HHS Public Access

Author manuscript

Annu Rev Biophys. Author manuscript; available in PMC 2017 July 25.

Published in final edited form as:

Annu Rev Biophys. 2016 July 05; 45: 253–278. doi:10.1146/annurev-biophys-062215-011113.

COMPUTATIONAL METHODOLOGIES for REAL-SPACE STRUCTURAL REFINEMENT of LARGE MACROMOLECULAR COMPLEXES

Boon Chong Goh[1,2,*], **Jodi A. Hadden**[1,3,*], **Rafael C. Bernardi**[1,3], **Abhishek Singharoy**[1], **Ryan McGreevy**[1], **Till Rudack**[1], **C. Keith Cassidy**[1,2], and **Klaus Schulten**[1,2,3,4,5]

[1]Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

[2]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

[3]Energy Biosciences Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

[4]Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

[5]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801

Abstract

The rise of the computer as a powerful tool for model building and refinement has revolutionized the field of structure determination for large biomolecular systems. Despite the wide availability of robust experimental methods capable of resolving structural details across a range of spatiotemporal resolutions, computational hybrid methods have the unique ability to integrate the diverse data from multimodal techniques such as X-ray crystallography and electron microscopy into consistent, fully atomistic structures. Here, commonly employed strategies for computational real-space structural refinement are reviewed, and their specific applications are illustrated for several large macromolecular complexes: ribosome, virus capsids, chemosensory array, and photosynthetic chromatophore. The increasingly important role of computational methods in large-scale structural refinement, along with current and future challenges, is discussed.

Keywords

integrative modeling; hybrid methods; molecular dynamics; flexible fitting; simulation; cryo-EM

INTRODUCTION

An array of diverse yet complementary methodologies represent the state-of-the-art repertoire for experimental structure determination of large-scale protein complexes and

*Co-first author

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

macromolecular assemblies. The most widely employed of these include cryo-electron microscopy (cryo-EM), X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, small-angle X-ray scattering (SAXS), and electron paramagnetic resonance spectroscopy with site-directed spin labeling (EPR-SDSL). Although each technique provides a powerful means for characterizing critical aspects of molecular and cellular architecture, the potential for a single experimental method to produce a comprehensive, all-atom description of a large biomolecular system in its native state remains limited.

X-ray crystallography provides biomolecular structures at atomic-level detail; however, owing to sample preparation conditions and artifacts of crystallization, the results do not necessarily capture physiologically relevant conformations of biomolecules. NMR spectroscopy can characterize atom-by-atom spatial and conformational relationships toward solution of a molecular structure, as well as provide information regarding dynamics, but NMR data signals become progressively difficult to detect and deconvolute with increasing system size. Unlike X-ray crystallography and NMR spectroscopy, cryo-EM can be applied to very large biomolecular systems, yielding detailed density maps that describe architectural features of native-state conformations. Although cryo-EM represents a rapidly advancing field that has recently produced density maps at near-atomic resolution (12, 49), the method is hampered considerably by inherent molecular flexibility, which translates to regions of low resolution and loss of structural detail. SAXS and EPR-SDSL are similarly applicable to systems of notable size and are capable of elucidating structural information even in the presence of flexibility; however, their resolution limits remain still too low to assign the positions of individual atoms.

Atomistic structure determination, which originally began with the construction of physical models over 50 years ago, has evolved dramatically with the rise of the computer as a powerful visualization and model-building tool (Figure 1). Since their introduction, computer-based refinement methods have enabled structural biologists to solve the structures of significantly larger and more complex biomolecules. Today, computational hybrid methods offer an effective strategy for addressing the limitations of experimental structure determination, facilitating the integration of complementary data from multiple sources to produce unified, refined, and fully atomistic structural models. Here, we review a collection of commonly employed approaches for computational structural refinement, along with specific applications of these hybrid methods to a selection of noteworthy biomolecular systems. Additionally, current and future challenges, as well as the anticipated direction of the field, are discussed, particularly with respect to the adaptation of computational techniques to support emerging experimental methods capable of increasingly fine structural resolution.

COMPUTATIONAL STRUCTURAL REFINEMENT METHODS

A major challenge for the study of large multiprotein complexes is obtaining and refining structural data that comprehensively describe the full system at atomistic detail. Hybrid structural biology approaches employing computational techniques to integrate data from diverse experiments provide a means to deliver all-atom structures.

The general strategy for applying hybrid methods, outlined in Figure 2, often involves fitting the structures of isolated subunits or modular components solved, for example, by X-ray crystallography or NMR spectroscopy to data obtained by cryo-EM to reassemble an entire multiprotein complex. Any missing structural features are modeled using in silico structure prediction tools. The complete model is further refined via conformational sampling while enforcing simulation restraints based on experimental data, such as those from NMR spectroscopy and EPR-SDSL.

The present review covers hybrid methods for large-scale structural refinement based on fitting to cryo-EM density maps, enhancing low-resolution X-ray and NMR data, applying experimentally based simulation restraints, employing in silico prediction routines to fill in unresolved structural features, and performing enhanced conformational sampling. Real-space refinement techniques involve a set of rigorous error-assessment criteria; however, a detailed discussion on error analysis is beyond the scope of this review, which seeks instead to overview the scientific prowess of hybrid methods. Error analysis of real-space refinement techniques is thoroughly reviewed elsewhere (111, 116).

Structural Refinement Based on Cryo-EM Data

Although the most widely used method for acquiring biomolecular structures is X-ray crystallography, crystallization of very large biomolecules, macromolecular complexes, and membrane proteins can be extremely challenging. In response, cryo-EM, which does not require the difficult crystallization step and allows imaging under physiologically relevant conditions, is increasingly becoming the central approach for structure determination of large systems. Notably, the two experimental techniques often reveal different levels of macromolecular architecture. X-ray crystallography generally produces atomic-resolution (<3 Å) structures, whereas many cryo-EM densities are resolved only at lower resolution. However, recent developments in cryo-EM detector technology have yielded high-resolution density maps (~2–5 Å) (11, 12, 20, 49, 80), though atomistic models derived purely from EM data are still very rare. Computational methods that combine data from both X-ray crystallography and cryo-EM effectively bridge the resolution gap between the two complementary techniques and afford physiologically accurate, atomic-resolution structures of biomolecular complexes.

Many such methods combining X-ray crystallography and cryo-EM data for structure determination have been developed in recent years. Some of these methods use rigid-fragment fitting (30, 108, 147), and others such as DireX (117), Flex-EM (133), Rosetta (40), and FRODA (64) perform flexible fitting, allowing conformational changes to better shape the structure to the data. Popular approaches employ low-frequency normal modes (132), deformable elastic networks (117), and cross-correlation (98) or least-squares difference between experimental and simulated maps (31) to drive the structure into the cryo-EM density. Fitting methods may be based on a Monte Carlo scheme (40) or on molecular dynamics (MD) simulation, such as molecular dynamics flexible fitting (MDFF) (135, 136).

The MDFF protocol requires, as a prerequisite, a complete atomistic model of reasonable initial quality that can be improved upon by attempting to match it to the EM data. The

starting model, which may be based on, for example, X-ray crystallography and in silico structure prediction, is subject to an MD simulation in NAMD (104) driven by a modified potential energy function that includes a term incorporating the cryo-EM density. Forces are computed from the added potential and applied to each atom, driving them into high-density regions and thus producing an atomic-resolution structure in the conformation captured by the cryo-EM map. Restraints imposed during the simulation help preserve the secondary structure, stereochemical correctness (115), and symmetry (28) of the proteins and nucleic acid components involved. The MD-based nature of MDFF allows for flexibility and sampling while maintaining a realistic structural geometry through application of the most advanced force fields. Both AMBER and CHARMM now also offer functionality to perform flexible fitting to cryo-EM density maps (98, 149). A similar feature is available in GROMACS through collaboration with MDfit (145).

An important step following any hybrid fitting method is the evaluation of the final model. Three types of evaluation are typically pursued. First, the quality of models produced by structure-building methods is commonly assessed using MolProbity (32), which is based on various statistical metrics such as Ramachandran analysis, rotamer outliers, and steric clashes. In the second type of evaluation, the quality of the resulting model is determined by its fit to the cryo-EM density. Although one of the most common scoring methods for such fit is the cross-correlation coefficient (CCC) between experimental and simulated density maps, the global CCC analysis is prone to producing false-positive information and the result is inherently degenerate. That is, two completely different structures fitted into a density map can produce the same CCC, and even though the CCC is close to 1 (perfect fit), the fitted structure could be severely distorted. Instead, it is more informative to calculate the correlation at a finer decomposition, for example, per residue, which specifically reveals the parts of a structure that fit well versus the parts that require additional attention. Fast parallel algorithms make this assessment computationally feasible, even for large structures and long-fitting trajectories (130). In the third evaluation, Fourier shell correlations (FSCs), used to calculate the resolution of a cryo-EM density, are employed to quantify the model-to-map fit. When applying the gold standard method of FSC, which utilizes two independent half-maps (59), a cross-validation protocol can be followed during the process of model refinement (42, 112) to evaluate problems due to overfitting. The protocol involves fitting the model to one half-map while calculating the real-space or FSC with respect to the other, similar to the R_{free} concept in crystallography (21).

