8   Meyerson, J.R. et al. (2014) Structural mechanism of glutamate receptor activation and desensitization. *Nature* **514**, 328–334

9   Zhang, R., Alushin, G.M., Brown, A. and Nogales, E. (2015) Mechanistic origin of microtubule dynamic instability and its modulation by EB proteins. *Cell* **162**, 849–859

10 Banerjee, S. et al. (2016) 2.3 Å resolution cryo-EM structure of human p97 and mechanism of allosteric inhibition. *Science* **351**, 871–875

11 Fernandez-Leiro, R. and Scheres, S.H.W. (2016) Unravelling biological macromolecules with cryo-electron microscopy. *Nature* **537**, 339–346

12 Danev, R. and Baumeister, W. (2017) Expanding the boundaries of cryo-EM with phase plates. *Curr. Opin. Struct. Biol.* **46**, 87–94

13 Danev, R., Tegunov, D. and Baumeister, W. (2017) Using the volta phase plate with defocus for cryo-em single particle analysis. *Elife* **6**, 1–9

14 J. Grant (2105), Part 6: Single Particle Analysis, https://www.youtube.com/watch?v=lwR3MmnIpJI&index=38&list=PL8_xPU5epJdctoHdQjpfHmd_z9WvGxK8-

15 Liu, H. and Cheng, L. (2015) cryo-EM shows the polymerase structures and a nonspooled genome within a dsRNA virus. *Science* **349**, 1347–1350

16 Zhang, X. et al. (2015) In situ structures of the segmented genome and RNA polymerase complex inside a dsRNA virus. *Nature* **527**, 531–534

17 Asano, S. et al. (2015) A molecular census of 26S proteasomes in intact neurons. *Science* **347**, 439–443

18 Mahamid, J. et al. (2016) Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972

19 Stuart, D.I., Subramaniam, S. and Abrescia, N.G.A. (2016) The democratization of cryo-EM. *Nat. Methods* **13**, 607–608

20 Wang, H.W., Lei, J. and Shi, Y. (2017) Biological cryo-electron microscopy in China. *Protein Sci.* **26**, 16–31

21 Titan Krios, https://www.fei.com/products/tem/titan-krios-for-life-sciences/

22 Htherey, J.H. and Nogales, E. (2016) Next-generation electron microscopy in autophagy research. *Curr. Opin. Struct. Biol.* **41**, 211–216

23 Han, B.-G., Watson, Z., Cate, J.H.D. and Glaeser, R.M. Monolayer-crystal streptavidin support films provide an internal standard of cryo-EM image quality. *J. Struct. Biol.*, **in press**

24 Feng, X. et al. (2017) A fast and effective microfluidic spraying-plunging method for high-resolution single-particle cryo-EM. *Structure* **25**, 663–670

25 Kimanius, D., Forsberg, B.O., Scheres, S.H.W. and Lindahl, E. (2016) Accelerated cryo-EM structure determination with parallelisation using GPUS in RELION-2. *Elife* **5**, 1–21

26 Nguyen, T.H.D. et al. (2016) cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution. *Nature* **530**, 298–302

27 Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296

28 Zhang, K. (2016) Gctf: real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12

29 Zheng, S., Palovcak, E., Armache, J.-P., Cheng, Y. and Agard, D. (2017) Anisotropic correction of beam-induced motion for improved single-particle electron cryo-microscopy. *bioRxiv* 1–30

30 Nvidia's newest GPU, https://www.nvidia.com/en-us/data-center/tesla-v100/

31 Intel's optane memory, https://www.intel.com/OptaneMemory

32 Fernandez-leiro, R. and Scheres, S.H.W. (2017) A pipeline approach to single-particle processing in RELION. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 1–14

33 Ripstein, Z.A., Huang, R., Augustyniak, R., Kay, L.E. and Rubinstein, J.L. (2017) Structure of a AAA+ unfoldase in the process of unfolding substrate. *Elife* **6**, 1–14

34 Lopez-Blanco, J.R. and Chacon, P. (2015) Structural modeling from electron microscopy data. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **5**, 62–81

35 Fu, Z. et al. (2016) Key Intermediates in Ribosome Recycling Visualized by Time-Resolved cryoelectron Microscopy. *Structure* **24**, 2092–2101

36 Pandurangan, A.P., Shakeel, S., Butcher, S.J. and Topf, M. (2014) Combined approaches to flexible fitting and assessment in virus capsids undergoing conformational change. *J. Struct. Biol.* **185**, 427–439

37 Wu, X. et al. (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

38 Brimacombe, R. et al. (2000) *The Ribosome: Structure, Function, Antibiotics, and Cellular Interactions*, pp. 151–163, ASM press, Washington, D.C.

