Advanced Review

Structural modeling from electron microscopy data

José Ramón López-Blanco and Pablo Chacón*

Cryo-electron microscopy is a powerful technique for the determination of three-dimensional (3D) structures of macromolecular machines, as it provides functional snapshots of biologically relevant complexes under near-physiological *in vitro* conditions. In this study, we review the computational algorithms developed to build macromolecular models from the information encoded in cryo-electron microscopy (EM) density maps. These modeling tools include fitting strategies to localize atomic structures into 3D maps, *de novo* methods to identify structural elements, and hybrid methods for the combination of multiple structural data from complementary biophysical techniques and other experimental sources. We also illustrate the power of EM-derived models in the atomic-level interpretation of the conformational changes of relevant macromolecular assemblies.

How to cite this article:
WIREs Comput Mol Sci 2015, 5:62–81. doi: 10.1002/wcms.1199

INTRODUCTION

In recent decades, the explosive growth of structural biology research has resulted in an exponential increase of the amount of information available for important biological systems. Detailed knowledge of protein structures is essential for our understanding of their functional mechanisms. To date, nearly 100,000 structures have been deposited in the Protein Data Bank (PDB). The major technique for the determination of atomic structures of proteins and other biomolecules is X-ray crystallography, although nuclear magnetic resonance spectroscopy is a powerful technique for relatively small proteins. The limiting factors of X-ray crystallography include protein expression, the amount of sample, crystallization, and the stability and homogeneity of the structure. These limitations are particularly troublesome for large macromolecular assemblies such as DNA and RNA polymerases, ribosomes, ATP synthases, and viral capsids. Moreover, these macromolecular machines adopt several conformations in solution, and their motions are directly coupled to numerous essential cellular functions. Electron microscopy (EM) techniques and, in particular, cryo-EM, have established themselves as the major structural biology techniques capable of imaging macromolecular complexes in their native aqueous environment.[1–3] Cryo-EM, and single particle analysis primarily, has demonstrated its capacity to distinguish coexisting conformational states, providing unique information on the functional mechanism and motions of macromolecular machines. Although the majority of the structures determined by EM range from medium (5–10 Å) to low (>10 Å) resolutions, recent advances in cryo-EM will lead to an increasing number of determined structures, particularly at medium and high (<5 Å) resolutions. In addition to the progress made in automated EM data acquisition and image processing software, the introduction of direct electron detectors (DED) has increased the quality of recorded EM images.[4,5] The substantial increase in the signal-to-noise ratio offered by DED represents an important contribution to this technique. For example, Bai et al.[6] recently demonstrated that near-atomic resolution structures can be obtained from only 20,000–30,000 single particles with only 2 days of data collection. Thus, although only 2000 structures determined by EM are currently available in the EM data bank (EMDB[7]), we strongly believe this technique will become a

*Correspondence to: pablo@chaconlab.org
Department of Biological Physical Chemistry, Rocasolano Physical Chemistry Institute, CSIC, Madrid, Spain
Conflict of interest: The authors have declared no conflicts of interest for this article.

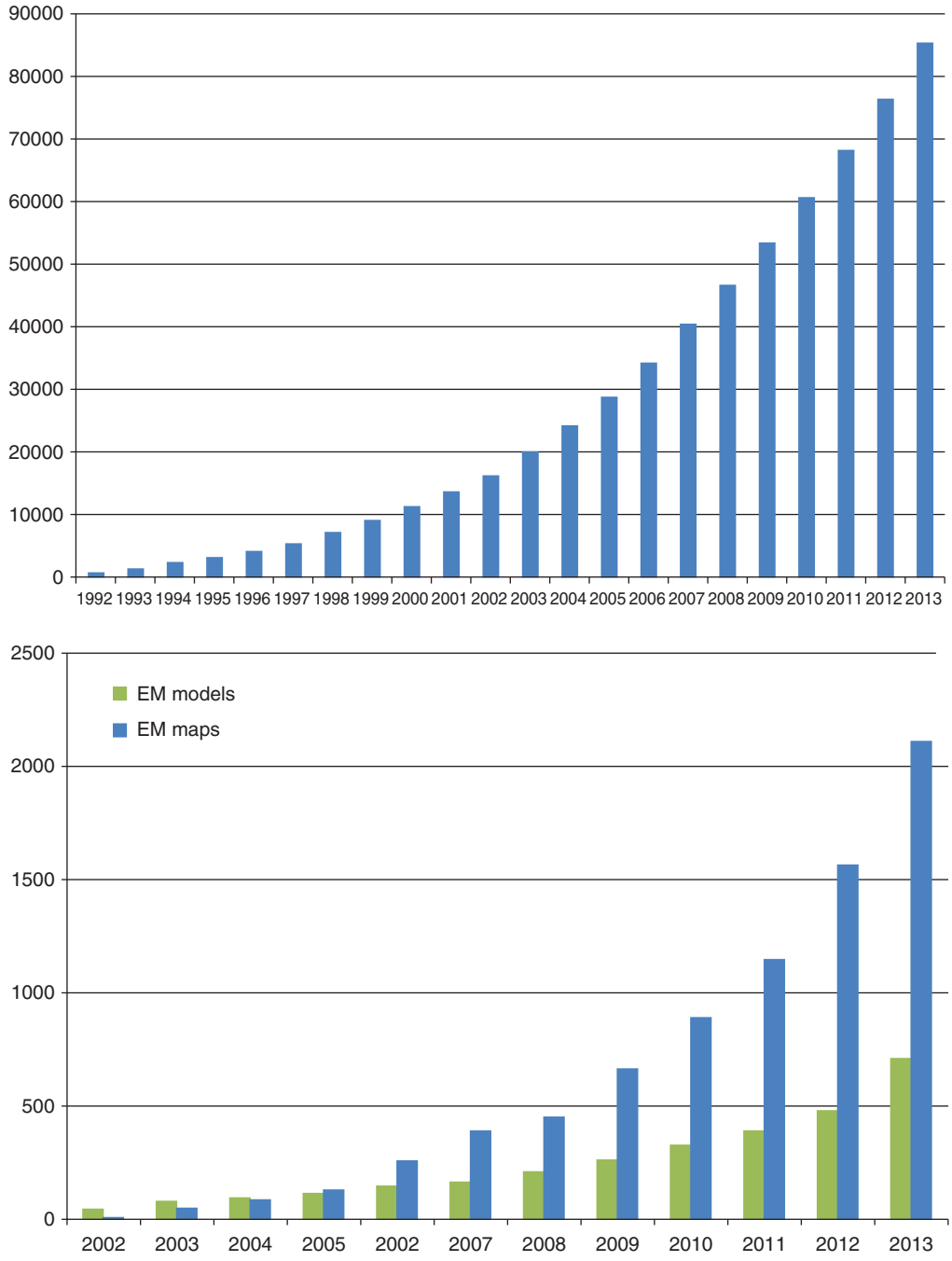

FIGURE 1 | Trends of (a) deposited atomic structures in the Protein Data Bank (PDB) and (b) deposited electron microscopy (EM) maps in the EM data bank (EMDB). Note that more than 80% of the deposited maps correspond to single-particle reconstructions.

major contributor of structural knowledge in the near future. Figure 1 shows the exponential growth of atomic structures deposited in the PDB (top panel) and the substantial growth of deposited EMDB maps, although at more modest levels (bottom panel). Noteworthy, the number of deposited structures modeled from EM data (green bars) follows an increasing trend, as observed for the experimentally determined structures (blue bars). This illustrates the relevance of the modeling in the near-atomic interpretation of EM three-dimensional (3D) reconstructions.[1,8–10] Despite considerable progress has been achieved in the last decades, EM is still a field under development. Major limiting factors include the assessment of the EM map quality and the lack of standards for validating model accuracy. Fourier shell correlation (FSC) is

the standard metric to estimate the map resolution. However, many of the published works overestimated the resolution with FSC due to overfitting of the high-frequency noise during the reconstruction process.[11] Moreover, the resolution is heterogeneous across the EM map and therefore should be locally estimated. There are promising methods for quantifying the local resolution[12,13] that we believe they will be quickly adopted by the community. Another resolution-dependent problem is the model validation, precautions need to be taken to avoid model overfitting by introducing too many refinable parameters relative to the experimental EM data. Initiatives such as 'Microscopy Validation Task Force Meeting'[14] where the experts in the field convened to collect recommendations and develop consensus on validation task for 3D maps and models will help to raise the standards to levels of X-ray crystallography.

Comprehensive reviews of structure determination using EM, including cryo-EM and electron tomography, can be found elsewhere. Here, we review the computational methods for modeling 3D-EM data and for bridging the resolution gap between atomic and low-medium resolution structural information. First, we describe the methodologies developed to fit the atomic structures of proteins and nucleic acids into an EM density map. These methodologies range from early rigid-body approximations to recent methods to characterize conformational changes observed by EM. In a second modeling scenario, in the absence of available atomic structures, different computational methods can be applied for *de novo* modeling of the EM density map. Once the resolution is better than approximately 10 Å, secondary structure elements (SSEs) such as α-helices can be computationally identified. At resolutions below approximately 5 Å, crystallographic refinement techniques can be directly borrowed to produce atomic models.[15–17] Finally, hybrid methods that combine the different available structural information represent a promising approach for extracting structural information from 3D EM reconstructions. These integrative approaches incorporate modeling constraints from other complementary biophysical techniques (e.g., small-angle X-ray scattering, fluorescence resonance energy transfer, etc.) or any other source of structural information (e.g., crosslinking, mutagenesis, protein–protein predictions, etc.). Figure 2 illustrates these different modeling scenarios, which will be addressed in the following sections.

MACROMOLECULAR FITTING

The first attempts to interpret EM maps beyond their nominal resolution fitted crystal structures into maps by visual inspection. However, computational tools have been developed to perform the fitting in a reliable and reproducible manner. The most frequently used approach utilizes an exhaustive rigid-body search of the best relative conformation of a given atomic structure inside the EM density. This relatively simple problem becomes cumbersome if more than one atomic structure must be considered. Multi-body fitting includes specific strategies to fit several atomic components into a density map while avoiding collisions. All of these rigid-body approaches are sufficient if the conformational differences between atomic structures and the macromolecule visualized by EM are small. In contrast, to characterize larger conformational changes, conformational flexibility must be considered. To introduce conformational variability into the fitting process, several flexible fitting tools have been developed. In the following subsections, we will describe the central principles underlying the most commonly used methods to fit macromolecular structures into EM maps and we will refer to the original articles for the details.

RIGID-BODY FITTING

Rigid-body docking is a well-established technique with several advanced approaches available. Comprehensive reviews of rigid-body docking methods have been previously presented.[18,19] In essence, these tools perform an automated 6D search of all possible relative rotations and translations to maximize an electron density-based cross-correlation (CC) function. This CC function is typically defined as the scalar product between the EM experimental map, ρ^{EM}, and a low-pass-filtered version of the atomic structure, ρ^{Model}, which can be mathematically represented as follows:

$$\mathrm{CC} = \frac{\sum \left(\rho_i^{\mathrm{EM}} \cdot \rho_i^{\mathrm{Model}} \right)}{\sigma^{\mathrm{EM}} . \sigma^{\mathrm{Model}}} \qquad (1)$$

where the sigmas correspond to the standard deviation and i is the volumetric map element (voxel). Several of the exhaustive 6D fitting methods, such as Colores,[20] COAN,[21] EMfit,[22] BCL:EM-fit,[23] and FoldHunter,[24] include the CC as quality-of-fit criteria. Other fitting tools, such as DockEM[25] and Mod-EM,[24] use CC variants that emphasize local rather than overall shape features, which can be particularly useful to avoid false positives or when a small component is fitted into a larger map.