As previously noted, recent advances in electron detectors and imaging software have improved the resolution obtainable by cryo-EM (77, 95), leading to the solution of a number of structures from near atomic-resolution (~2–5 Å) cryo-EM densities (11, 12, 20, 49, 80). Such resolution permits de novo model building by assigning sequence into the density, creating accurate structures without the use of homologous proteins (73, 143). Even though high-resolution cryo-EM data are becoming more readily obtainable, resolution is not always uniform throughout a map. Flexible regions of the structure may still produce local resolutions lower than that of the overall map, as evaluated with tools such as ResMap (72). Modelers can adjust refinement protocols to account for such local variations and better inform the process of model validation. Local resolution analysis can be especially

important for determining the parts of a high-resolution map that realistically contain side chain information and the parts that do not, preventing overinterpretation of the latter.

INTERACTIVE MOLECULAR DYNAMICS FLEXIBLE FITTING

MDFF can be run interactively to incorporate user expertise into the fitting process, especially for complicated, ambiguous structural regions where automated procedures may fail. This interactive feature allows a user to manipulate the target structure during the MDFF calculation by manually pulling it to the desired regions of density (129, 134). A new parallel implementation of cross-correlation analysis in visual molecular dynamics (VMD) (61, 130) provides real-time quality-of-fit estimates during interactive simulations. Together, these features enable improved, efficient fitting of structures to density maps that incorporate user intuition.

Structural Refinement Based on X-Ray Crystallography Data

X-ray crystallography is arguably the most versatile and dominant technique for delivering atomistic structures of biomolecules. An increasing number of structures are submitted each year to the Protein Data Bank (PDB) (e.g., 748 in 1995, compared with 8,895 in 2014), with approximately 90% of the current total entries coming from X-ray crystallography (93,667 by the end of 2014). This increase is partially attributable to the advancement of automated computational methods used in analysis of diffraction data and structural refinement (2, 3). Furthermore, next-generation synchrotron radiation sources, such as X-ray free-electron lasers (XFELs), are enhancing significantly the capabilities of X-ray crystallography (57). For example, XFELs can be used for collecting diffraction data from microcrystals, which for many protein complexes are more easily obtainable than a single large crystal that is traditionally required. Additionally, XFELs have been shown to be capable of imaging single particles (e.g., proteins or viruses) (13) similar to cryo-EM.

Despite the recent advances in the field of X-ray crystallography, investigating the structure of large biomolecular complexes remains a significant challenge for traditional crystallography techniques. The inherent flexibility embodied in large systems, as well as the presence of disordered solvent, lipids, or ligands, often causes large-system crystals to diffract at low resolution (5–7 Å). At moderate-to-low resolution (3–7 Å), knowledge of the stereochemistry of the system must be incorporated to achieve accurate atomic positions (118).

Solving structures from low-resolution diffraction data is a difficult, time-consuming process. Low-resolution (beyond 5 Å) data sets are often simply discarded (66). However, novel methods developed to better handle low-resolution data have resulted in a rapid increase in the number of low-resolution X-ray structures presented in recent years (66). For example, DEN refinement combines deformable elastic network models with generic stereochemistry and homology information (118, 119). AMBER includes a framework to perform refinement via an interface with the Crystallography and NMR System (CNS) software suite (22). Other notable recent developments include normal mode refinement (36), the Rosetta implementation of physical energy functions (41) and its combination with

reciprocal space X-ray refinement in Phenix (39), torsional optimization protocols (58), external structural restraints or jelly body refinement in REFMAC (97), and xMDFF (92).

The xMDFF protocol (92) is based on a modified version of the MDFF approach, discussed in the previous section. Instead of using densities from cryo-EM, xMDFF works with model-phased densities, which include information taken from both a tentative model and X-ray diffraction data. The tentative model is flexibly fitted into the density, with restraints applied to preserve secondary structure, stereochemical correctness (115), and symmetry (28) of proteins and nucleic acids. The resulting fitted structure, together with the experimental diffraction data, is used to regenerate a new electron density. The fitted structure is then employed as an updated model to be driven into the new electron density and further refined. This process is repeated iteratively until a convergence test is passed (e.g., the R_{work} and R_{free} values reach a minimum or become lower than a predefined tolerance).

Historically, real-space refinement methods are expected to have a wide convergence radius, as has been formally shown for the case that initial phases are of good quality (37). MD protocols have served to further improve the convergence radius of real-space refinements (23), and xMDFF leverages this advantage toward its refinement of low-resolution diffraction data. In particular, the incorporation of electrostatics during refinement, as accounted for by the xMDFF protocol, can improve the resulting structures (48). Furthermore, MolProbity scores indicate that the MD force fields and restraints used during xMDFF refinement produce structures with good geometry (92). Application of MD force fields can be an important factor in reducing the degrees of freedom required for fitting low-resolution density and, thus, the experimental measurements needed for refinement (3).

THE PHASE PROBLEM OF X-RAY CRYSTALLOGRAPHY

The classical problem in crystallography, the phase problem, refers to the necessity of knowing phase factors corresponding to the measured diffraction spots in order to reconstruct an electron density. Normally, crystallographic diffraction patterns provide only the amplitude of the diffracted X-ray and not the phase. Some solutions to this problem rely on experimental phase determination using the measured phase shift (anomalous diffraction) from multiple (MAD) or single (SAD) wavelengths. Direct phase determination from continuous diffraction patterns obtainable by XFELs is possible through use of the oversampling method with iterative algorithms (114). Other methods obtain approximate phases calculated from known similar models, which are then subsequently refined (69), such as in molecular replacement (MR). Similar to MR (91), xMDFF (molecular dynamics flexible fitting for low-resolution X-ray crystallography) uses model-phased densities as a target for refinement (92).

Structural Refinement Based on Experimental Restraints

Just as biomolecular structures can be refined through simulations that drive their conformations to match cryo-EM or X-ray diffraction densities, experimentally derived restraints based on data from a variety of other biophysical techniques can also be used to

guide computational structure refinement. Indeed, modern protein structure determination, particularly in the field of NMR spectroscopy, relies heavily on the use of programs designed to produce structural solutions that best satisfy data-based restraints. Widely used software for enforcing experimental restraints to solve NMR structures include CNS (22) and Xplor-NIH (122). Experimental sources for restraint data may also include, among others, EPR-SDSL, SAXS, fluorescence resonance energy transfer (FRET) microscopy, and cross-linking/mass spectrometry (XL-MS). More advanced approaches to restraint-based structure determination incorporate fitting to data derived from a combination of complementary techniques. For example, the combination of restraints obtained from NMR and SAXS experiments provides an established and powerful strategy for investigating structure and dynamics in large, multidomain proteins (139). Further, scoring functions based on a variety of proteomics data can be applied as an acceptance criterion to select computationally derived structures that satisfy experimentally determined spatial restraints; such an approach was employed to determine the molecular architecture of the yeast nuclear pore complex (6, 7).

Restraints based on experimentally derived data can, likewise, be used to bias the conformational dynamics of proteins during MD simulations for the purpose of structure refinement. For example, the pair-wise distances describing the spatial relationships of atoms in macromolecular complexes, determined by NMR spectroscopy and EPR-SDSL, can be employed as structural restraints during MD simulations. These restraints, imposed as harmonic potentials, effectively force atoms to adopt experimentally known relative positions to enhance the physical validity of structural models. Alternatively, restraints based on other NMR spectroscopy measurements, including angular and torsional relationships, chemical shifts, pseudocontact shifts, and dipolar couplings, can also be used. Restraints can further be imposed to force the average over multiple independent MD simulation trajectories to match a given experimental value. For example, a recent study employed NMR data in restrained-ensemble simulations to refine the structure of a membrane protein within the complex biological environment of a lipid bilayer (33). Additionally, restrained-ensemble simulations using data from EPR double electron-electron resonance (EPR-DEER) experiments have been adapted for protein structure refinement, as demonstrated in an application to T4 lysozyme (62, 109). Although functionality to impose standard distance and torsional restraints is commonly available in MD simulation software suites, restrained-ensemble MD simulations for structural refinement have only been implemented thus far in AMBER (150) and CHARMM (109), or are otherwise available through the open-source PLUMED plug-in (88).