39 Volkmann, N. (2014) The joys and perils of flexible fitting. *Adv. Exp. Med. Biol.* **805**, 137–155

40 Dourado, D.F.A.R. and Flores, S.C. (2016) Modeling and fitting protein-protein complexes to predict change of binding energy. *Sci. Rep.* 1–10

41 He, Y. et al. (2016) Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365

42 Plaschka, C., Hantsche, M., Dienemann, C., Burzinski, C. and Plitzko, J. (2016) Transcription initiation complex structures elucidate DNA opening. *Nature* **533**, 353–358

43 Villa, E. and Lasker, K. (2014) Finding the right fit: chiseling structures out of cryo-electron microscopy maps. *Curr. Opin. Struct. Biol.* **25**, 118–125

44 Mindell, J.A. and Grigorieff, N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347

45 Bartesaghi, A. et al. (2015) 2.2 Å resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. **348**, 1147–1152

46 Li, H. et al. (2016) Structure- and function-based design of Plasmodium-selective proteasome inhibitors. *Nature* **530**, 233–236

47 Eskelinen, E.L., Reggiori, F., Baba, M., Kovács, A.L. and Seglen, P.O. (2011) Seeing is believing: the impact of electron microscopy on autophagy research. *Autophagy* **7**, 935–956

48 Ahmed, A., Whitford, P.C., Sanbonmatsu, K.Y. and Tama, F. (2012) Consensus among flexible fitting approaches improves the interpretation of cryo-EM data. *J. Struct. Biol.* **177**, 561–570

49 Singharoy, A. et al. (2016) Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps. *Elife* **5**, 1–33

50 Ahmed, A. and Tama, F. (2013) Consensus among multiple approaches as a reliability measure for flexible fitting into cryo-EM data. *J. Struct. Biol.* **182**, 67–77

51 Liu, X., Li, M., Xia, X., Li, X. and Chen, Z. (2017) Mechanism of chromatin remodelling revealed by the Snf2-nucleosome structure. *Nature* **544**, 440–445

52 Bertram, K. et al. (2017) cryo-EM structure of a human spliceosome activated for step 2 of splicing. *Nature* **542**, 318–323

53 Tao, X., Hite, R.K. and MacKinnon, R. (2017) cryo-EM structure of the open high-conductance Ca^{2+}-activated K^+ channel. *Nature* **541**, 46–51

54 T. Terwilliger, Model-building into cryo-EM and low-resolution maps with map_to_model, https://www.phenix-online.org/version_docs/dev-2486/reference/map_to_model.html

55 Frenz, B., Walls, A.C., Egelman, E.H., Veesler, D. and DiMaio, F. (2017) RosettaES: a sampling strategy enabling automated interpretation of difficult cryo-EM maps. *Nat. Methods*, **in press**

56 Chen, M., Baldwin, P.R., Ludtke, S.J. and Baker, M.L. (2016) De Novo modeling in cryo-EM density maps with Pathwalking. *J. Struct. Biol.* **196**, 289–298

57 Miyashita, O., Kobayashi, C., Mori, T., Sugita, Y. and Tama, F. (2017) Flexible fitting to cryo-EM density map using ensemble molecular dynamics simulations. *J. Comput. Chem.* **38**, 1447–1461

58 Wu, X. and Brooks, B.R. (2016) Structure and dynamics of macromolecular assemblies from electron microscopy maps. *Intech* 243–262

59 Tan, R.K.Z., Devkota, B. and Harvey, S.C. (2008) YUP.SCX: coaxing atomic models into medium resolution electron density maps. *J. Struct. Biol.* **163**, 163–174