The major disadvantage of these exhaustive search methods is their high computational cost. Nevertheless, the exhaustive CC-based search can be dramatically accelerated by taking advantage of

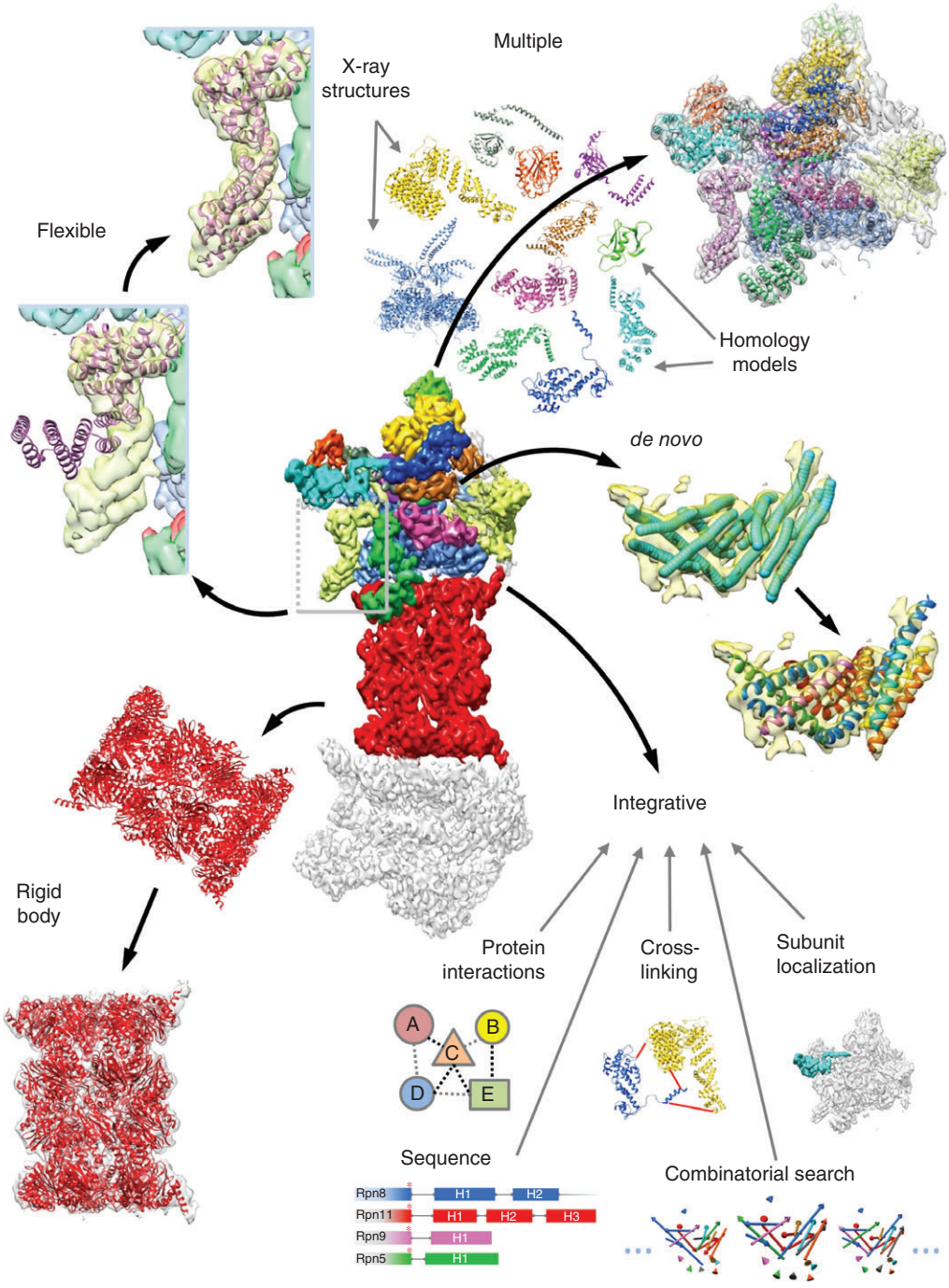

FIGURE 2 | Scheme of possible modeling scenarios. The different cryo-electron microscopy modeling alternatives are illustrated using a proteasome map.

the convolution theorem and the fast Fourier transform algorithms. CoLoRes[20] and URO[26] accelerate the translational part of the search and reduce the runtimes to hours or minutes. In addition, CoLoRes introduces a Laplacian image-processing filter to enhance the fitting contrast. Alternatively, ADP_EM[27] accelerates the rotational part of the search using spherical harmonics, which reduces the fitting to a few minutes. Examples of rigid-body fitting using this approach are shown in Figure 3. The more recently developed gEMfitter[30] accelerates the fitting even more using the texture memory of graphics processing units (GPUs), including a local CC scoring function. Other exhaustive methods use alternative strategies to reduce

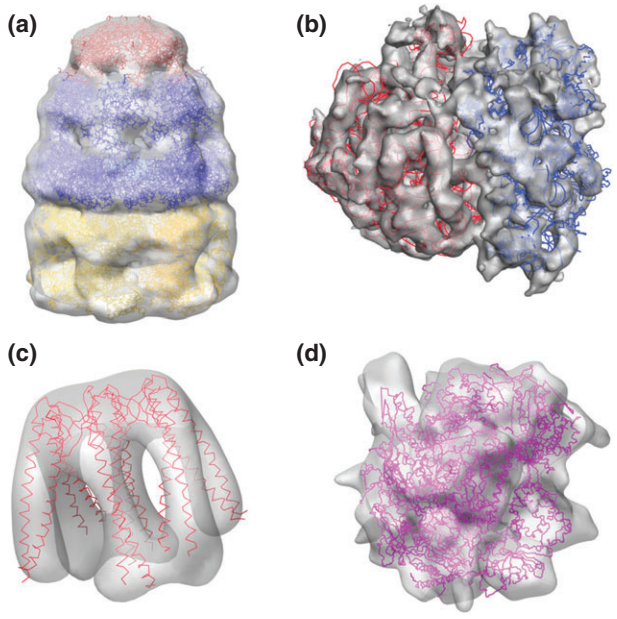

FIGURE 3 | Rigid-body fitting results with experimental electron microscopy (EM) data using ADP_EM. (a) *Escherichia coli* GroES-ADP7-GroEL-ATP7 at 23.5 Å (EMDB ID: 1046; PDB ID: 1ML5); ADP and ATP bound GroEL subunits have been independently fitted to reconstruct the *cis* and *trans* heptameric rings of the complex. For GroES, the entire heptamer was used. (b) Fitting of the 30S and 50S subunits into the *E. coli* ribosome map at 14 Å (EMDB ID: 1046; PDB ID: 1GIX/1GIY). Single-molecule docking of prefoldin (c) at 23 Å28 (PDB ID: 1L6H) and of yeast RNA polymerase II (d) at 15 Å29 (PDB ID: 1FXK). Reproduced with permission from Ref 19.

the computational time; e.g., COAN, EMfit, and FoldHunter perform an initial coarse exploration to reduce the search space, and then refine only the best identified conformations.

Although the majority of these approaches use CC-based scoring functions, alternatives exist. For example, the pioneering EM package Situs[31,32] uses a feature-points-based approach. After reducing both the atomic structure and the EM map into sets of characteristic points with the aid of neural network techniques, the fitting is performed by minimizing the distance between these points. Following also a reductionist strategy, several scores have been introduced by reducing the electron density into isosurface normals[33] or Gaussian functions[34] (for a review of fitting functions[35]). Finally, molecular visualization programs such as Chimera[36] or Sculptor[37] also include interactive fitting modules, together with segmentation and other useful modeling tools.

MULTI-BODY FITTING

A special case of rigid-body fitting occurs in cases when multiple assembly components must be localized into single cryo-EM maps. Particularly at low resolutions or with a high number of components, the search space becomes too large to be exhaustively explored. The common starting point for nearly all multi-body methods is the generation of a high number of candidate configurations that are subsequently refined, typically by using a CC variant. The *collage* tool[37] of Situs simultaneously maximizes the global CC of all of the components from an initial configuration using a conjugate gradient optimization method. The MultiFit program[38] exploits a combinatorial search to optimize both the local CC and the shape complementarity between interacting subunits (see Figure 4 for details). In MOSAEC,[39] approximately 1000 random configurations evolve via a genetic algorithm to optimize efficiently a fitness function based on neuronal network feature-point representations. Zhang et al.[40] also simplified input structures as sets of feature points but used an efficient mathematical programming procedure to generate an ensemble of candidate models that locally considered the geometry and the neighboring density. In the gmfit[34] program, input structures of the components are simplified as summations of a few 3D Gaussian functions. The Gaussian approximation allows the rapid steepest-descent minimization of initial configurations under a force field in which every component is attracted by the target density map and repelled by the other components. Symmetry can also be included as an additional restraining force. Finally, the EMLZerD[41] and ATTRACT-EM[42] methods rely on an initial protein–protein docking step to generate the initial candidate configurations. The docking step of EMLZerD generates pairs of components with high-surface complementarity that are subsequently fitted into the EM map using 3D Zernike descriptors, i.e., a compact rotationally invariant representation of 3D surfaces. In contrast, in ATTRACT-EM, all components are simultaneously docked, maximizing a smooth Gaussian overlap function between the subunits and the electron density map.

FLEXIBLE FITTING

When significant conformational differences are present between the atomic structures and the target map, the classical rigid-body approach is no longer valid. In this case, the number of fitting variables greatly exceeds the number of experimental observables encoded in the EM map; therefore, special attention must be given to overfitting. All flexible fitting algorithms require a method to explore new conformations. The sampling engines are typically based on either Molecular Dynamics[43] (MD) or

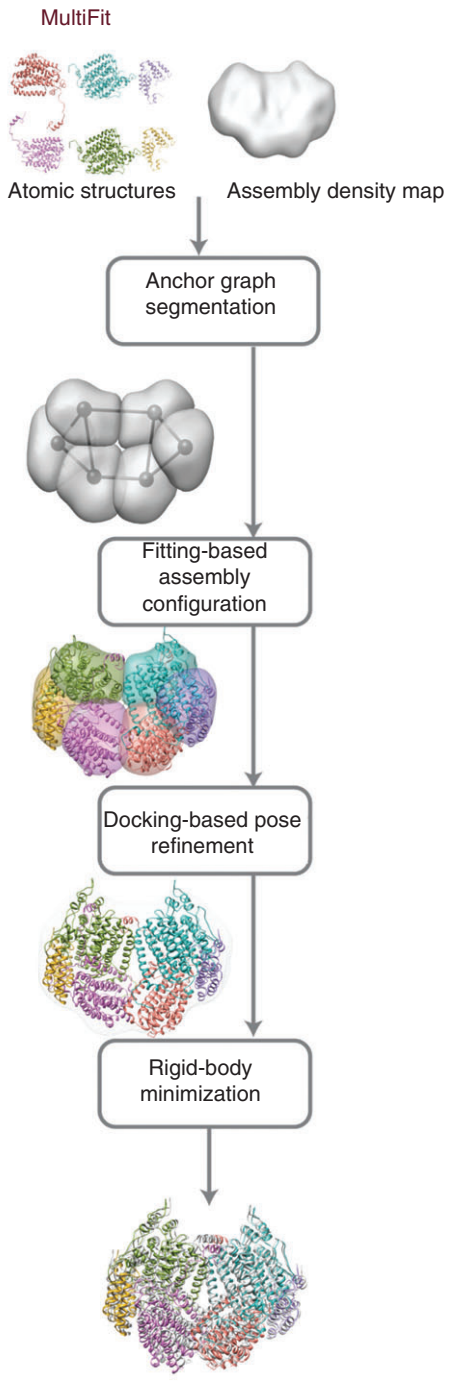

FIGURE 4 | Overview of the multi-fitting protocol used in MultiFit. The input is a simulated density map of the Methane monooxygenase (MMO) hydroxylase complex at 20 Å resolution (gray) and the atomic models of the subunits (colors). First, the map is segmented into six regions (gray transparency) and the corresponding anchor graph (black) is calculated. Subsequently, the subunits are assembled into segmented regions by rigid-body fitting (colors). The conformations are then refined using a docking strategy. Finally, the model is fitted into the map (gray transparency). The resulting model is superposed onto the native complex (gray) at the bottom. Reproduced with permission from Ref 44.