Structural Refinement Based on In Silico Structure Prediction

Regardless of whether structural data are obtained by cryo-EM, X-ray crystallography, or NMR spectroscopy, inherent molecular flexibility often means a biological structure cannot be fully resolved. The corresponding lack of structural segments is evident in many structures deposited in the PDB that have missing coordinates for inserted loops and chain termini, which represent the most highly disordered regions of a protein. In silico structure prediction tools can be utilized to rebuild missing structural features, providing complete initial input models for further structural refinement strategies. The available in silico

structure prediction methods can be divided into two major categories: those that draw upon the structural knowledge stored in the PDB and those that infer structural details solely from information encoded in MD force field parameters.

RESOLUTION DEPENDENCE OF STRUCTURAL PARAMETERS DERIVED FROM EXPERIMENTAL DATA

Spatial resolution of an experimental data set uniquely determines the total number of derivable structural parameters. Extraction of any number of parameters beyond those limited by the resolution of the data set would result in overfitting of the data, leading to poor structural statistics and spurious biochemical conclusions (112). It is therefore crucial to quantify the expected number of structural parameters that characterizes a given data set before invoking any real-space refinement technique.

The Matthews coefficient (90), a metric employed by crystallographers, can be used to estimate the number of structural parameters in a given experimental data set. Real-space refinement of protein side chains is recommended for data at a resolution better than 3 Å (65). Secondary structure fitting can be done reasonably well for data at a resolution up to 6 Å. However, if the connectivity information between atoms is accurately maintained, as with MD force fields, information on the protein backbone within a secondary structure can be extracted from data at a resolution up to 8 Å. Any data beyond resolution of 10 Å, in general, can only be fitted using rigid-body docking protocols.

Homology modeling, which takes direct advantage of data stored in the PDB, is the most widely employed approach for structure prediction. The application of homology modeling techniques requires, as a prerequisite, the availability of at least one protein structure of similar amino acid sequence (8). Assuming that structural homology is highly correlated with sequence homology, models are constructed according to alignment of the two sequences and mapping of the homologous template structure to fill the missing regions of the target structure. The quality of models produced is then evaluated on the basis of specific structural and energetic criteria to eliminate erroneous results (46). An alternative approach, which also uses knowledge from the PDB, involves building a local fragment library based on the target amino acid residue sequence. Implemented in Rosetta (128), this approach uses a Monte Carlo method guided by a knowledge-based scoring function to exchange and place the fragments into a partial model (68). Recently, in silico structure prediction by Rosetta was combined iteratively with MDFF to overcome conformational traps that can befall structure prediction methods (81). A detailed comparison of software options available for knowledge-based structure prediction is given by Dolan et al. (43).

The second category of in silico structure prediction methods, which draws solely on information encoded within the employed general (i.e., independent of any ad hoc assumptions) MD force field, is based on optimization and/or conformational sampling starting from an initial guess for the atomic coordinates of the missing structural regions. Sampling may be accomplished through stochastic Monte Carlo schemes or through MD simulations. Commonly used strategies for performing enhanced conformational sampling are discussed in the next section.

Regardless of the method employed to generate candidate conformations for experimentally unresolved structural segments or, indeed, for entire biomolecules, the large number of potential structural models produced must be filtered. A widely adopted strategy is model clustering, followed by ranking according to cluster size (127). Further ranking of structural models can be accomplished through the discrete optimized protein energy (DOPE) metric (125), Rosetta's knowledge-based energy scoring function (68), or structure evaluators such as MolProbity (32) and ProCheck (75). Although ranking algorithms can select reasonable structural models, they cannot be used for model validation. The accuracy of a computationally predicted structure can be notably improved by incorporating experimentally derived data into the model-building process (68). Some structure prediction programs, e.g., Rosetta, offer an option to perform structure prediction targeted to high-resolution (3.5 Å) cryo-EM data, adding an extra energy term to penalize disagreement with the density map (40, 143). Rosetta's implementation for incorporating cryo-EM data was recently applied to successfully obtain the structure of the type VI secretion system sheath (73).

Conformational Sampling for Structural Refinement

The energy landscapes that describe the conformational space of biological molecules are extremely rough, and as a result, MD simulations often become trapped in local minima, expending much of their effort exploring conformations that represent nonfunctional states (16). As conformational changes, often large ones, play an essential role in mediating the activity of many proteins (14, 16), the inability to cross energy barriers to visit other regions of conformational space constitutes a significant limitation. In particular, structural refinement simulations that require relaxation of a molecule into a new conformation, as in the case of fitting to a cryo-EM density map or performing MD simulation with experimental restraints, may necessitate the crossing of high-energy barriers to reach an optimally refined structure, as illustrated in Figure 3.

A number of methods facilitate the traversal of energy barriers during MD simulations to enhance conformational sampling, all of which include the introduction of a system bias. For example, the bias may take the form of increased kinetic energy tantamount to increased temperature. A basic protocol involves raising the system temperature to overcome barriers and then subsequently reducing it to arrive at low-energy conformations outside the original landscape well. This technique, referred to as simulated annealing, has long been applied in MD simulations to relax initial conformations and represents a common approach used in structural refinement simulations (67). More complex annealing methods that make use of randomly selected conformations, e.g., generalized simulated annealing (GSA), have also been employed for structure prediction of highly flexible protein linkers and loops (16). In particular, GSA is useful when applied to global optimization problems, because it can explore the conformational space more homogeneously and maintains the ability to make significant jumps even at low temperature (94). Combining sampling methods to overcome energetic barriers is also a common strategy; for example, to solve the structure of the yeast nuclear pore complex, Alber et al. (7) used an optimization protocol comprising simulated annealing, molecular dynamics, and conjugate gradient minimization.

Another robust and widely used enhanced sampling technique, replica exchange molecular dynamics (REMD) (131), runs multiple parallel trajectories at different discrete temperatures (T-REMD), including the physiological temperature; temperatures are spaced closely enough such that the thermal conformations arising exhibit overlapping energy distributions. During this type of simulation, molecular conformations are exchanged between systems at different temperatures according to a Monte Carlo scheme, facilitating efficient sampling of higher-energy conformations at physiological temperature. REMD has been incorporated efficiently into NAMD, permitting users to exploit the power of massively parallel computers to run large numbers of separate, yet communicating MD trajectories (104). Recent work has demonstrated that REMD can be applied to enhance conformational sampling of a broad range of systems, from the smallest peptides to large biomolecules (1, 103). Notably, large-scale REMD calculations were recently employed on a 0.6 million-atom subsystem of an immature retroviral lattice to investigate the mobility of a critical domain (54). The success of the REMD method has paved the way for the development of many variations on the approach, including reservoir REMD, constant-pH REMD, Hamiltonians REMD, and multiplexed REMD, which have been shown to achieve yet more appropriate sampling in shorter simulation times compared with T-REMD (16).

Notwithstanding the availability of advanced schemes for efficiently exploring conformational space, sampling methods, and MD simulations in general, are nevertheless limited by the quality of MD force fields employed for molecular descriptions. The wide use of REMD has demonstrated that different force fields tend to bias certain secondary structure, strongly indicating the necessity for force field improvement (76). As classical force fields fail at accurately reproducing the potential energy surface of proteins (38, 53, 82), the development of more reliable force fields represents a significant challenge for the field of computational structural refinement. A promising recent advance in this regard has been the development of a force field featuring atomic polarizability that is applicable to very large systems (83, 86).

APPLICATIONS OF COMPUTATIONAL STRUCTURAL REFINEMENT METHODS

The hybrid methods reviewed here comprise a toolkit of established approaches for integrating experimental data from diverse sources to perform computational structural refinement. Although a general strategy governs the way data are typically integrated (Figure 2), there is no precise recipe dictating the use of the techniques described in any particular case. As biological systems are complex and uniquely different from each other (Figure 4), one-size-fits-all application of any refinement method would produce suboptimal results. Therefore, the structural refinement protocol employed for any particular project should reflect the nuances of the system under study as well as the experimental information available. We review a collection of representative, yet distinct, examples in which hybrid methods were used to determine all-atom structures of the following biomolecular systems: ribosome, bacterial chemosensory array, virus capsids, and photosynthetic chromatophore.

Ribosome

The ribosome is a large (2.5–4.5 MDa) molecular machine responsible for translating genetic material into functional proteins. Because of its sheer size and complexity, the ribosome presents an outstanding challenge for traditional high-resolution structure determination methods such as X-ray crystallography and NMR spectroscopy. Cryo-EM has provided an alternative source of structural data, producing density maps at resolutions of 3–12 Å for the *Escherichia coli* ribosome in different functional states (4, 20, 50, 137). MDFF was employed to generate an atomistic structural model from the medium-resolution density maps (7–12 Å) (123, 140) by applying a multistep protocol to refine the ribosomal RNA, proteins, and various ligands. Notwithstanding the prowess of cryo-EM structure determination for large system sizes, X-ray crystallography also proved successful in providing detailed structural insights on a complete ribosome (10, 152, 153), only marginally later in time.