60 Tama, F., Miyashita, O. and Brooks, C.L. (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

61 Tama, F., Miyashita, O. and Brooks, C.L. (2004) Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J. Struct. Biol.* **147**, 315–326

62 McGreevy, R., Teo, I., Singharoy, A. and Schulten, K. (2016) Advances in the molecular dynamics flexible fitting method for cryo-EM modeling. *Methods* **100**, 50–60

63 Budkevich, T.V. et al. (2014) Regulation of the mammalian elongation cycle by subunit rolling: A eukaryotic-specific ribosome rearrangement. *Cell* **158**, 121–131

64 Sharma, M.R. et al. (2015) cryo-EM structure of the small subunit of the mammalian mitochondrial ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2736–E2736

65 U. of I. at Urbana-Champaign, Tutorial of Molecular Dynamics Flexible Fitting, http://www.ks.uiuc.edu/Training/Tutorials/science/mdff/tutorial_mdff-html/

66 Stone, J.E., Gullingsrud, J. and Schulten, K. (2001) A System for Interactive Molecular Dynamics Simulation. *ACM Symp. Interact. 3D Graph. ACM SIGGRAPH, New York* 191–194

67 Ratje, A.H. et al. (2010) Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* **468**, 713–716

68 Whitford, P.C. et al. (2011) Excited states of ribosome translocation revealed through integrative molecular modeling. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18943–18948

69 Muhs, M. et al. (2015) cryo-EM of ribosomal 80s complexes with termination factors reveals the translocated cricket paralysis virus IRES. *Mol. Cell.* **57**, 422–433

70 Schröder, G.F., Brunger, A.T. and Levitt, M. (2007) Combining Efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* **15**, 1630–1641

71 Wang, Z. and Schroder, G.F. (2012) Real-space refinement with Direx: from global fitting to side-chain improvements. *Biopolymers* **97**, 687–697

72 Song, F. et al. (2014) cryo-EM study of the chromatin fiber tetranucleosomal units. *Science* **334**, 376–381

73 Stone, J.E., McGreevy, R., Isralewitz, B. and Schulten, K. (2014) GPU-accelerated analysis and visualization of large structures solved by molecular dynamics flexible fitting. *Faraday Discuss* **169**, 265–283

74 Kirmizialtin, S., Loerke, J., Behrmann, E., Spahn, C.M.T. and Sanbonmatsu, K.Y. (2015) Using molecular simulation to model high-resolution cryo-EM reconstructions. *Methods Enzymol.* **558**, 497–514

75 Yokom, A.L. et al. (2016) Spiral architecture of the Hsp104 disaggregase reveals the basis for polypeptide translocation. *Nat. Struct. Mol. Biol.* **23**, 830–837

76 T. Terwilliger, Auto-sharpening cryo-EM maps with auto_sharpen, https://www.phenix-online.org/documentation/reference/auto_sharpen.html

77 Murshudov, G.N. (2016) Refinement of atomic structures against cryo-EM maps. *Methods Enzymol.* **579**, 277–305

78 Pettersen, E.F. et al. (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612

79 van Zundert, G.C.P. and Bonvin, A.M.J.J. (2016) Defining the limits and reliability of rigid-body fitting in cryo-EM maps using multi-scale image pyramids. *J. Struct. Biol.* **195**, 252–258

80 DiMaio, F. et al. (2015) Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nat. Methods* **12**, 361–365

81 Berman, H. et al. (2000) Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242

82 Pavel Afonine, Phenix.real_space_refine, https://www.phenix-online.org/documentation/reference/real_space_refine.html

83 Wang, R.Y.R. et al. (2016) Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta. *Elife* **5**, 1–22

84 Trabuco, L.G., Villa, E., Mitra, K., Frank, J. and Schulten, K. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683

85 Xu, X., Yan, C., Wohlhueter, R. and Ivanov, I. (2015) Integrative modeling of macromolecular assemblies from low to near-atomic resolution. *Comput. Struct. Biotechnol. J.* **13**, 492–503