Normal Mode Analysis[45] (NMA) methodologies. In MD-based approaches, the new conformations are generated by numerically integrating the Newtonian equations of motion under different flavors of physics-based force fields. Following this strategy, the earliest tool for flexible fitting, Situs,[46] used a classical simulation software (XPLOR) to minimize the differences between the reduced feature-point representations of both the atomic structure and the EM map. The conformational flexibility of RNAP has been successfully modeled using this approach to reveal functional implications in the initiation of transcription.[47] Based on standard MD packages, the MDFF[48] and MDFIT[49] tools perform flexible fitting by including an additional potential energy term in the force field to morph the atomic model into the target EM map. In MDFF, the gradient of the density map drives the fitting with additional harmonic constraining terms to preserve the stereochemical quality and prevent overfitting. In contrast, MDFIT retains the tertiary contacts present in the initial structure without special constraints using a simpler structure-derived force field. Impressive results have been obtained for the ribosome, revealing several functional conformational states of translational events[49–53] (see Figures 5 and 6). MDFF is the most popular approach and it has been used for flexible fitting with other key systems including, among others, the proteasome,[55,56] the HIV-capsid,[57] adenovirus,[58] and even DNA-origami nanotechnologic objects.[59] The major disadvantage of MD-based methods is their high computational cost. In this context, several approaches have been developed to reduce the computational burden. For example, YUP.SCX[60] uses an efficient MD-based simulated annealing protocol that minimizes a simplified potential energy with restraints, ICFF[61] is formulated in internal coordinates, and TAMDFF[62] enhances the conformational sampling of MDFF. Moreover, more efficient alternatives have been developed using NMA as sampling engine. In this case, the macromolecular system is considered an elastic network of harmonic springs that holds the atoms together and vibrates around a given equilibrium conformation. According to this simple mechanical model, the exact solutions for Newton's equations can be analytically obtained as a set of atomic displacement vectors. Such vectors, or normal modes, are naturally sorted according to their vibration frequency. High-frequency (high-energy) modes represent localized displacements, whereas low-frequency (low-energy) modes correspond to collective conformational changes. All NMA-based approaches are focused on these collective modes to reduce the number of search variables. NMFF,[63] which was the first NMA-based flexible fitting tool,

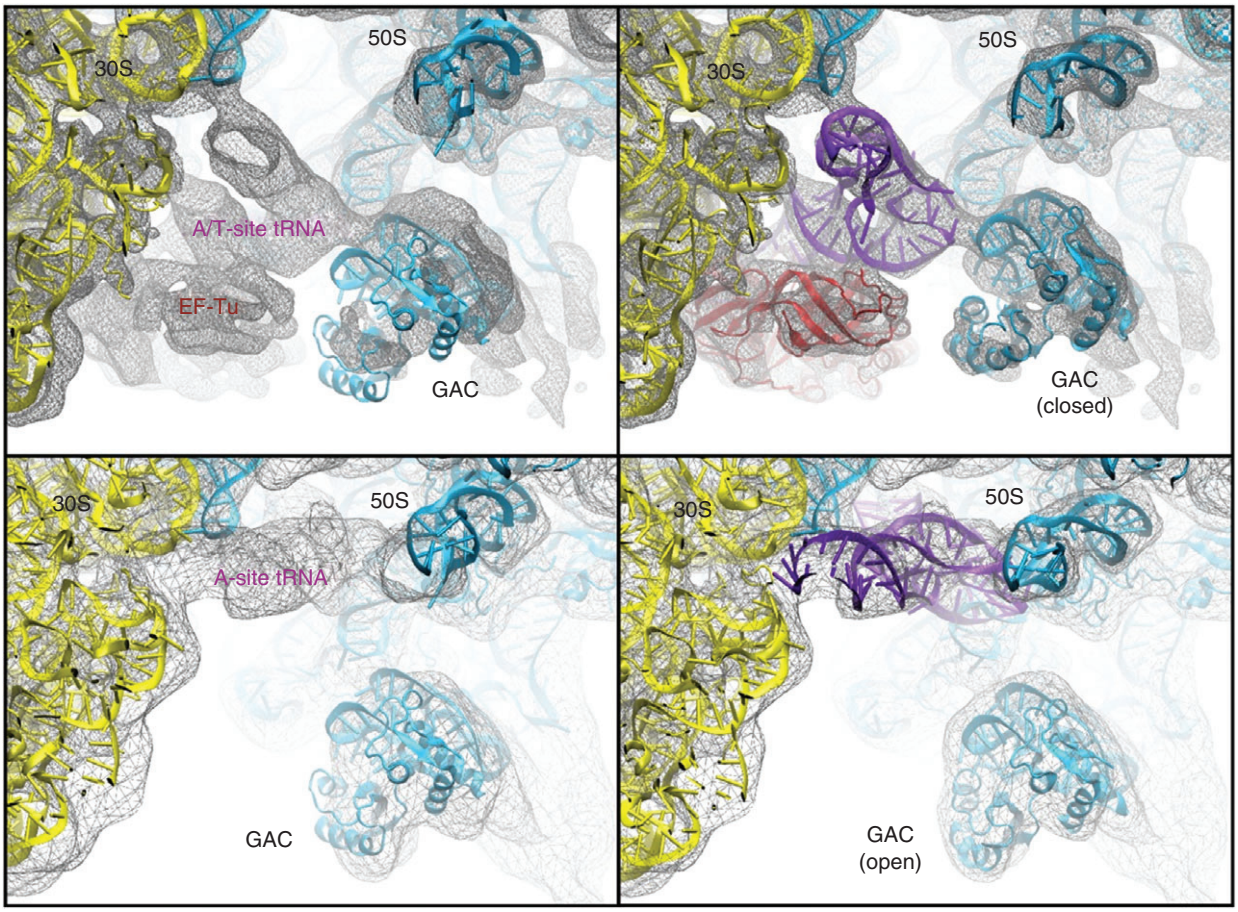

FIGURE 5 | Fitting into the topology constraint (TC)-bound ribosome cryo-EM map at 6.7 Å resolution by means of MDFF. Conformational dynamics of the GTPase-associated center. Shown are differences in the conformation of the GTPase-associated center between the TC-bound ribosome [electron microscopy (EM) map at 6.7-Å resolution, top], and the accommodated ribosome (EM map at 9 Å resolution, bottom). Rigid-body docked structures into the corresponding maps, used as initial coordinates for flexible fitting, are shown on the left; flexibly fitted structures are shown on the right. Reproduced with permission from Ref 37.

employed a linear combination of the first 20 or 30 lowest frequency modes to conform the target map iteratively. This approach has been applied to several large and functionally relevant conformational changes of the elongation factor G bound to the ribosome, the *Escherichia coli* RNA polymerase, and the cowpea chlorotic mottle virus.[64] Hinsen et al.[65] used a map-derived force to select the best modes that transform the atomic structures of the calcium ATPase into the target conformation. Although these Cartesian coordinates-based approaches yield satisfactory results, their straight-line molecular motions lead to unfeasible distortions in bond lengths and angles that eventually can produce overfitting problems.[54] In the recently released iMODFIT tool[54], the normal modes are defined in internal coordinates (dihedral angles), providing a more natural and effective approach for modeling the conformational changes. Consideration of internal coordinates minimizes the potential distortions by implicitly preserving the model geometry while substantially reducing the number of variables.[66] In iMODFIT, the lowest-frequency modes are randomly selected to generate trial conformations. Only those conformations that increase the CC are accepted. Different groups have utilized this tool to obtain accurate flexible fittings for ATP synthase,[67] yeast vacuolar ATPase,[68] and coxsackievirus.[69] We also illustrate iMODFIT performance with additional experimental maps in Figure 7.

Other alternative methods provide the necessary conformational variability and retain the correct stereochemistry using different approaches. Jolley et al.[70] initially produced these conformations using constrained geometric simulations, which are subsequently accepted or rejected according to a Monte Carlo algorithm with a pseudo-energy derived from the CC coefficient. In addition to a Monte

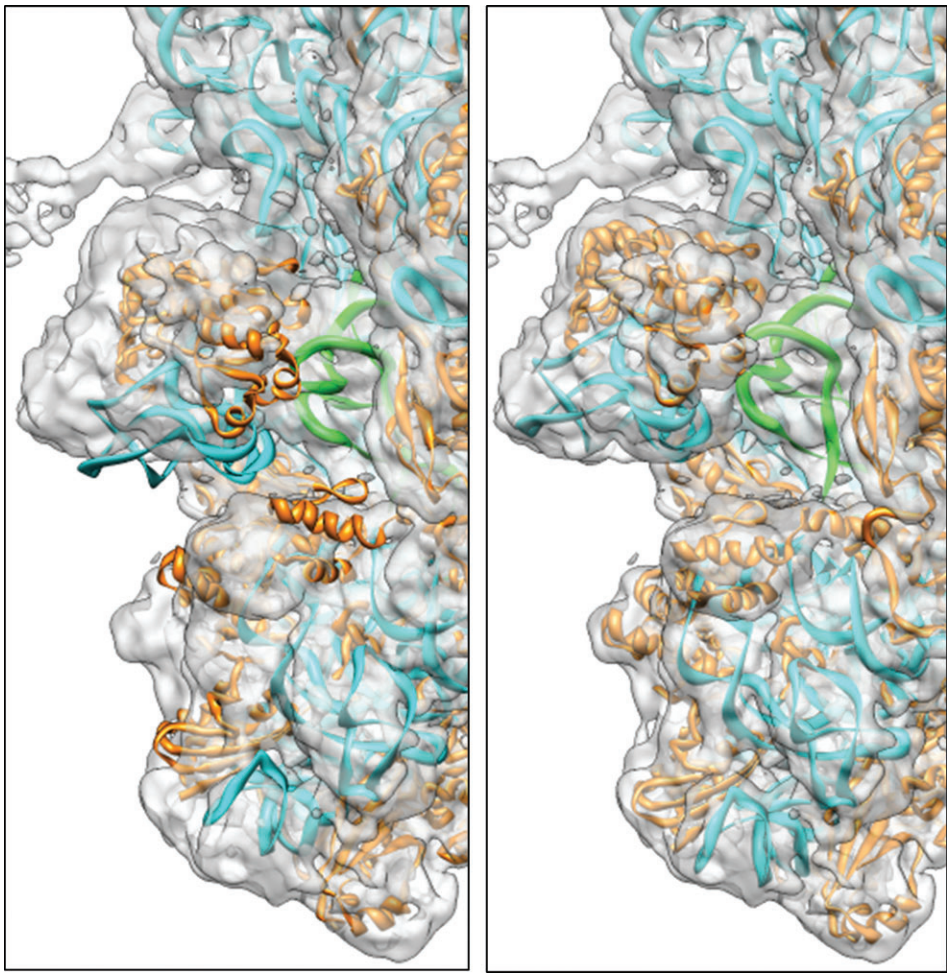

FIGURE 6 | Ribosome flexible fitting. On the left, the initial rigid-body fitting of the TIpre atomic structure is shown inside an electron microscopy map of the TIpost conformation (EMDB ID: 1799) with a resolution of 7.6 Å. On the right, the final fitted results using MDfit are shown.[49] An essentially identical fitted structure was obtained using iMODFIT.[54] The images represent the regions of the 30S head and L1 of the ribosome, where major differences were observed. The color codes are orange for protein, blue for rRNA, green for tRNA, and red for EFG. Reproduced with permission from Ref 53.