Nascent peptide chains synthesized and released by the ribosome are met at the ribosome exit by chaperone proteins, such as the trigger factor (TF) (60), that protect and fold the peptides. TF folds cytoplasmic proteins into their native states, and membrane proteins are assisted in their membrane placement by insertases such as YidC (70) and SecY (44, 51) shown in Figure 4*a*. TF was long thought to bind to the ribosome as a rigid molecule (71, 110); however, recent structures of TF bound to ribosomes, generated by applying MDFF to subnanometer cryo-EM density maps, reveal distinct degrees of flexibility and suggest a conformational transition induced by ribosome binding within the different TF domains.

The YidC model was derived using MDFF and evolutionary covariance analysis (146). Independent X-ray crystallography (74), mutational, and *N*-ethyl-acetamide-mediated water accessibility studies (126) have validated the model. After three decades of research, investigators ascertained that a YidC monomer is also sufficient to insert small hydrophobic helices into the membrane (74, 146).

Bacterial Chemosensory Array

Bacteria utilize a fundamental signaling process known as chemotaxis to interpret environmental chemical gradients and, in response, place themselves within the nutrient-optimal portion of their habitats (45, 141). Central to their chemotactic ability, bacteria possess a universally conserved, supramolecular protein complex known as the chemosensory array, which binds chemicals in the environment and processes the resulting information to direct cellular swimming behavior (19, 47). A wealth of multiscale structural data characterizing the chemosensory array are available, including atomistic structures of the individual components from X-ray crystallography (17, 78,102) and NMR spectroscopy (56, 79), as well as lower-resolution images of the array's global architecture from cryo-electron tomography (cryo-ET) (18, 84). Nevertheless, a high-resolution description of the intact and extended array structure has remained elusive, hindering the investigation of molecular mechanisms underlying signal transduction and regulation within the array.

Recently, hybrid modeling techniques were employed to combine existing structural data with new, highly resolved cryo-ET density maps, producing an atomistic model of the

chemosensory array (27). The model construction process is outlined in Figure 5. Briefly, MDFF simulations with symmetry restraints (28) were used to refine the tertiary structure of the array's component proteins to match the array-bound conformations represented in cryo-ET data, allowing the characterization of novel interaction interfaces between the kinase and chemoreceptor proteins. The array model was further interrogated using unbiased all-atom MD simulations, revealing a novel conformational change in the catalytic domain of the kinase and highlighting key residues affecting its conformational dynamics. Future studies combining computation and experiment are expected to yield additional significant insights into the chemosensory array's structure and function.

Virus Capsids

Virus capsids, made of protein, are shells that encase and protect the viral genome. Computational methods, particularly MDFF, have been extensively applied to refine the structures of virus capsids with icosahedral symmetry (99, 144). Whereas it is increasingly routine to experimentally obtain high-resolution structures for spherical or cylindrical capsids by taking advantage of their symmetric architecture (85), it remains challenging to obtain high-resolution structures for asymmetric capsids. For example, HIV-1 capsids are large (~30 MDa) and notably irregular, exhibiting mature states that are polymorphic in nature (156) and immature states that form incomplete shells (121). MDFF played an instrumental role in refining a critical trimeric interface between HIV-1 mature capsid proteins, which suggested mutations that directly resulted in the formation of the mature HIV-1 core in vitro (156). An all-atom model of the mature HIV-1 capsid (Figure 1, right) was subsequently constructed and characterized by combining data from cryo-EM, NMR spectroscopy, and MD simulations.

Besides driving MDFF for structural refinement, cryo-EM density maps have also been employed as templates for homology modeling of virus capsids. In particular, an atomistic model of the immature Rous sarcoma virus (RSV) capsid (Figure 4*d*) lattice was proposed to have a similar quaternary arrangement as the immature capsid lattice of Mason-Pfizer monkey virus (MPMV) (54). The construction of the model involved docking the X-ray and NMR structures of mature RSV capsid proteins into their corresponding regions of the immature MPMV cryo-EM density map. Notably, the spacer peptide (SP) and part of the nucleocapsid (NC) domain were previously proposed to oligomerize into a six-helix bundle (6HB) structure (148). The 6HB model, originally generated using an ion channel as a structural homolog (24), was refined by adjusting the packing between the helices, and its stability was probed by MD simulation (54). Microsecond-long REMD simulations revealed that the 6HB adopts multiple orientations, accounting for the lower resolution seen in the corresponding domains of the cryo-EM density maps (54, 121). The final immature lattice, which was selected using a clustering method, exhibits key interactions that are consistent with mutagenesis experiments (105, 151). An immature RSV capsid recently solved by cryo-EM experiments displays an architecture consistent with the computationally derived lattice model (120). The atomistic models of the mature and immature virus capsids serve as a platform to investigate the delicate interactions of the capsid with cellular host factors (83, 87), as well as with numerous assembly-inhibiting and maturation-inhibiting molecules (142).

Chromatophore

Chromatophores are bacterial photosynthetic organelles that form in certain bacteria as extensions of the cytoplasmic membrane upon transition to phototrophic growth. The chromatophores house an array of membrane proteins (96) that harvest sunlight and utilize it in a series of electron and proton transfer processes to synthesize ATP from ADP, thereby providing the energy that drives the regular functions of a bacterial cell. The shape of these chromatophores varies among species; the two most common forms are stacks of flat lamellar folds and spherical vesicles, the latter of which are approximately 700 Å in diameter.

Figure 6 outlines the construction process of the all-atom model of a lamellar (flat) chromatophore (29) based on X-ray crystallography and experimental microscopy data (113). The macromolecular assembly consists of circular light-harvesting complex 1 surrounding a photosynthetic reaction center (LH1-RC) and circular light-harvesting complex 2 (LH2) proteins embedded in a lipid membrane. Whereas X-ray structures of LH2 were available (100), for LH1-RC a homology model had to be constructed and refined by MDFF to match available cryo-EM data (124). Thirty-nine LH2 and 7 LH1-RC proteins were placed in a membrane patch (113) by means of rigid-body mapping into an atomic force microscopy (AFM) image (29), which clearly displayed the relative positions of small LH2 and large LH1-RC ring-shaped complexes. As the experimental image represented a shriveled system, steric clashes arose and were removed manually. The membrane patch, comprising the empty space between and around the planar protein system, was constructed as a bilayer consisting of a representative mixture of four lipid types, the distributions of which were obtained from chromatography and titration data (5, 89). Finally, 150 charge-carrier quinone molecules were added throughout the membrane patch, and the resulting system was solvated in an aqueous, ionic solution. As placement according to AFM data could not account for distortion of the flexible LH1-RC rings, or the tilt of the LH1-RC and LH2 complexes in the membrane, an MD simulation was performed to refine the overall system by relaxing it into a stable and stationary state.

The chromatophore membrane model (29) depicted in Figure 6 demonstrates that hybrid structural refinement techniques are very flexible regarding the type of experimental data that can be integrated. A model for spherical chromatophores composed of LH2, LH1, RC, and ATP synthase has also been achieved by additionally incorporating into the refinement process data from optical spectroscopy, nanogold-labeling, and mass spectroscopy (26). Atomic-level chromatophore models permit detailed investigations on how the system utilizes initially absorbed sunlight. For example, critical excitation transfer rates of the chromatophore can be calculated quantum-mechanically, where the accuracy of the calculation is sensitive to the atomic-level geometries for nearest LH1-RC and LH2 neighbors.

FUTURE CHALLENGES OF COMPUTATIONAL STRUCTURAL REFINEMENT

The hybrid methods reviewed here have already played essential roles in the construction of fully atomistic models for the biomolecular systems presented above. In particular, the MDFF method is routinely used not only by the authors of this article (e.g., 27, 140, 146,

156), but also by other researchers (e.g., 9, 101, 121, 157), and has resulted in over 80 publications. As experimental techniques continue to improve and provide better-quality, higher-resolution structural data for large systems, computational approaches must evolve to match these experimental advancements. Below, we discuss future challenges and the anticipated direction of the large-scale structural refinement field, with particular emphasis on MDFF.

Integrating Higher-Resolution Experimental Data

The advent of the direct electron detector, combined with technological advances in image processing, has significantly improved the quality of cryo-EM maps to near-atomic resolution (11, 12, 20, 49, 80). However, most high-resolution maps presented today do not display uniform resolution throughout the entire molecule. Hence, in silico structure prediction and homology modeling methods remain essential for generating complete, all-atom structures for the foreseeable future. Furthermore, the ability to incorporate data from diverse experimental sources (6, 7, 29), beyond X-ray crystallography and cryo-EM, to produce the best possible structural model remains a unique strength of computational hybrid methods, which are therefore anticipated to continue to play a critical role in structural refinement, even in the era of high-resolution cryo-EM.