86 Chen, V.B. et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21

87 Uversky, V. (2016) Dancing protein clouds: the strange biology and chaotic physics of intrinsically disordered proteins. *J. Biol. Chem.* **291**, 6681–6688

88 Yang, G., Zhou, R. and Shi, Y. cryo-EM structures of human g-secretase. *Curr. Opin. Struct. Biol.* **46**, 55–64

89 Narayanan, C., Bafna, K., Roux, L.D., Agarwal, P.K. and Doucet, N. Applications of NMR and computational methodologies to study protein dynamics. *Arch. Biochem. Biophys.*, **in press**

90 Fraser, J.S. et al. (2011) Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16247–16252

91 Vashisth, H., Skiniotis, G. and Brooks, C.L. (2013) Enhanced sampling and overfitting analyses in structural refinement of nucleic acids into electron microscopy maps. *J. Phys. Chem. B* **11**, 3738–3746

92 Qian, B. et al. (2007) High-resolution structure prediction and the crystallographic phase problem. *Nature* **450**, 259–264

93 Wu, X., Subramaniam, S., Case, D.A., Wu, K.W. and Brooks, B.R. (2013) Targeted conformational search with map-restrained self-guided Langevin dynamics: application to flexible fitting into electron microscopic density maps. *J. Struct. Biol.* **183**, 429–440

94 Gasteiger, J. (2016) Chemoinformatics: achievements and challenges, a personal view. *Molecules* **21**

95 Woods, S.P. et al. (2006) A battery approach for measuring neuropsychological change. *Arch. Clin. Neuropsychol.* **21**, 83–89

96 Chen, H. et al. (2013) A Combined Bioinformatics and chemoinformatics approach for developing asymmetric bivalent AMPA receptor positive allosteric modulators as neuroprotective agents. *Chem. Med. Chem.* **8**, 226–230

97 Lo Piparo, E. and Worth, A. (2010) Review of QSAR models and software tools for predicting developmental and reproductive toxicity. http://tcsweb3.jrc.it/DOCUMENTS/QSAR/EUR_24522_EN.pdf

98 Muratov, E.N., Lewis, M., Fourches, D., Tropsha, A. and Cox, W.C. (2017) Computer-assisted decision support for student admissions based on their predicted academic performance. *Am. J. Pharm. Educ.* **81**, 1–9

99 Chan, K.-Y., Trabuco, L.G., Schreiner, E. and Schulten, K. (2012) cryo-electron microscopy modeling by the molecular dynamics flexible fitting method. *Biopolymers* **97**, 678–686

100 Leaver-fay, A. et al. (2011) ROSETTA 3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574

101 Phenix, https://www.phenix-online.org/documentation/reference/cryo_fit.html

102 Maji, S., Shahoei, R., Schulten, K. and Frank, J. (2017) Quantitative characterization of domain motions in molecular machines. *J. Phys. Chem. B* **121**, 3747–3756

103 Rakesh, R. and Srinivasan, N. (2016) Improving the accuracy of fitted atomic models in cryo-EM density maps of protein assemblies using evolutionary information from aligned homologous proteins. *Methods Mol. Biol.* **1415**, 193–209

104 Joseph, A.P., Swapna, L.S., Rakesh, R. and Srinivasan, N. (2016) Use of evolutionary information in the fitting of atomic level protein models in low resolution cryo-EM map of a protein assembly improves the accuracy of the fitting. *J. Struct. Biol.* **195**, 294–305

105 Ovchinnikov, S. et al. (2017) Protein structure determination using metagenome sequence data. *Science* **355**, 294–298

106 Ovchinnikov, S., Kamisetty, H. and Baker, D. (2014) Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **2014**, 1–21

107 Bai, F., Morcos, F., Cheng, R.R., Jiang, H. and Onuchic, J.N. (2016) Elucidating the druggable interface of protein−protein interactions using fragment docking and coevolutionary analysis. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E8051–E8058

108 Chan, K.Y. et al. (2011) Symmetry-restrained flexible fitting for symmetric em maps. *Structure* **19**, 1211–1218

109 MDFF Cloud Computing, https://www-s.ks.uiuc.edu/Research/mdff/cloud.html