Carlo search, Flex-EM[71] applies conjugate-gradient minimization and simulated annealing stages to a series of rigid-body subdivisions of the structure. In DireX,[72] the conformations generated using a geometric-based sampling[73] are biased toward the target map with a random force. This method utilizes a deformable elastic network to prevent overfitting. EMFF[74] is also based on a modified elastic network model to maintain pseudo-bonds and secondary structures while permitting large-scale conformational changes. The target conformation is iteratively attained by Newton-Raphson minimization of the elastic network energy, which includes an extra term to penalize the squared differences between the target map and the atomic structure. S-flexfit[75] uses the structural variability within a protein superfamily to build models that are then selected by CC with the target map. The MOSAICS-EM[76] approach represents the macromolecule as a small number of rigid segments connected with flexible loops to fit atomic structures directly into 2D class averages, i.e., disregarding the 3D reconstruction stage. Finally, other authors propose a consensus fit among different fitting approaches.[77]

HOMOLOGY MODELING

Comparative modeling platforms can generate reliable models from homologous structures with a sequence identity greater than approximately 50%.[78,79] Homology models can be fitted into the density map using the approaches detailed in the sections above when the corresponding atomic structure is not available. For example, the Moulder-EM[80] protocol of Topf

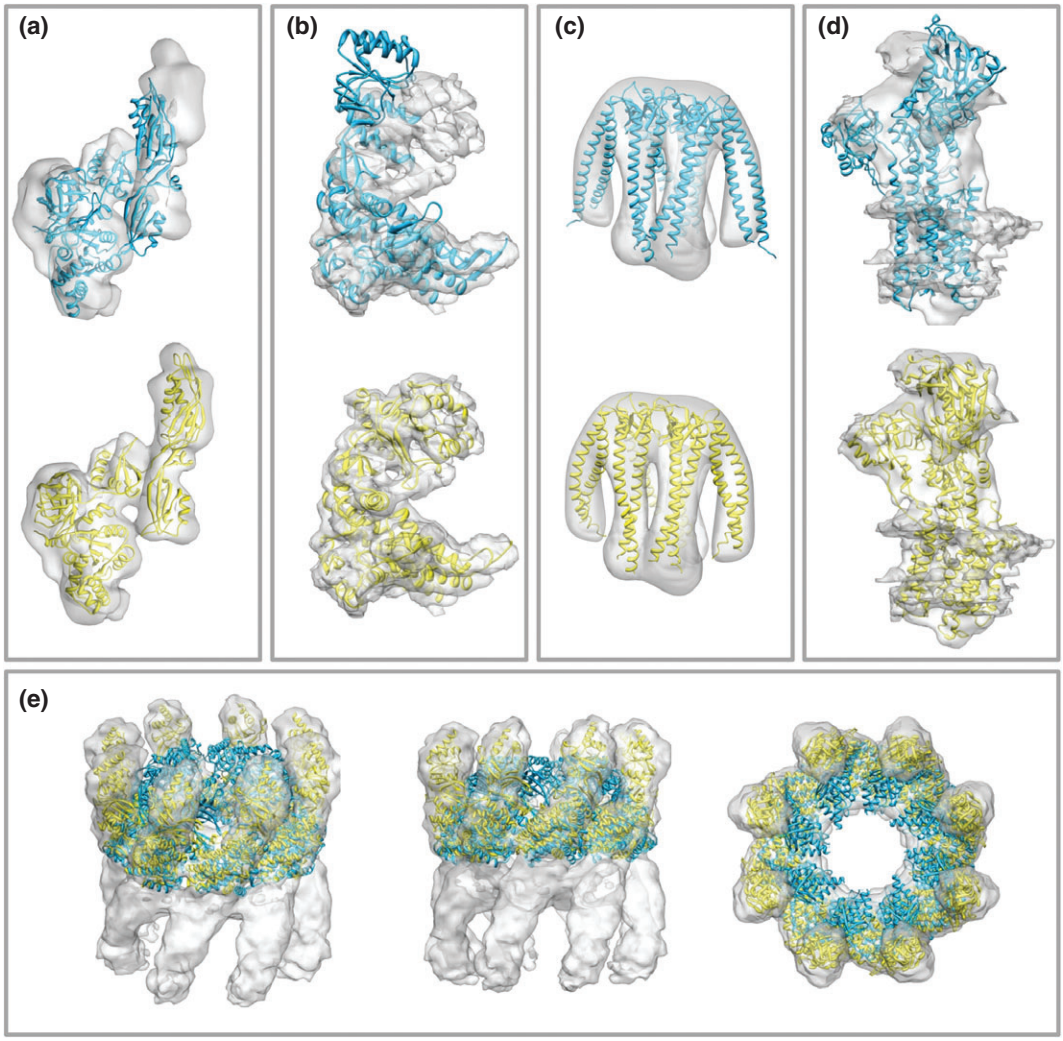

FIGURE 7 | Flexible fitting of representative experimental maps using iMODFIT. The initial conformation was obtained by rigid-body fitting between the initial atomic structures (cyan) and their corresponding experimental electron microscopy (EM) maps (transparent) using ADP_EM.[27] The final fitted models are represented as yellow ribbons. The examples shown include the following: (a) elongation factor G (PDB ID: 1FNM; EMDB ID: 1364); (b) the GroEL monomer (PDB ID: 1SX4; EMDB ID: 1181); (c) prefoldin (PDB ID: 1FXK); (d) calcium ATPase (PDB ID: 1SU4); (e) the thermosome (PDB ID: 1A6D; EMDB ID: 1396); and (f) RNA polymerase II (PDB ID: 1YLV; EMDB ID: 1283). Reproduced with permission from Ref 53.

et al. has been used for the refinement of a homology model built for the upper domain of the P8 capsid protein of rice dwarf virus within a 6.8-Å resolution cryo-EM map. These authors also demonstrated that the fitting can be helpful in improving the accuracy of comparative models.[24] Other approaches combine both supervised and automated structure refinement protocols to improve locally the homology models of large macromolecular assemblies. Rosetta[81] software includes protocols to refine comparative models and low-resolution Cα traces using density maps as a guide. In the multi-scale method EM-IMO,[82] the problematic regions are first identified by visual inspection and then iteratively refined using the density map information, e.g., by searching low-energy conformations of rigid SSEs connected by flexible loops. As a final step, MDFF is applied to refine the structural details. Although homology modeling is a widely used alternative, important errors can also be produced in regions of the protein that share little identity; such errors can jeopardize the EM modeling procedure.

When suitable templates cannot be detected by sequence-based or threading techniques, it is still possible to identify representative structural scaffolds suitable for fitting. Using the EM map as a template, SPI-EM[83] and FOLD-EM[84] screen the CATH or SCOP protein domain databases, respectively, to identify structural homologs. In contrast, EMatch[85] uses a previously detected SSEs (see *De novo modeling*

section) to obtain the greatest common alignment with the SCOP database. This method explicitly considers errors in the α-helix extraction stage and has been validated with experimental cryo-EM maps of GroEL.[86,87]

DE NOVO MODELING

When atomic models are not available and the map resolution is sufficiently high, computational algorithms can identify the SSEs in EM maps. Once the resolution is greater than approximately 10 Å, the cylindrical densities of α-helices become readily recognizable; for β-sheets, this occurs at approximately 6 Å. Alternatively, such high-resolution density maps can contain sufficient information to trace the backbone. If the resolution is high enough (<4 Å), it is possible to construct reliable atomic models directly from the cryo-EM density map with the help of X-ray crystallographic tools such as COOT,[88] REFMAC,[89] or Phenix[90] (see Table 1). For example, high-resolution structures of the Mm-cpn chaperonin,[108] the TRPV1 channel,[15] the F420-reducing [NiFe] dehydrogenase,[16] and the yeast mitochondrial large ribosomal subunit[17] have been recently solved from EM data using COOT.

SECONDARY STRUCTURE DETECTION

HelixHunter[96] was the first automated tool for α-helix detection. This method is based on the correlation of a prototypical straight helix with the map to enhance the identification of helical densities. Alternatively, in HelixTracer,[109] the α-helices are detected by gradient analysis and modeled as quadratic splines to consider their curvature explicitly. Using a genetic algorithm, the VolTrac[107] tool optimally characterizes the curvature and length of the helical regions by placing into the map short cylindrical templates that are then enlarged in an adaptive bidirectional expansion stage. The practical applicability of VolTrac has been demonstrated in the modeling of the N-terminal domain of the ryanodine receptor[110] and in the helical bundle of the proteasome.[111] Once the map resolution reaches approximately 6 Å, the β-sheets become distinguishable as thin planar surfaces. The Sheetminer method recognizes disk-shaped regions to identify β-sheets.[104] At resolutions below 5 Å, the constituent individual strands are discernible and can be resolved by the Sheettracer tool.[105]

Additional approaches enable the simultaneous identification of SSEs; e.g., SSELearner[100] uses a machine-learning approach to identify automatically α-helices and β-sheets using information from existing EM maps in the EMDB. Other approaches include the skeletonization of the EM map, i.e., the simplified geometric representation of the overall density shape, as a key step in the SSE identification process. The method described by Yu and Bajaj[112] rapidly generates skeletons of helices and sheets by considering the overall distribution of the gradient vectors within a local spherical window. SSEhunter[97] integrates several scores, skeletonization, local geometry, and CC with a prototypical helix to obtain α-helix or β-sheet propensity measures. Figure 8 shows representative examples of SSE identification using the SSEhunter approach with several experimental EM maps. Based on the local shape characteristics of α-helices and β-sheets, SSEtracer[101] performs a series of local feature analyses together with the sparseness and local thickness of externally generated skeleton points to detect the SSEs. Finally, backbone traces can be obtained with SSEhunter[98] once the resolution is greater than 5 Å using a combination of sequence-based secondary structure predictions with an improved skeletonization algorithm. This approach facilitated the direct *de novo* backbone tracing from 4 to 5 Å cryo-EM maps of the GroEL chaperonin[113] and the capsid of the infectious ϵ15 particle.[114] An alternative algorithm, *Pathwalking*,[99] can trace Cα backbones by solving the 'Traveling Salesman Problem' from a set of pseudo-atoms clustered inside the EM densities.