As experimental methods improve further, established computational approaches must also evolve. Flexible fitting programs such as MDFF were originally designed to refine structures from medium-resolution cryo-EM density maps (~5–8 Å) (135, 136); today's higher-resolution maps display more distinct structural features that can trap the computed structure in non-native conformations during the MDFF procedure. Therefore, enhanced sampling methods such as REMD (131) and GSA (16) may need to be incorporated into the MDFF protocol to overcome local minima and explore conformational space more efficiently. Additionally, high-resolution tools such as Phenix (39), which were initially designed for dealing with data from X-ray crystallography, are being employed for structural refinement for highly resolved domains in cryo-EM density maps. Low local resolution in cryo-EM density maps can often be attributed to sample heterogeneity and inherent flexibility of the molecule studied. If a mixture of conformational states is present, the averaging of these states during image reconstruction results in blurring and ultimately loss of structural detail. This loss might be prevented by Bayesian analysis, in which raw cryo-EM images are processed directly to identify distinct ensembles (classes) of states (35, 111). A potential strategy for incorporating Bayesian analysis into the MDFF protocol involves identifying distinct conformational states sampled during an MD simulation and using representative structures from each state to better interpret the experimental maps prior to flexible fitting. Conformational variety can also be explored by individual-particle electron tomography (IPET), although the resolution is at best 15–20 Å (154). In a recent study, high-resolution models for distinct conformational states were constructed by applying MDFF to flexibly fit X-ray structures of an antibody to corresponding IPET data (155). Overall, the interface between real-space computational refinement methods and experimental data-processing protocols represents an area of untapped potential to improve interpretation of cryo-EM data (116).

Advancements on the Computational Front

Computational refinement methods will benefit in the future from significant advancements in protocol automation and efficiency, as well as from user-friendliness. For example, the current MDFF protocol places the onus solely on the user to assess the quality of model fit to target density; however, moving forward, quality assessment could be carried out routinely through real-time in situ analysis of the local cross-correlation coefficient (CCC) (130) or the elastic deformational energy accumulated during refinement. Improved MDFF implementation would discern the fitting quality of structural segments that converge quickly from structural segments that are hard to fit, thereby channeling the computational effort to the latter. More robust metrics for determining quality of fit, free of the inherent shortcomings of CCC, should also be developed for this purpose. To improve the accuracy of structural models, more realistic force fields should be employed during the refinement and conformational sampling processes. Although present fixed-charge force fields have been applied successfully to model and refine large structures (144, 156), they struggle to properly account for molecular interactions that involve highly charged components. In particular, specific protein–ion interactions found in metalloproteins are often poorly described and require empirical treatment to reproduce experimental detail (55). The use of force fields that include atomic polarization, such as those based on Drude oscillator models (86) or the AMOEBA force field (106), or those that alternatively employ explicit hydrogen-bonding terms or ghost atoms representing lone pair charge sites (53), should notably improve structural refinement and sampling. For example, the Drude polarizable force field, which is implemented and optimized in NAMD (63), has been recently applied very successfully for cryo-EM density map fitting and refinement simulations to study the interaction between the HIV-1 virus capsid and host cell factors (83).

Addressing the Bigger Issue: Reproducibility

Reproducibility is a topic of intense debate in the biomedical and life sciences community (15, 52). One primary reason why the reproducibility rate of scientific results is low is often lack of detail in protocol documentation. Indeed, the typical modus of presenting research procedures within published Methods and Supplementary Information text often proves insufficient to guide accurate reproduction of an experiment or a calculation. In that case, what is the minimum requirement for reproducibility? For the field of computational structural refinement, utter transparency, demonstrated by providing the structural models both before and after refinement, as well as listing all key parameters used during the refinement process, should be considered indispensable. Although results from computational techniques are relatively easy to reproduce compared with results from biochemical experiments owing to the deterministic nature of a computer, there currently exists no standard reporting framework to systematically document a computational procedure. In response to this dilemma, major modeling programs such as VMD are working toward providing options to comprehensively log the key modeling steps, printing extensive log files that can easily be attached to publications as supplementary material.

Because the PDB does not accept purely computationally derived structural models, many such models are made publicly available in the form of supplementary material (54, 107). This practice has yet to become common, as it is currently not required for publication. To

enhance research integrity across the board for the structure determination field, investigators should consider open availability of computationally derived models as critically important as the deposition of experimentally derived atomic structures in the PDB and cryo-EM maps in the Electron Microscopy Data Bank (EMDB). With the US National Institutes of Health (NIH) beginning to explore strategies to enhance reproducibility of scientific results (34), as well as top scientific journals making a concerted effort to establish guidelines for improved reporting practices (93), the field is poised to make dramatic strides toward increasing the transparency and credibility that underlie professional and public trust in science.

Dealing with Biological Systems of Increasing Size

A question that inevitably arises when discussing the future of computational structural refinement is the extension of hybrid methods to increasingly large biomolecular systems. With the expanding size of systems under study come expanding data sets and, therefore, increasing storage needs. Although the cost of storing data is relatively low, very large data sets become problematic when the need arises to transfer them between locations. For example, a supercomputer center may provide computational power and adequate disk space to perform large-scale structural refinement simulations; however, limited bandwidth and storage capacity might prohibit the transfer of results to researchers' desktops for evaluation. Fortunately, this problem can be addressed through powerful remote visualization solutions that analyze computational data where they are generated, namely at the computer centers, and stream only the resulting graphics to researchers' local commodity computers such as laptops (25).

Given the recent rate of advancement in both experimental and computational fields, which now enable researchers to describe systems comprising hundreds of millions of atoms, we expect to carry out billion-atom structural refinement within less than a decade. Supporting the feasibility of this goal is the development of exascale supercomputers, which is projected to be completed by the year 2023. At the exascale, the types of colossal macromolecular assemblies that can be structurally investigated, e.g., the interaction of virus capsids with the nuclear pore complex, will answer key biological questions related to human health and the treatment of disease. Further, the billion-atom regime will enable the study of entire cellular organelles, or even small whole cells, in full atomistic detail. For example, the structural composition of the pathogenic bacterium *Mycoplasma genitalium*, a cofactor in HIV transmission, is a whole-cell target placed within reach for the next decade of large-scale structure determination by computational hybrid methods. The leading MD program for large biomolecular simulation, NAMD (104), is currently capable of handling systems of up to 2 billion atoms. Forthcoming optimizations in the NAMD code, particularly with respect to the capabilities of MDFF, as well as data analysis efficiency in VMD (61), will render hybrid methods well prepared to take advantage of exascale supercomputers for billion-atom structural refinement as these powerful computational resources become available.

CONCLUDING REMARKS

In conclusion, computational approaches will continue to play an essential role in the large-scale structural refinement field, strongly complementing recent far-reaching advancements in experimental structure determination. The unique ability of hybrid methods to incorporate data from diverse sources to produce the highest-quality structural models, as well as the critical initial requirement of rebuilding unresolved structural features, renders them invaluable to the process of elucidating all-atom structures for large multiprotein complexes now and into the foreseeable future. Following upcoming improvements and new developments in molecular mechanics force field parameterization, automated model quality assessment strategies, and the powerful features of structural refinement tools such as MDFF, computational hybrid methods will be applied toward solving increasingly large and more complex biomolecular systems, helping to gradually reveal, atom-by-atom, the awe-inspiring architecture of living cells in their entirety.

Acknowledgments

The authors thank Juan R. Perilla, John E. Stone, and James C. Phillips for insightful discussions that contributed to this review, as well as Bo Liu for kindly providing a ribosome figure. The authors gratefully acknowledge funding from the National Institutes of Health (NIH, 9P41GM104601, R01-GM067887, U54 GM087519, 5 R01 GM098243-02), the National Science Foundation (NSF, MCB-1157615, PHY1430124), and the Energy Biosciences Institute (EBI, 231 UCB BP 2014OO4J01). T.R. is supported by the Alexander von Humboldt Foundation.

KEY TERMS AND DEFINITIONS

Atomic force microscopy (AFM)
an imaging technique used to mechanically probe the surface structure of a material at near-atomic resolution

Cryo-electron microscopy (cryo-EM)
a method to determine the structure of biological complexes in their native environment, typically at the temperature of liquid nitrogen

Cryo-electron tomography (cryo-ET)
a cryo-EM technique that produces 3D reconstructions of macromolecular complexes from a series of images collected at different angles relative to an irradiating electron beam

Electron paramagnetic resonance (EPR) spectroscopy
a method that characterizes the distances between unpaired, site-directed spin labels introduced by recombinant protein expression

Fluorescence resonance energy transfer (FRET) microscopy
a method that measures the time evolution of the distance between two fluorophores based on their transfer efficiency

HIV
a retrovirus that causes AIDS in humans

Molecular dynamics (MD) simulation

a technique to compute the positions and velocities of particles using Newton's equations of motion and forces derived from a heuristic or quantum-chemically determined potential energy function

Nuclear magnetic resonance (NMR) spectroscopy
a method that yields structures and dynamics of proteins based on the magnetic properties of atomic nuclei

Protein Data Bank (PDB)
a database of 3D, atomistic protein structures solved mainly by X-ray crystallography and NMR spectroscopy

Replica exchange molecular dynamics (REMD)
an enhanced sampling method to efficiently explore the energy landscape of a biomolecule

Rous sarcoma virus (RSV)
a retrovirus that causes cancer in chicken

Small-angle X-ray scattering (SAXS)
a method that determines shape and size of a macromolecule in the resolution range of 1 to 5 nm

X-ray crystallography
a technique to obtain the 3D structure of a molecule in its crystalline form based on X-ray diffraction patterns;

LITERATURE CITED

1. Abrams C, Bussi G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. Entropy. 2013; 16:163–99.
2. Adams PD, Afonine PV, Grosse-Kunstleve RW, Read RJ, Richardson JS, et al. Recent developments in phasing and structure refinement for macromolecular crystallography. Curr. Opin. Struct. Biol. 2009; 9:566–72.
3. Adams PD, Baker D, Brunger AT, Das R, DiMaio F, et al. Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems. Annu. Rev. Biophys. 2013; 42:265–87. [PubMed: 23451892]
4. Agrawal RK, Heagle AB, Penczek P, Grassucci RA, Frank J. EF-G-dependent GTP hydrolysis induces translocation accompanied by large conformational changes in the 70S ribosome. Nat. Struct. Biol. 1999; 6:643–47. [PubMed: 10404220]
5. Al-Bayatti KK, Takemoto JY. Phospholipid topography of the photosynthetic membrane of *Rhodopseudomonas sphaeroides*. Biochemistry. 1981; 20:5489–95. [PubMed: 6975121]
6. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, et al. Determining the architectures of macromolecular assemblies. Nature. 2007; 450:683–94. [PubMed: 18046405]
7. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, et al. The molecular architecture of the nuclear pore complex. Nature. 2007; 450:695–701. [PubMed: 18046406]
8. Arnold K, Bordoli L, Kopp J, Schwede T. The Swiss-model workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006; 22:195–201. [PubMed: 16301204]
9. Asano S, Fukuda Y, Beck F, Aufderheide A, Förster F, et al. A molecular census of 26S proteasomes in intact neurons. Science. 2015; 347:439–42. [PubMed: 25613890]
10. Bashan A, Yonath A. Ribosome crystallography: from early evolution to contemporary medical insights. Ribosomes. 2011; 1:3–18.

11. Bartesaghi A, Matthies D, Banerjee S, Merk A, Subramaniam S. Structure of β-galactosidase at 3.2 Å resolution obtained by cryo-electron microscopy. PNAS. 2014; 111:11709–14. [PubMed: 25071206]
12. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, et al. 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. Science. 2015; 348:1147–51. [PubMed: 25953817]
13. Barty A, Küpper J, Chapman HN. Molecular imaging using X-ray free-electron lasers. Annu. Rev. Phys. Chem. 2013; 64:415–35. [PubMed: 23331310]
14. Batista PR, Costa MGDS, Pascutti PG, Bisch PM, de Souza W. High temperatures enhance cooperative motions between CBM and catalytic domains of a thermostable cellulase: mechanism insights from essential dynamics. Phys. Chem. Chem. Phys. 2011; 13:13709–20. [PubMed: 21713261]
15. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. Nature. 2012; 483:531–33. [PubMed: 22460880]
16. Bernardi RC, Melo MCR, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim. Biophys. Acta. 2015; 1850:872–77. [PubMed: 25450171]
17. Bilwes A, Alex L, Crane B, Simon M. Structure of CheA, a signal-transducing histidine kinase. Cell. 1999; 96:131–41. [PubMed: 9989504]
18. Briegel A, Li X, Bilwes AM, Hughes KT, Jensen GJ, Crane BR. Bacterial chemoreceptor arrays are hexagonally packed trimers of receptor dimers networked by rings of kinase and coupling proteins. PNAS. 2012; 109:3766–71. [PubMed: 22355139]
19. Briegel A, Ortega DR, Tocheva EI, Wuichet K, Li Z, et al. Universal architecture of bacterial chemoreceptor arrays. Nat. Rev. Mol. Cell Biol. 2009; 106:17181–86.
20. Brown A, Shao S, Murray J, Hegde RS, Ramakrishnan V. Structural basis for stop codon recognition in eukaryotes. Nature. 2015; 524:493–96. [PubMed: 26245381]
21. Brünger AT. The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature. 1992; 355:472–74. [PubMed: 18481394]
22. Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D. 1998; 54:905–21. [PubMed: 9757107]
23. Brünger AT, Kuriyan J, Karplus M. Crystallographic *R* factor refinement by molecular dynamics. Science. 1987; 235:458–60. [PubMed: 17810339]
24. Bush DL, Monroe EB, Bedwell GJ, Prevelige PE, Phillips JM, Vogt VM. Higher order structure of the Rous sarcoma virus SP assembly domain. J. Virol. 2014; 88:5617–29. [PubMed: 24599998]
25. Carroll MD, Hadzic I, Katsak WA. 3D rendering in the cloud. Bell Labs Technical J. 2012; 17:55–66.
26. Cartron ML, Olsen JD, Sener M, Jackson PJ, Brindley AA, et al. Integration of energy and electron transfer processes in the photosynthetic membrane of *Rhodobacter sphaeroides*. Biochim. Biophys. Acta. 2014; 1837:1769–80. [PubMed: 24530865]
27. Cassidy CK, Himes BA, Alvarez FJ, Ma J, Zhao G, et al. Cryo-electron tomography and all-atom molecular dynamics simulations reveal a novel kinase conformational switch in bacterial chemotaxis signaling. eLife. 2015; 4:e08419. [PubMed: 26583751]
28. Chan KY, Gumbart J, McGreevy R, Watermeyer JM, Sewell BT, Schulten K. Symmetry-restrained flexible fitting for symmetric EM maps. Structure. 2011; 19:1211–18. [PubMed: 21893283]
29. Chandler D, Strümpfer J, Sener M, Scheuring S, Schulten K. Light harvesting by lamellar chromatophores in *Rhodospirillum photometricum*. Biophys. J. 2014; 106:2503–10. [PubMed: 24896130]
30. Chapman MS. Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function. Acta Crystallogr A. 1995; 51:69–80.
31. Chapman MS, Trzynka A, Chapman BK. Atomic modeling of cryo-electron microscopy reconstructions—joint refinement of model and imaging parameters. J. Struct. Biol. 2013; 182:10–21. [PubMed: 23376441]

32. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D. 2010; 66:12–21. [PubMed: 20057044]
33. Cheng X, Jo S, Qi Y, Marassi FM, Im W. Solid-state NMR-restrained ensemble dynamics of a membrane protein in explicit membranes. Biophys. J. 2015; 108:1954–62. [PubMed: 25902435]
34. Collins FS, Tabak LA. NIH plans to enhance reproducibility. Nature. 2014; 505:612–13. [PubMed: 24482835]
35. Cossio P, Hummer G. Bayesian analysis of individual electron microscopy images: towards structures of dynamic and heterogeneous biomolecular assemblies. J. Struct. Biol. 2013; 184:427–37. [PubMed: 24161733]
36. Delarue M. Dealing with structural variability in molecular replacement and crystallographic refinement through normal-mode analysis. Acta Crystallogr. D. 2008; 64:40–48. [PubMed: 18094466]
37. Diamond R. A real-space refinement procedure for proteins. Acta Crystallogr. A. 1971; 27:436–52.
38. Dill KA, MacCallum JL. The protein-folding problem, 50 years on. Science. 2012; 338:1042–6. [PubMed: 23180855]
39. DiMaio F, Echols N, Headd JJ, Terwilliger TC, Adams PD, Baker D. Improved low-resolution crystallographic refinement with Phenix and Rosetta. Nat. Methods. 2013; 10:1102–4.
40. DiMaio F, Song Y, Li X, Brunner MJ, Xu C, et al. Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. Nat. Methods. 2015; 12:361–65. [PubMed: 25707030]
41. DiMaio F, Terwilliger TC, Read RJ, Wlodawer A, Oberdorfer G, et al. Increasing the radius of convergence of molecular replacement by density- and energy-guided protein structure optimization. Nature. 2011; 473:540–43. [PubMed: 21532589]
42. DiMaio F, Zhang J, Chiu W, Baker D. Cryo-EM model validation using independent map reconstructions. Protein Sci. 2013; 22:865–68. [PubMed: 23592445]
43. Dolan MA, Noah JW, Hurt D. Comparison of common homology modeling algorithms: application of user-defined alignments. Methods Mol. Biol. 2012; 857:399–414. [PubMed: 22323232]
44. Duong F, Wickner W. Distinct catalytic roles of the SecYE, SecG and SecDFyajC subunits of preprotein translocase holoenzyme. EMBO J. 1997; 16:2756–68. [PubMed: 9184221]
45. Eisenbach, M. Chemotaxis. London: Imperial College Press; 2004.
46. Eswar N, Webb B, Marti-Renom MA, Madhusudhan M, Eramian D, et al. Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinform. 2006; 5 5.6.
47. Falke JJ, Piasta KN. Architecture and signal transduction mechanism of the bacterial chemosensory array: progress, controversies, and challenges. Curr. Opin. Struct. Biol. 2014; 29:85–94. [PubMed: 25460272]
48. Fenn TD, Schnieders MJ, Mustyakimov M, Wu C, Langan P, et al. Reintroducing electrostatics into macromolecular crystallographic refinement: application to neutron crystallography and DNA hydration. Structure. 2011; 19:523–33. [PubMed: 21481775]
49. Fischer N, Neumann P, Konevega AL, Bock LV, Ficner R, et al. Structure of the *E. coli* ribosome-EF-Tu complex at <3 Å resolution by *cs*-corrected cryo-EM. Nature. 2015; 520:567–70. [PubMed: 25707802]
50. Frank J, Agrawal RK. A ratchet-like inter-subunit reorganization of the ribosome during translocation. Nature. 2000; 406:318–22. [PubMed: 10917535]
51. Frauenfeld J, Gumbart J, van der Sluis EO, Funes S, Gartmann M, et al. Cryo-EM structure of the ribosome-SecYE complex in the membrane environment. Nat. Struct. Mol. Biol. 2011; 18:614–21. [PubMed: 21499241]
52. Freedman LP, Cockburn IM, Simcoe TS. The economics of reproducibility in preclinical research. PLOS Biol. 2015; 13:e1002165. [PubMed: 26057340]
53. Freddolino PL, Harrison CB, Liu Y, Schulten K. Challenges in protein folding simulations. Nat. Phys. 2010; 6:751–58. [PubMed: 21297873]