Once the SSEs are extracted from the map, their directionality and connectivity, i.e., their topology, can be determined. However, the number of potential topologies satisfying a given set of SSEs can be enormous. Despite several efforts to register the SSEs into EM maps (e.g., Wu et al.,[115] Al Nasr et al.,[116] and Biswas et al.[117]) the solutions only appear reliable for protein transmembrane regions (TM). These proteins constitute a highly populated group that accounts for approximately 50% of the contemporary drug targets. The α-helical nature of their transmembrane domains facilitates the automated registration of the sequence segments and map-detected helices. The method described by Enosh et al.[118] produces a small number of feasible configurations upon ranking all possible assignments with a score function derived from the loop structures found in soluble α-helix bundles. This method accurately predicted the structures of TM proteins within 1.5–3.5 Å from the native state in all evaluated cases.[119] An alternative method[120] also successfully recognizes TM helices with an RMSD of 3–5 Å by first identifying the TM regions from the sequence, then assigning buried and lipid-exposed faces of these regions, and finally assembling a helical bundle based on the geometric

TABLE 1 | Available Tools for Structural Modeling Based on EM Data.

	Method	Features	Availability
Rigid-body fitting	3SOM[33]	Surface-based	www.russelllab.org/3SOM
	ADP_EM[27]	Fast exhaustive spherical-harmonics accelerated	chaconlab.org/methods/fitting/adpem
	BCL:EM-fit[23]	CC-based (web server)	www.meilerlab.org/bclcommons
	Chimera[36]	CC-based (GUI)	www.cgl.ucsf.edu/chimera
	COAN[21]	CC-based with confidence intervals	coan.burnham.org
	DOCKEM[25]	Local CC-based	www.msg.ucsf.edu/local/programs/DockEM/DockEM_Instructions.htm
	EMAN	CC-based (foldhunter[24]) includes a parallel version	www.msg.ucsf.edu/local/programs/eman
	EMfit[22]	Multiple fitting criteria	bilbo.bio.purdue.edu/viruswww/Rossmann_home/softwares/emfit.php
	gEMfitter[30]	Ultra-fast exhaustive multi-processor/GPU accelerated	gem.loria.fr/gEMfitter
	MODELLER[91]	Local CC-based (*Mod-EM*)[24]	topf-group.ismb.lon.ac.uk/Software.html
	Sculptor[37]	CC-based (GUI)	sculptor.biomachina.org
	SITUS[92]	Feature-points (*matchpt*, *qvol*, *qpdb*)[31] CC based (*colores*)[20]	situs.biomachina.org
	UROX/VEDA[93]	CC and reciprocal space based (GUI)	sites.google.com/site/xsiebert/urox
Multi-body fitting	ATTRACT-EM[42]	Gaussian-CC-based	Available upon request
	EMLZerD[41]	Surface complementarity based	Available upon request
	Gmfit[34]	Gaussian-CC-based	strcomp.protein.osaka-u.ac.jp/gmfit
	MOSAEC[39]	Genetic algorithm optimization (GUI)	sculptor.biomachina.org
	MultiFit[38]	CC-based with docking refinement (web server)	modbase.compbio.ucsf.edu/multifit
	SITUS[92]	Off-lattice multi-body Powell optimization (*collage*)[37]	situs.biomachina.org
	Zhang's[40]	Mathematical programming based	Available upon request
Flexible fitting	DireX[72]	Geometric based sampling with an elastic network	www.schroderlab.org/software/direx
	EMFF[74]	Elastic network and Newton-Raphson based (web server)	enm.lobos.nih.gov/emff/start_emff.html
	MODELLER	CC-based MC/Conjugate gradients/SA (*Flex-EM*)[71]	http://salilab.org/Flex-EM
	ICFF/MMB[61]	Internal coordinates with fully atomic forces (fast)	simtk.org/home/rnatoolbox
	iMODFIT[54]	Internal coordinates NMA and CC-based (fast)	chaconlab.org/imodfit
	Jolley's[70]	Constrained geometric simulations and MC-based	http://flexweb.asu.edu
	MDFF[48]	MD-based with constraints	www.ks.uiuc.edu/Research/mdff
	MDFIT[49]	MD-based under simplified structure-based force field	mdfit.lanl.gov
	MOSAICS-EM[76]	Direct flexible fitting into 2D class-averages	www.cs.ox.ac.uk/mosaics
	NMFF[63]	Cartesian NMA and CC gradient based	mmtsb.org/software/nmff.html
	NORMA[94]	Cartesian NMA and CC based	www.igs.cnrs-mrs.fr/elnemo/NORMA

TABLE 1 | Continued

	Method	Features	Availability
	S-flexfit[75]	Based on Structural variability and principal component analysis	biocomp.cnb.csic.es/Sflexfit
	SITUS[92]	Feature-points (*qplasty*)[46]	situs.biomachina.org
	YUP.SCX[60]	MD-SA-based with a simplified potential energy and restraints (fast)	www.harvey.gatech.edu/YammpWeb/userman/yupscx/yupscx.shtml
Homology modeling	EM-IMO[82]	Supervised and automated structure refinement	wiki.c2b2.columbia.edu/honiglab_public/index.php/Software:cryoEM
	EMatch[85]	Fold recognition based on secondary structure elements	Available upon request
	Fold-EM[84]	Find structural homologs from SCOP domain database	robotics.stanford.edu/mitul/foldEM
	MODELLER[91]	Genetic algorithm optimizes sequence alignment and CC (*Moulder-EM*)[80]	salilab.org/modeller
	ROSETTA[81]	Comparative modeling assisted by electron density maps	www.rosettacommons.org
	SPI-EM[83]	Find structural homologs from CATH domain database	Available upon request
De novo modeling	EM-Fold[95]	MC-based topology determination from externally identified SSEs (web server)	www.meilerlab.org/bclcommons
	EMAN	*Helixhunter*[96]: α-helix detection (EMAN1) *SSEhunter*[97]: α-helix, β-sheet, and trace[98] detection (EMAN1 & Gorgon) *SSEBuilder*: builds α-helices and β-sheets (EMAN1 & Chimera) *Pathwalking*[99]: Trace detection (EMAN2)	blake.bcm.edu/emanwiki
	Gorgon[99]	Interactive modeling which includes *SSEHunter* (GUI)	gorgon.wustl.edu
	He's	*SSELearner*[100] (α-helix and β-sheet) *SSEtracer*[101] (α-helix, β-sheet, and backbone) *StrandTwister* (β-sheet)	www.cs.odu.edu/dsi/software.html
	IMP[102,103]	Multi-scale integrative modeling platform	salilab.org/imp
	Ma's	*Sheetminer*[104] (β-sheet)*Sheettracer*[105] (β-strand)	
	PyRy3D[106]	MC-based integrative modeling platform	genesilico.pl/pyry3d
	SITUS	Genetic algorithm based α-helix detection (*voltrac*)[107]	situs.biomachina.org
High resolution	COOT[88]	Macromolecular model building, completion, and validation, including rigid-body fitting (GUI)	www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot
	Phenix[90]	Macromolecular model building, completion, and validation, including ligand fitting (GUI)	www.phenix-online.org
	REFMAC[89]	Refinement of macromolecular structures using maximum likelihood and Bayesian statistics	www2.mrc-lmb.cam.ac.uk/groups/murshudov

CC, cross-correlation; GUI, graphical user interface; MD, molecular dynamics; MC, Monte Carlo; SA, simulated annealing.

restraints provided by the EM data. Kovacs et al.[121] generated an ensemble of α-helices with alternative orientations that are compatible with density rods of the EM map. In an alternative method, EM-Fold[95] directly assembles predicted SSEs into the target electron densities using a Metropolis Monte Carlo algorithm with a knowledge-based energy function. By bending, translating, and resizing the predicted

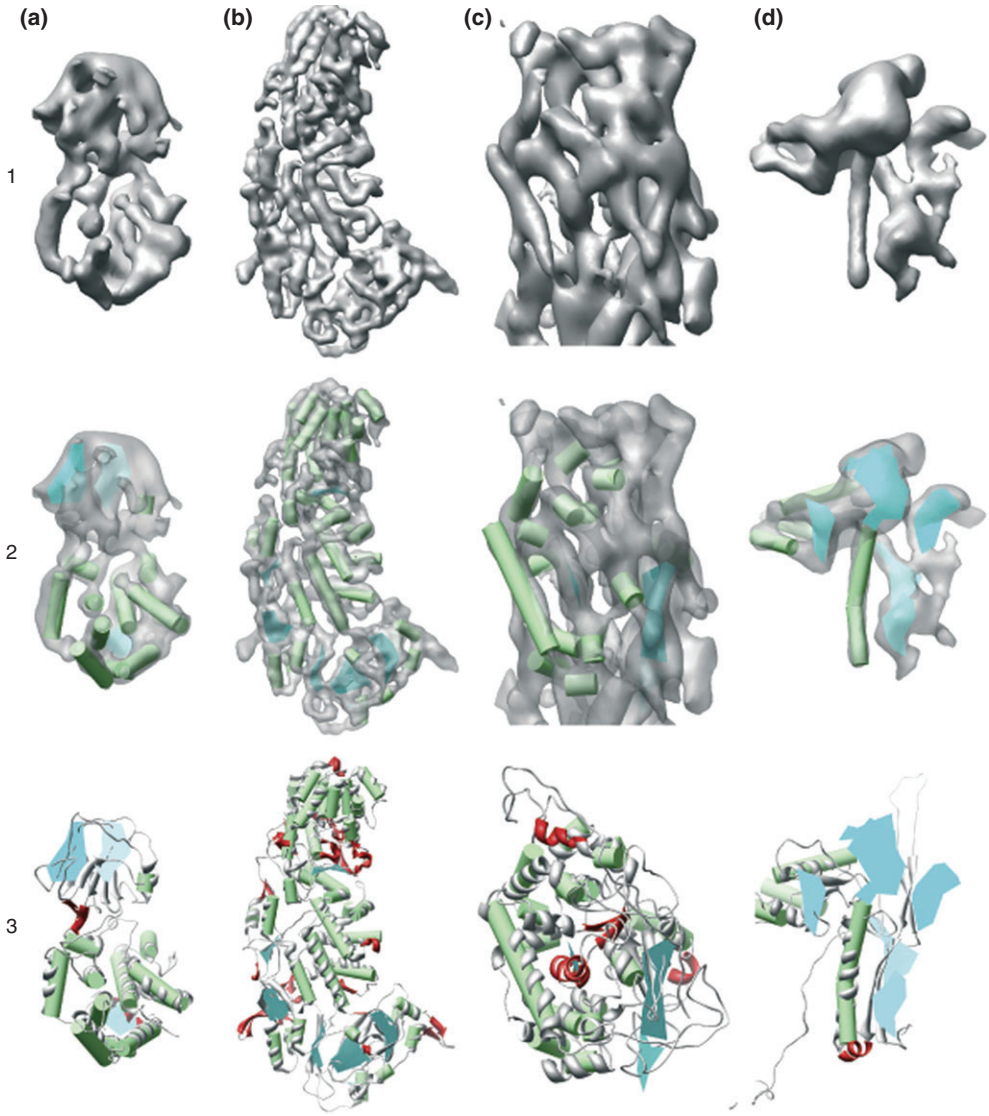

FIGURE 8 | Identification of secondary structure elements in experimental cryo-electron microscopy (EM) density maps. The experimental maps of several viral proteins are shown in the first row: the 6.8 Å resolution capsid proteins P8 and P3 from rice dwarf virus (EMDB ID: 1060) in columns (a) and (b), respectively; the upper domain of a hexon subunit from the 8 Å resolution Herpes simplex virus 1 (HSV-1) cryo-EM density map in column (c); and a gp5 monomer from the 9.5 Å resolution structure of the P22 phage (EMDB ID: 1101) in column (d). The α-helices (green cylinders) and β-sheets (cyan polygons) extracted using SSEhunter are illustrated in the middle row. In the last row, the corresponding X-ray structures (PDB IDs: 1UF2, 1NO7, and 1OHG) are superposed on the SSEhunter results with the discrepancies in identification highlighted in red. Reproduced with permission from Ref 85.