54. Goh BC, Perilla JR, England MR, Heyrana KJ, Craven RC, Schulten K. Atomic modeling of an immature retroviral lattice using molecular dynamics and mutagenesis. Structure. 2015; 23:1414–25. [PubMed: 26118533]
55. Goh BC, Rynkiewicz MJ, Cafarella TR, White MR, Hartshorn KL, et al. Molecular mechanisms of inhibition of influenza by surfactant protein D revealed by large-scale molecular dynamics simulation. Biochemistry. 2013; 52:8527–38. [PubMed: 24224757]
56. Griswold IJ, Zhou H, Matison M, Swanson RV, McIntosh LP, et al. The solution structure and interactions of CheW from *Thermotoga maritima*. Nat. Struct. Biol. 2002; 9:121–25. [PubMed: 11799399]
57. Gruner SM, Lattman EE. Biostructural science inspired by next-generation X-ray sources. Annu. Rev. Biophys. 2015; 44:33–51. [PubMed: 25747590]
58. Haddadian EJ, Gong H, Jha AK, Yang X, DeBartolo J, et al. Automated real-space refinement of protein structures using a realistic backbone move set. Biophys. J. 2011; 101:899–909. [PubMed: 21843481]
59. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, et al. Outcome of the first electron microscopy validation task force meeting. Structure. 2012; 20:205–14. [PubMed: 22325770]
60. Hoffmann A, Bukau B, Kramer G. Structure and function of the molecular chaperone trigger factor. Biochim. Biophys. Acta. 2010; 1803:650–61. [PubMed: 20132842]
61. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J. Mol. Graph. 1996; 14:33–38. [PubMed: 8744570]
62. Islam SM, Stein RA, Mchaourab HS, Roux B. Structural refinement from restrained-ensemble simulations based on EPR/DEER data: application to T4 lysozyme. J. Phys. Chem. B. 2013; 117:4740–54. [PubMed: 23510103]
63. Jiang W, Hardy D, Phillips J, MacKerell A, Schulten K, Roux B. High-performance scalable molecular dynamics simulations of a polarizable force field based on classical Drude oscillators in NAMD. J. Phys. Chem. Lett. 2011; 2:87–92. [PubMed: 21572567]
64. Jolley CC, Wells SA, Fromme P, Thorpe MF. Fitting low-resolution cryo-EM maps of proteins using constrained geometric simulations. Biophys. J. 2008; 94:1613–21. [PubMed: 17993504]
65. Kantardjieff KA, Rupp B. Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. Protein Sci. 2003; 12:1865–71. [PubMed: 12930986]
66. Karmali AM, Blundell TL, Furnham N. Model-building strategies for low-resolution X-ray crystallographic data. Acta Crystallogr. D. 2009; 65:121–27. [PubMed: 19171966]
67. Karplus M, Petsko GA. Molecular dynamics simulations in biology. Nature. 1990; 347:631–39. [PubMed: 2215695]
68. Kaufmann KW, Lemmon GH, DeLuca SL, Sheehan JH, Meiler J. Practically useful: what the Rosetta protein modeling suite can do for you. Biochemistry. 2010; 49:2987–98. [PubMed: 20235548]
69. Koepke J, Hu X, Muenke C, Schulten K, Michel H. The crystal structure of the light-harvesting complex II (B800-850) from *Rhodospirillum molischianum*. Structure. 1996; 4:581–97. [PubMed: 8736556]
70. Kol S, Nouwen N, Driessen AJM. Mechanisms of YidC-mediated insertion and assembly of multimeric membrane protein complexes. J. Biol. Chem. 2008; 283:31269–73. [PubMed: 18658156]
71. Krishtal O. The ASICs: signaling molecules? Modulators? Trends Neurosci. 2003; 26:477–83. [PubMed: 12948658]
72. Kucukelbir A, Sigworth FJ, Tagare HD. Quantifying the local resolution of cryo-EM density maps. Nat. Methods. 2014; 11:63–65. [PubMed: 24213166]
73. Kudryashev M, Wang RYR, Brackmann M, Scherer S, Maier T, et al. Structure of the type VI secretion system contractile sheath. Cell. 2015; 160:952–62. [PubMed: 25723169]
74. Kumazaki K, Kishimoto T, Furukawa A, Mori H, Tanaka Y, et al. Crystal structure of *Escherichia coli* YidC, a membrane protein chaperone and insertase. Sci. Rep. 2014; 4:7299. [PubMed: 25466392]

75. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. *PROCHECK*: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 1993; 26:283–91.
76. Li E, Wimley WC, Hristova K. Transmembrane helix dimerization: beyond the search for sequence motifs. Biochim. Biophys. Acta. 2012; 1818:183–93. [PubMed: 21910966]
77. Li X, Mooney P, Zheng S, Booth CR, Braunfeld MB, et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. Nat. Methods. 2013; 10:584–90. [PubMed: 23644547]
78. Li X, Ortega DR, Bilwes AM, Falke JJ, Zhulin IB, Crane BR. The 3.2 Å resolution structure of a receptor:CheA:CheW signaling complex defines overlapping binding sites and key residue interactions within bacterial chemosensory arrays. Biochemistry. 2013; 52:3866–80. [PubMed: 23668882]
79. Li Y, Hu Y, Fu W, Xia B, Jin C. Solution structure of the bacterial chemotaxis adaptor protein CheW from *Escherichia coli*. Biochem. Biophys. Res. Commun. 2007; 360:863–67. [PubMed: 17631272]
80. Liao M, Cao E, Julius D, Cheng Y. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. Nature. 2013; 504:107–12. [PubMed: 24305160]
81. Lindert S, McCammon JA. Improved cryoEM-guided iterative molecular dynamics–Rosetta protein structure refinement protocol for high precision protein structure prediction. J. Chem. Theory Comput. 2015; 11:1337–46. [PubMed: 25883538]
82. Lindorff-Larsen K, Piana S, Dror RO, Shaw DE. How fast-folding proteins fold. Science. 2011; 334:517–20. [PubMed: 22034434]
83. Liu C, Perilla JR, Ning J, Lu M, Hou G, et al. Cyclophilin A stabilizes the HIV-1 capsid through a novel non-canonical binding site. Nat. Commun. 2016; 7:10714. [PubMed: 26940118]
84. Liu J, Hu B, Morado DR, Jani S, Manson MD, Margolin W. Molecular architecture of chemoreceptor arrays revealed by cryoelectron tomography of *Escherichia coli* minicells. PNAS. 2012; 109:E1481–88. [PubMed: 22556268]
85. Liu Y, Sheng J, Fokine A, Meng G, Shin WH, et al. Structure and inhibition of EV-D68, a virus that causes respiratory illness in children. Science. 2015; 347:71–74. [PubMed: 25554786]
86. Lopes PEM, Huang J, Shim J, Luo Y, Li H, et al. Polarizable force field for peptides and proteins based on the classical Drude oscillator. J. Chem. Theor. Comp. 2013; 9:5430–49.
87. Lu M, Hou G, Zhang H, Suiter CL, Ahn J, et al. Dynamic allostery governs cyclophilin A-HIV capsid interplay. PNAS. 2015; 112:14617–22. [PubMed: 26553990]
88. Marinelli F, Faraldo-Gómez JD. Ensemble-biased metadynamics: a molecular simulation method to sample experimental distributions. Biophys. J. 2015; 108:2779–82. [PubMed: 26083917]
89. Marinetti GV, Cattieu K. Lipid analysis of cells and chromatophores of *Rhodopseudomonas sphaeroides*. Chem. Phys. Lipids. 1981; 28:241–51.
90. Matthews BW. Solvent content of protein crystals. J. Mol. Biol. 1968; 33:491–97. [PubMed: 5700707]
91. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J. Appl. Crystallogr. 2007; 40:658–74. [PubMed: 19461840]
92. McGreevy R, Singharoy A, Li Q, Zhang J, Xu D, et al. xMDFF: molecular dynamics flexible fitting of low-resolution X-ray structures. Acta Crystallogr. D. 2014; 70:2344–55. [PubMed: 25195748]
93. McNutt M. Journals unite for reproducibility. Science. 2014; 346:679. [PubMed: 25383411]
94. Melo MC, Bernardi RC, Fernandes TV, Pascutti PG. GSAFold: a new application of GSA to protein structure prediction. Proteins. 2012; 80:2305–10. [PubMed: 22622959]
95. Milazzo AC, Cheng A, Moeller A, Lyumkis D, Jacovetty E, et al. Initial evaluation of a direct detection device detector for single particle cryo-electron microscopy. J. Struct. Biol. 2011; 176:404–8. [PubMed: 21933715]
96. Monger T, Parson W. Singlet-triplet fusion in *Rhodopseudomonas sphaeroides* chromatophores. A probe of the organization of the photosynthetic apparatus. Biochim. Biophys. Acta. 1977; 460:393–407. [PubMed: 301747]