SSEs, the program builds physically realistic models that are compatible with the densities of the map.

INTEGRATIVE APPROACHES

In particular cases, the unambiguous elucidation of the fitting conformation, the identification of SSEs, or the characterization of the topology is not possible unless additional information from complementary sources provides the necessary constraints. These structural constraints can be extracted from biophysical techniques complementary to EM such as small-angle X-ray scattering (SAXS), fluorescence resonance energy transfer (FRET), NMR spectroscopy, and double electron-electron resonance (DEER), among others, but any structural experimental (e.g., mutagenesis, proteomics data) or theoretical (e.g., structure prediction) data can be used as well.[9,122] To facilitate the integration of these data from disparate sources, an integrative modeling platform (IMP) has been developed by Russel et al.[102,103] In this approach, the models can be encoded at

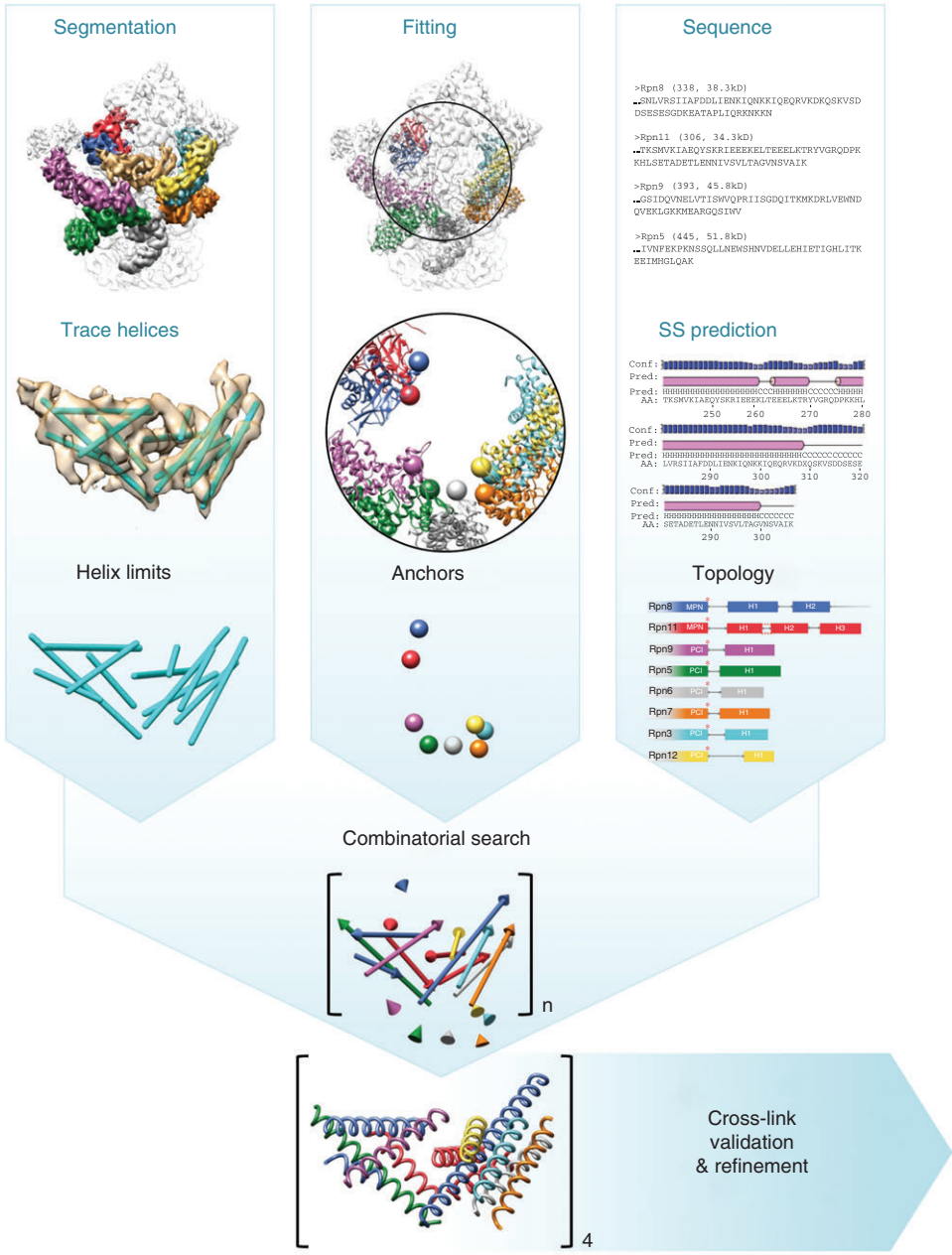

FIGURE 9 | Modeling workflow for the integrative combinatorial approach to elucidate the topology of the proteasome helical bundle. The helical bundle region (in gold) was segmented from a 7-Å resolution proteasome map (gray transparency; EMDB ID: 2165[131]), with surrounding subunits individually colored. Helix limits (cyan sticks) were automatically traced using the VolTrac tool[107] from the segmented map (gold transparency). The anchor points for the C-terminal helix linkers were extracted from the available fitted models of the neighboring subunits (PDB ID: 4B4T[131]) and are represented by colored spheres. The secondary structure prediction of the relevant subunits was obtained from their sequences using the PSIPRED server.[132] Colored arrows represent one example of the helical bundle configurations evaluated during the combinatorial search. Note that the anchor points are now shown as cones pointing toward the connected helix. Resultant models for the helical bundle with individual helices colored according to the respective lid subunits. These models were further validated with existing crosslinking data and finally refined by flexible fitting. Reproduced with permission from Ref 80.

different coarse-graining levels in consonance with the available data. Modeling is considered an assembly process in which the known restraints must be fulfilled. Several conformational sampling strategies are considered to build different candidate models. Once models are generated, they are evaluated by a potential energy that integrates the restraints derived from the EM density maps and any other experimental

source. Finally, IMP provides a variety of tools for the comparison, clustering, and analysis of such models. IMP has been applied to the nuclear pore complex,[123] ribosomes,[124] the 26S proteasome,[55] the Hsp90 and TRiC/CCT chaperonins,[125,126] the actin complex,[127] and chromatin.[128]

The 26S proteasome constitutes an excellent example to illustrate the benefits of integrative approaches. This extremely large macromolecular complex of approximately 2.5 MDa is the last effector of the ubiquitin–proteasome protein degradation pathway and is composed of the 20S core particle (CP) and the 19S regulatory particle (RP). The RP consists of 6 AAA-ATPases and at least 13 non-ATPase subunits. An atomic model of the AAA-ATPase–CP subcomplex was derived from a 20-Å resolution cryo-EM map by integrating experimental information from protein–protein interactions and homologous structures.[129] This model was improved using IMP to obtain the complete molecular architecture of the 26S holocomplex merging 7 Å resolution data from cryo-EM, X-ray crystallographic structures, chemical crosslinking, and other proteomic data.[55] The results obtained in these studies fully agreed with the most recent atomic models.[130,131]

We obtained an *ab initio* model of the key helical bundle of the RP that governs the ordered self-assembly process of the entire proteasome using the novel hybrid approach illustrated in Figure 9.[111] The helical bundle is composed of 12 helical densities that belong to the C-terminal tails (whose atomic structures have not been determined) of the eight proteins that form the so-called proteasome lid. To model the helices and establish the relationship between the helices and lid proteins, we developed a hybrid strategy to match recursively a set of geometric constraints (GCs) with a set of topology constraints (TCs). The GCs are provided by the locations of the helices detected in the map and the position of the C-terminal anchors of the surrounding lid proteins that served as starting points for modeling. The TCs include the lengths of individual helices and linkers estimated from the secondary structure predictions. The extremely large number of possible configurations of the 12 helices inside of the bundle densities ($\sim 2 \times 10^{12}$) was dramatically reduced to a handful of configurations by these constraints. Moreover, a single configuration accounted for all obtained crosslinking data and biochemical results. The final model of the helical bundle allowed us to demonstrate how this structural framework serves as a hub for lid assembly prior to its incorporation into the proteasome.

CONCLUSION

Modeling 3D-EM reconstructions with computational tools currently enables the interpretation at near-atomic resolution of different functional states of macromolecules, thereby deciphering the functional mechanism of biologically relevant complexes. In summary, in Table 1, we provide links to the reviewed available modeling tools with a short description. Although considerable progress has been achieved in using these modeling tools, there is still room for improvement. Major limiting factors include the lack of standards for assessing the quality of the EM maps and the lack of criteria for validating model accuracy. Given the current improvements in the 3D EM field, such as detectors, specimen preparation, data automation, and new software, we are confident that the number of 3D EM maps available in the EMDB, as well as the corresponding model structures, will considerably increase in the near future. This progress will help the cryo-EM community to develop tools for the critical assessment of maps and models at all resolution ranges.[14] While the community raises the quality standards to rival X-ray crystallography, carefully validated models extracted from EM maps will positively contribute to current structural biology research. In a few years' time, thanks for the joint effort of microscopists and modeling experts we will witness the explosive growth in the size and complexity of cryo-microscopy-based imaging data that will reveal the inner mechanisms of many key cellular macromolecular machines.

REFERENCES

1. Förster F, Villa E, Thomas D, Korinek A, Baumeister W. Structure determination of macromolecular complexes by cryo-electron microscopy in vitro and in situ. In: Dyson HJ, ed. *Biophysical Techniques for Structural Characterization of Macromolecules (Comprehensive Biophysics)*, vol. 1. Amsterdam: Elsevier Inc.; 2012, 245–276.
2. Frank J. Single-particle reconstruction of biological macromolecules in electron microscopy – 30 years. *Q Rev Biophys* 2009, 42:139–158.

3. Wang L, Sigworth FJ. Cryo-EM and single particles. *Physiology* 2006, 21:13–18.
4. Bammes BE, Rochat RH, Jakana J, Chen DH, Chiu W. Direct electron detection yields cryo-EM reconstructions at resolutions beyond 3/4 Nyquist frequency. *J Struct Biol* 2012, 177:589–601.
5. Milazzo AC, Cheng A, Moeller A, Lyumkis D, Jacovetty E, Polukas J, Ellisman MH, Xuong NH, Carragher B, Potter CS. Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. *J Struct Biol* 2011, 176:404–408.
6. Bai XC, Fernandez IS, McMullan G, Scheres SHW. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* 2013, 2:e00461.
7. Lawson CL, Baker ML, Best C, Bi C, Dougherty M, Feng P, Van Ginkel G, Devkota B, Lagerstedt I, Ludtke SJ, et al. EMDataBank.org: unified data resource for CryoEM. *Nucleic Acids Res* 2011, 39:D456–D464.
8. Volkmann N. Putting structure into context: fitting of atomic models into electron microscopic and electron tomographic reconstructions. *Curr Opin Cell Biol* 2012, 24:141–147.
9. Lander GC, Saibil HR, Nogales E. Go hybrid: EM, crystallography, and beyond. *Curr Opin Struct Biol* 2012, 22:627–635.
10. Esquivel-Rodríguez J, Kihara D. Computational methods for constructing protein structure models from 3D electron microscopy maps. *J Struct Biol* 2013, 184:93–102.
11. Henderson R. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. *Proc Natl Acad Sci U S A* 2013, 110:18037–18041.
12. Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* 2014, 11:63–65.
13. Cardone G, Heymann JB, Steven AC. One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J Struct Biol* 2013, 184:226–236.
14. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N, et al. Outcome of the first electron microscopy validation task force meeting. *Structure* 2012, 20:205–214.
15. Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature* 2013, 504:107–112.
16. Allegretti M, Mills DJ, McMullan G, Kühlbrandt W, Vonck J. Atomic model of the F420-reducing [NiFe] hydrogenase by electron cryo-microscopy using a direct electron detector. *eLife* 2014, 3:e01963.
17. Amunts A, Brown A, Bai XC, Llácer JL, Hussain T, Emsley P, Long F, Murshudov G, Scheres SHW, Ramakrishnan V. Structure of the yeast mitochondrial large ribosomal subunit. *Science* 2014, 343:1485–1489.
18. Fabiola F, Chapman MS. Fitting of high-resolution structures into electron microscopy reconstruction images. *Structure* 2005, 13:389–400.
19. Wriggers W, Chacón P. Modeling tricks and fitting techniques for multiresolution structures. *Structure* 2001, 9:779–788.
20. Chacón P, Wriggers W. Multi-resolution contour-based fitting of macromolecular structures. *J Mol Biol* 2002, 317:375–384.
21. Volkmann N, Hanein D. Quantitative fitting of atomic models into observed densities derived by electron microscopy. *J Struct Biol* 1999, 125:176–184.
22. Rossmann MG, Bernal R, Pletnev SV. Combining electron microscopic with X-ray crystallographic structures. *J Struct Biol* 2001, 136:190–200.
23. Woetzel N, Lindert S, Stewart PL, Meiler J. BCL::EM-Fit: rigid body fitting of atomic structures into density maps using geometric hashing and real space refinement. *J Struct Biol* 2011, 175:264–276.
24. Topf M, Baker ML, John B, Chiu W, Sali A. Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy. *J Struct Biol* 2005, 149:191–203.
25. Roseman AM. Docking structures of domains into maps from cryo-electron microscopy using local correlation. *Acta Crystallogr D Biol Crystallogr* 2000, 56:1332–1340.
26. Navaza J, Lepault J, Rey FA, Álvarez-Rúa C, Borge J. On the fitting of model electron densities into EM reconstructions: a reciprocal-space formulation. *Acta Crystallogr D Biol Crystallogr* 2002, 58:1820–1825.
27. Garzón JI, Kovacs J, Abagyan R, Chacón P. ADP_EM: fast exhaustive multi-resolution docking for high-throughput coverage. *Bioinformatics* 2007, 23:427–433.
28. Martín-Benito J, Boskovic J, Gómez-Puertas P, Carrascosa JL, Simons CT, Lewis SA, Bartolini F, Cowan NJ, Valpuesta JM. Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. *EMBO J* 2002, 21:6377–6386.
29. Craighead JL, Chang WH, Asturias FJ. Structure of yeast RNA polymerase II in solution: implications for enzyme regulation and interaction with promoter DNA. *Structure* 2002, 10:1117–1125.
30. Hoang TV, Cavin X, Ritchie DW. gEMfitter: a highly parallel FFT-based 3D density fitting tool with GPU texture memory acceleration. *J Struct Biol* 2013, 184:348–354.
31. Wriggers W, Milligan RA, McCammon JA. Situs: a package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* 1999, 125:185–195.

32. Wriggers W, Chacón P, Kovacs JA, Tama F, Birmanns S. Topology representing neural networks reconcile biomolecular shape, structure, and dynamics. *Neurocomputing* 2004, 56:365–379.
33. Ceulemans H, Russell RB. Fast fitting of atomic structures to low-resolution electron density maps by surface overlap maximization. *J Mol Biol* 2004, 338:783–793.
34. Kawabata T. Multiple subunit fitting into a low-resolution density map of a macromolecular complex using a Gaussian mixture model. *Biophys J* 2008, 95:4643–4658.
35. Vasishtan D, Topf M. Scoring functions for cryoEM density fitting. *J Struct Biol* 2011, 174:333–343.
36. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera – a visualization system for exploratory research and analysis. *J Comput Chem* 2004, 25:1605–1612.
37. Birmanns S, Rusu M, Wriggers W. Using Sculptor and Situs for simultaneous assembly of atomic components into low-resolution shapes. *J Struct Biol* 2011, 173:428–435.
38. Lasker K, Topf M, Sali A, Wolfson HJ. Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly. *J Mol Biol* 2009, 388:180–194.
39. Rusu M, Birmanns S. Evolutionary tabu search strategies for the simultaneous registration of multiple atomic structures in cryo-EM reconstructions. *J Struct Biol* 2010, 170:164–171.
40. Zhang S, Vasishtan D, Xu M, Topf M, Alber F. A fast mathematical programming procedure for simultaneous fitting of assembly components into cryoEM density maps. *Bioinformatics* 2010, 26:i261–i268.
41. Esquivel-Rodríguez J, Kihara D. Fitting multimeric protein complexes into electron microscopy maps using 3D zernike descriptors. *J Phys Chem B* 2012, 116:6854–6861.
42. de Vries SJ, Zacharias M. ATTRACT-EM: a new method for the computational assembly of large molecular machines using cryo-EM maps. *PLoS One* 2012, 7:e49733.
43. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol* 2002, 9:646–652.
44. Lasker K, Sali A, Wolfson HJ. Determining macromolecular assembly structures by molecular docking and fitting into an electron density map. *Proteins* 2010, 78:3205–3211.
45. Miyashita O, Tama F. Normal mode analysis techniques in structural biology. *eLS* 2007. doi: 10.1002/9780470015902.a0020204.
46. Wriggers W, Birmanns S. Using Situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J Struct Biol* 2001, 133:193–202.
47. Darst SA, Opalka N, Chacon P, Polyakov A, Richter C, Zhang G, Wriggers W. Conformational flexibility of bacterial RNA polymerase. *Proc Natl Acad Sci U S A* 2002, 99:4296–4301.
48. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* 2008, 16:673–683.
49. Ratje AH, Loerke J, Mikolajka A, Brünner M, Hildebrand PW, Starosta AL, Dönhöfer A, Connell SR, Fucini P, Mielke T, et al. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* 2010, 468:713–716.
50. Becker T, Franckenberg S, Wickles S, Shoemaker CJ, Anger AM, Armache JP, Sieber H, Ungewickell C, Berninghausen O, Daberkow I, et al. Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* 2012, 482:501–506.
51. Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R. Structures of the human and Drosophila 80S ribosome. *Nature* 2013, 497:80–85.
52. Agirrezabala X, Liao HY, Schreiner E, Fu J, Ortiz-Meoz RF, Schulten K, Green R, Frank J. Structural characterization of mRNA-tRNA translocation intermediates. *Proc Natl Acad Sci U S A* 2012, 109:6094–6099.
53. Hashem Y, Des Georges A, Fu J, Buss SN, Jossinet F, Jobe A, Zhang Q, Liao HY, Grassucci RA, Bajaj C, et al. High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. *Nature* 2013, 494:385–389.
54. López-Blanco JR, Chacón P. iMODFIT: efficient and robust flexible fitting based on vibrational analysis in internal coordinates. *J Struct Biol* 2013, 184:261–270.
55. Lasker K, Förster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, Beck F, Aebersold R, Sali A, Baumeister W. Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proc Natl Acad Sci U S A* 2012, 109:1380–1387.
56. Śledź P, Unverdorben P, Beck F, Pfeifer G, Schweitzer A, Förster F, Baumeister W. Structure of the 26S proteasome with ATP-γS bound provides insights into the mechanism of nucleotide-dependent substrate translocation. *Proc Natl Acad Sci U S A* 2013, 110:7264–7269.
57. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn AM, Schulten K, Aiken C, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* 2013, 497:643–646.
58. Doronin K, Flatt JW, Di Paolo NC, Khare R, Kalyuzhniy O, Acchione M, Sumida JP, Ohto U, Shimizu T, Akashi-Takamura S, et al. Coagulation factor X activates innate immunity to human species C adenovirus. *Science* 2012, 338:795–798.

59. Bai XC, Martin TG, Scheres SHW, Dietz H. Cryo-EM structure of a 3D DNA-origami object. *Proc Natl Acad Sci U S A* 2012, 109:20012–20017.
60. Tan RKZ, Devkota B, Harvey SC. YUP.SCX: coaxing atomic models into medium resolution electron density maps. *J Struct Biol* 2008, 163:163–174.
61. Flores SC. Fast fitting to low resolution density maps: elucidating large-scale motions of the ribosome. *Nucleic Acids Res* 2014, 42:e9.
62. Vashisth H, Skiniotis G, Brooks CL. Enhanced sampling and overfitting analyses in structural refinement of nucleic acids into electron microscopy maps. *J Phys Chem B* 2013, 117:3738–3746.
63. Tama F, Miyashita O, Brooks CL III. Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis. *J Mol Biol* 2004, 337:985–999.
64. Tama F, Miyashita O, Brooks CL III. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. *J Struct Biol* 2004, 147:315–326.
65. Hinsen K, Reuter N, Navaza J, Stokes DL, Lacapère JJ. Normal mode-based fitting of atomic structure into electron density maps: application to sarcoplasmic reticulum Ca-ATPase. *Biophys J* 2005, 88:818–827.
66. López-Blanco JR, Garzón JI, Chacón P. iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics* 2011, 27:2843–2850.
67. Lau WCY, Rubinstein JL. Subnanometre-resolution structure of the intact *Thermus thermophilus* H+-driven ATP synthase. *Nature* 2012, 481:214–219.
68. Oot RA, Huang LS, Berry EA, Wilkens S. Crystal structure of the yeast vacuolar ATPase heterotrimeric EGC head peripheral stalk complex. *Structure* 2012, 20:1881–1892.
69. Shakeel S, Seitsonen JJT, Kajander T, Pasi L, Hyypi T, Susi P, Butcher SJ. Structural and functional analysis of coxsackievirus a9 integrin $\alpha v\beta 6$ binding and uncoating. *J Virol* 2013, 87:3943–3951.
70. Jolley CC, Wells S, Fromme P, Thorpe MF. Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations. *Biophys J* 2008, 94:1613–1621.
71. Topf M, Lasker K, Webb B, Wolfson H, Chiu W, Sali A. Protein structure fitting and refinement guided by cryo-EM density. *Structure* 2008, 16:295–307.
72. Schröder GF, Brunger AT, Levitt M. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure* 2007, 15:1630–1641.
73. De Groot BL, Van Aalten DMF, Scheek RM, Amadei A, Vriend G, Berendsen HJC. Prediction of protein conformational freedom from distance constraints. *Proteins* 1997, 29:240–251.
74. Zheng W. Accurate flexible fitting of high-resolution protein structures into cryo-electron microscopy maps using coarse-grained pseudo-energy minimization. *Biophys J* 2011, 100:478–488.
75. Velazquez-Muriel JA, Valle M, Santamaría-Pang A, Kakadiaris IA, Carazo JM. Flexible fitting in 3D-EM guided by the structural variability of protein superfamilies. *Structure* 2006, 14:1115–1126.
76. Zhang J, Minary P, Levitt M. Multiscale natural moves refine macromolecules using single-particle electron microscopy projection images. *Proc Natl Acad Sci U S A* 2012, 109:9845–9850.
77. Ahmed A, Tama F. Consensus among multiple approaches as a reliability measure for flexible fitting into cryo-EM data. *J Struct Biol* 2013, 182:67–77.
78. Ginalski K. Comparative modeling for protein structure prediction. *Curr Opin Struct Biol* 2006, 16:172–177.
79. Gront D, Kmiecik S, Blaszczyk M, Ekonomiuk D, Koliński A. Optimization of protein models. *Wiley Interdiscip Rev Comput Mol Sci* 2012, 2:479–493.
80. Topf M, Baker ML, Marti-Renom MA, Chiu W, Sali A. Refinement of protein structures by iterative comparative modeling and cryoEM density fitting. *J Mol Biol* 2006, 357:1655–1668.
81. DiMaio F, Tyka MD, Baker ML, Chiu W, Baker D. Refinement of protein structures into low-resolution density maps using Rosetta. *J Mol Biol* 2009, 392:181–190.
82. Zhu J, Cheng L, Fang Q, Zhou ZH, Honig B. Building and refining protein models within cryo-electron microscopy density maps based on homology modeling and multiscale structure refinement. *J Mol Biol* 2010, 397:835–851.
83. Velázquez-Muriel JA, Sorzano COS, Scheres SHW, Carazo JM. SPI-EM: towards a tool for predicting CATH superfamilies in 3D-EM maps. *J Mol Biol* 2005, 345:759–771.
84. Saha M, Morais MC. FOLD-EM: automated fold recognition in medium- and low-resolution (4-15 Å) electron density maps. *Bioinformatics* 2012, 28:3265–3273.
85. Lasker K, Dror O, Nussinov R, Wolfson H. Discovery of protein substructures in EM maps. In: Casadio R, Myers G, eds. *5th International Workshop on Algorithms in Bioinformatics, WABI 2005 (Lecture Notes in Computer Science)*. 2005, 423–434.
86. Dror O, Lasker K, Nussinov R, Wolfson H. EMatch: an efficient method for aligning atomic resolution subunits into intermediate-resolution cryo-EM maps of large macromolecular assemblies. *Acta Crystallogr D Biol Crystallogr* 2006, 63:42–49.
87. Lasker K, Dror O, Shatsky M, Nussinov R, Wolfson HJ. EMatch: discovery of high resolution structural

homologues of protein domains in intermediate resolution Cryo-EM maps. *IEEE/ACM Trans Comput Biol Bioinform* 2007, 4:28–39.

88. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010, 66:486–501.
89. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 2011, 67:355–367.
90. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Echols N, Headd JJ, Hung LW, Jain S, Kapral GJ, Grosse Kunstleve RW, et al. The Phenix software for automated determination of macromolecular structures. *Methods* 2011, 55:94–106.
91. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 2007, Chapter 2: Unit 2.9. doi: 10.1002/0471140864.ps0209s50.
92. Wriggers W. Conventions and workflows for using Situs. *Acta Crystallogr D Biol Crystallogr* 2012, 68:344–351.
93. Siebert X, Navaza J. UROX 2.0: an interactive tool for fitting atomic models into electron-microscopy reconstructions. *Acta Crystallogr D Biol Crystallogr* 2009, 65:651–658.
94. Suhre K, Navaza J, Sanejouand YH. NORMA: A tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps. *Acta Crystallogr D Biol Crystallogr* 2006, 62:1098–1100.
95. Lindert S, Alexander N, Wötzel N, Karakaş M, Stewart PL, Meiler J. EM-Fold: de novo atomic-detail protein structure determination from medium-resolution density maps. *Structure* 2012, 20:464–478.
96. Jiang W, Baker ML, Ludtke SJ, Chiu W. Bridging the information gap: computational tools for intermediate resolution structure interpretation. *J Mol Biol* 2001, 308:1033–1044.
97. Baker ML, Ju T, Chiu W. Identification of secondary structure elements in intermediate-resolution density maps. *Structure* 2007, 15:7–19.
98. Ju T, Baker ML, Chiu W. Computing a family of skeletons of volumetric models for shape description. *Comput Aided Des* 2007, 39:352–360.
99. Baker ML, Baker MR, Hryc CF, Ju T, Chiu W. Gorgon and pathwalking: macromolecular modeling tools for subnanometer resolution density maps. *Biopolymers* 2012, 97:655–668.
100. Si D, Ji S, Nasr KA, He J. A machine learning approach for the identification of protein secondary structure elements from electron cryo-microscopy density maps. *Biopolymers* 2012, 97:698–708.
101. Si D, He J: Beta-sheet detection and representation from medium resolution Cryo-EM density maps. In *Proceedings of the 4th ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, Washington, DC, 22–25 September 2013, 764–770.
102. Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* 2012, 10:e1001244.
103. Velázquez-Muriel J, Lasker K, Russel D, Phillips J, Webb BM, Schneidman-Duhovny D, Sali A. Assembly of macromolecular complexes by satisfaction of spatial restraints from electron microscopy images. *Proc Natl Acad Sci U S A* 2012, 109:18821–18826.
104. Kong Y, Ma J. A structural-informatics approach for mining β-sheets: locating sheets in intermediate-resolution density maps. *J Mol Biol* 2003, 332:399–413.
105. Kong Y, Zhang X, Baker TS, Ma J. A structural-informatics approach for tracing β-sheets: building pseudo-Cα traces for β-strands in intermediate-resolution density maps. *J Mol Biol* 2004, 339:117–130.
106. Tuszynska I, Matelska D, Magnus M, Chojnowski G, Kasprzak JM, Kozlowski LP, Dunin-Horkawicz S, Bujnicki JM. Computational modeling of protein-RNA complex structures. *Methods* 2014, 65:310–319.
107. Rusu M, Wriggers W. Evolutionary bidirectional expansion for the tracing of alpha helices in cryo-electron microscopy reconstructions. *J Struct Biol* 2012, 177:410–419.
108. Baker ML, Zhang J, Ludtke SJ, Chiu W. Cryo-EM of macromolecular assemblies at near-atomic resolution. *Nat Protoc* 2010, 5:1697–1708.
109. Dal Palù A, He J, Pontelli E, Lu Y. Identification of alpha-helices from low resolution protein density maps. *Comput Syst Bioinformatics Conf* 2006, 89-98.
110. Zhu L, Zhong X, Chen SRW, Banavali N, Liu Z. Modeling a ryanodine receptor N-terminal domain connecting the central vestibule and the corner clamp region. *J Biol Chem* 2013, 288:903–914.
111. Estrin E, López-Blanco JR, Chacón P, Martin A. Formation of an intricate helical bundle dictates the assembly of the 26S proteasome Lid. *Structure* 2013, 21:1624–1635.
112. Yu Z, Bajaj C. Computational approaches for automatic structural analysis of large biomolecular complexes. *IEEE/ACM Trans Comput Biol Bioinform* 2008, 5:568–582.
113. Ludtke SJ, Baker ML, Chen DH, Song JL, Chuang DT, Chiu W. De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure* 2008, 16:441–448.

114. Jiang W, Baker ML, Jakana J, Weigele PR, King J, Chiu W. Backbone structure of the infectious ϵ15 virus capsid revealed by electron cryomicroscopy. *Nature* 2008, 451:1130–1134.

115. Wu Y, Chen M, Lu M, Wang Q, Ma J. Determining protein topology from skeletons of secondary structures. *J Mol Biol* 2005, 350:571–586.

116. Al Nasr K, Chen L, Si D, Ranjan D, Zubair M, He J. Building the initial chain of the proteins through de novo modeling of the cryo-electron microscopy volume data at the medium resolutions. In: *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB '12)*, Orlando, Florida, 2012, 490–497.

117. Biswas A, Si D, Al Nasr K, Ranjan D, Zubair M, He J. Improved efficiency in Cryo-Em secondary structure topology determination from inaccurate data. *J Bioinform Comput Biol* 2012, 10:1242006.

118. Enosh A, Fleishman SJ, Ben-Tal N, Halperin D. Assigning transmembrane segments to helices in intermediate-resolution structures. *Bioinformatics* 2004, 20:i122–i129.

119. Fleishman SJ, Harrington S, Friesner RA, Honig B, Ben-Tal N. An automatic method for predicting transmembrane protein structures using cryo-EM and evolutionary data. *Biophys J* 2004, 87:3448–3459.

120. Beuming T, Weinstein H. Modeling membrane proteins based on low-resolution electron microscopy maps: a template for the TM domains of the oxalate transporter OxIT. *Protein Eng Des Sel* 2005, 18:119–125.

121. Kovacs JA, Yeager M, Abagyan R. Computational prediction of atomic structures of helical membrane proteins aided by EM maps. *Biophys J* 2007, 93:1950–1959.

122. Förster F, Villa E. Integration of cryo-EM with atomic and protein–protein interaction data. *Methods Enzymol* 2010, 483:47–72.

123. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT, et al. Determining the architectures of macromolecular assemblies. *Nature* 2007, 450:683–694.

124. Taylor DJ, Devkota B, Huang AD, Topf M, Narayanan E, Sali A, Harvey SC, Frank J. Comprehensive molecular structure of the eukaryotic ribosome. *Structure* 2009, 17:1591–1604.

125. Krukenberg KA, Förster F, Rice LM, Sali A, Agard DA. Multiple conformations of *E. coli* Hsp90 in solution: insights into the conformational dynamics of Hsp90. *Structure* 2008, 16:755–765.

126. Booth CR, Meyer AS, Cong Y, Topf M, Sali A, Ludtke SJ, Chiu W, Frydman J. Mechanism of lid closure in the eukaryotic chaperonin TRiC/CCT. *Nat Struct Mol Biol* 2008, 15:746–753.

127. Cong Y, Topf M, Sali A, Matsudaira P, Dougherty M, Chiu W, Schmid MF. Crystallographic conformers of actin in a biologically active bundle of filaments. *J Mol Biol* 2008, 375:331–336.

128. Baù D, Sanyal A, Lajoie BR, Capriotti E, Byron M. The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat Struct Biol* 2010, 18:107–114.

129. Förster F, Lasker K, Beck F, Nickell S, Sali A, Baumeister W. An atomic model AAA-ATPase/20S core particle sub-complex of the 26S proteasome. *Biochem Biophys Res Commun* 2009, 388:228–233.

130. Unverdorben P, Beck F, Śledź P, Schweitzer A, Pfeifer G, Plitzko JM, Baumeister W, Förster F. Deep classification of a large cryo-EM dataset defines the conformational landscape of the 26S proteasome. *Proc Natl Acad Sci U S A* 2014, 111:5544–5549.

131. Beck F, Unverdorben P, Bohn S, Schweitzer A, Pfeifer G, Sakata E, Nickell S, Plitzko JM, Villa E, Baumeister W, et al. Near-atomic resolution structural model of the yeast 26S proteasome. *Proc Natl Acad Sci U S A* 2012, 109:14870–14875.

132. Buchan DWA, Ward SM, Lobley AE, Nugent TCO, Bryson K, Jones DT. Protein annotation and modelling servers at University College London. *Nucleic Acids Res* 2010, 38:W563–W568.

FURTHER READING

Frank J. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in their Native State*. 2nd ed. New York: Oxford University Press; 2006.