97. Murshudov G, Skubák P, Lebedev A, Pannu N, Steiner R, et al. *REFMAC5* for the refinement of macromolecular crystal structures. Acta Crystallogr. D. 2011; 67:355–67. [PubMed: 21460454]

98. Orzechowski M, Tama F. Flexible fitting of high-resolution X-ray structures into cryo electron microscopy maps using biased molecular dynamics simulations. Biophys. J. 2008; 95:5692–705. [PubMed: 18849406]

99. Pandurangan AP, Shakeel S, Butcher SJ, Topf M. Combined approaches to flexible fitting and assessment in virus capsids undergoing conformational change. J. Struct. Biol. 2014; 185:427–39. [PubMed: 24333899]

100. Papiz MZ, Prince SM, Howard T, Cogdell RJ, Isaacs NW. The structure and thermal motion of the B800-850 LH2 complex from *Rps. acidophila* at 2.0 Å resolution and 100 K: new structural features and functionally relevant motions. J. Mol. Biol. 2003; 326:1523–38. [PubMed: 12595263]

101. Park E, Ménétret JF, Gumbart JC, Ludtke SJ, Li W, et al. Structure of the SecY channel during initiation of protein translocation. Nature. 2014; 506:102–6. [PubMed: 24153188]

102. Park SY, Borbat PP, Gonzalez-Bonet G, Bhatnagar J, Pollard AM, et al. Reconstruction of the chemotaxis receptor-kinase assembly. Nat. Struct. Mol. Biol. 2006; 13:400–7. [PubMed: 16622408]

103. Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, et al. Molecular dynamics simulations of large macromolecular complexes. Curr. Opin. Struct. Biol. 2015; 31:64–74. [PubMed: 25845770]

104. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, et al. Scalable molecular dynamics with NAMD. J. Comp. Chem. 2005; 26:1781–802. [PubMed: 16222654]

105. Phillips JM, Murray PS, Murray D, Vogt VM. A molecular switch required for retrovirus assembly participates in the hexagonal immature lattice. EMBO J. 2008; 27:1411–20. [PubMed: 18401344]

106. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, et al. Current status of the AMOEBA polarizable force field. J. Phys. Chem. B. 2010; 114:2549–64. [PubMed: 20136072]

107. Reddy T, Shorthouse D, Parton DL, Jefferys E, Fowler PW, et al. Nothing to sneeze at: a dynamic and integrative computational model of an influenza A virion. Structure. 2015; 23:584–97. [PubMed: 25703376]

108. Roseman AM. Docking structures of domains into maps from cryo-electron microscopy using local correlation. Acta Crystallogr. D. 2000; 56:1332–40. [PubMed: 10998630]

109. Roux B, Islam SM. Restrained-ensemble molecular dynamics simulations based on distance histograms from double electron-electron resonance spectroscopy. J. Phys. Chem. B. 2013; 117:4733–39. [PubMed: 23510121]

110. Rutkowska A, Mayer MP, Hoffmann A, Merz F, Zachmann-Brand B, et al. Dynamics of trigger factor interaction with translating ribosomes. J. Biol. Chem. 2008; 283:4124–32. [PubMed: 18045873]

111. Scheres SHW. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. 2012; 180:519–30. [PubMed: 23000701]

112. Scheres SHW, Chen S. Prevention of overfitting in cryo-EM structure determination. Nat. Methods. 2012; 9:853–54. [PubMed: 22842542]

113. Scheuring S, Sturgis JN. Atomic force microscopy of the bacterial photosynthetic apparatus: plain pictures of an elaborate machinery. Photosyn. Res. 2009; 102:197–211. [PubMed: 19266309]

114. Schlichting I, Miao J. Emerging opportunities in structural biology with X-ray free-electron lasers. Curr. Opin. Struct. Biol. 22:613–26.

115. Schreiner E, Trabuco LG, Freddolino PL, Schulten K. Stereochemical errors and their implications for molecular dynamics simulations. BMC Bioinform. 2011; 12:190.

116. Schröder GF. Hybrid methods for macromolecular structure determination: experiment with expectations. Curr. Opin. Struct. Biol. 2015; 31:20–27. [PubMed: 25795086]

117. Schröder GF, Brunger AT, Levitt M. Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. Structure. 2007; 15:1630–41. [PubMed: 18073112]

118. Schröder GF, Levitt M, Brunger AT. Super-resolution biomolecular crystallography with low-resolution data. Nature. 2010; 464:1218–22. [PubMed: 20376006]

119. Schröder GF, Levitt M, Brunger AT. Deformable elastic network refinement for low-resolution macromolecular crystallography. Acta Crystallogr. D. 2014; 70:2241–55. [PubMed: 25195739]
120. Schur FKM, Dick RA, Hagen WJH, Vogt VM, Briggs JAG. The structure of immature virus-like Rous sarcoma virus Gag particles reveals a structural role for the p10 domain in assembly. J. Virol. 2015; 89:10294–302. [PubMed: 26223638]
121. Schur FKM, Hagen WJH, Rumlová M, Ruml T, Müller B, et al. Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution. Nature. 2015; 517:505–8. [PubMed: 25363765]
122. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. J. Magn. Reson. 2003; 160:65–73. [PubMed: 12565051]
123. Seidelt B, Innis CA, Wilson DN, Gartmann M, Armache JP, et al. Structural insight into nascent polypeptide chain-mediated translational stalling. Science. 2009; 326:1412–15. [PubMed: 19933110]
124. Sener MK, Hsin J, Trabuco LG, Villa E, Qian P, et al. Structural model and excitonic properties of the dimeric RC-LH1-PufX complex from *Rhodobacter sphaeroides*. Chem. Phys. 2009; 357:188–97. [PubMed: 20161332]
125. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006; 15:2507–24. [PubMed: 17075131]
126. Shimokawa-Chiba N, Kumazaki K, Tsukazaki T, Nureki O, Ito K, Chiba S. Hydrophilic microenvironment required for the channelin-dependent insertase function of YidC protein. PNAS. 2015; 112:5063–68. [PubMed: 25855636]
127. Shortle D, Simons KT, Bakér D. Clustering of low-energy conformations near the native structures of small proteins. PNAS. 1998; 95:11158–62. [PubMed: 9736706]
128. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J. Mol. Biol. 1997; 268:209–25. [PubMed: 9149153]
129. Stone, JE., Gullingsrud, J., Grayson, P., Schulten, K. A system for interactive molecular dynamics simulation. In: Hughes, JF., Séquin, CH., editors. 2001 ACM Symposium on Interactive 3D Graphics. New York: ACM Siggraph; 2001. p. 191-94.
130. Stone JE, McGreevy R, Isralewitz B, Schulten K. GPU-accelerated analysis and visualization of large structures solved by molecular dynamics flexible fitting. Faraday Discuss. 2014; 169:265–83. [PubMed: 25340325]
131. Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. Chem. Phys. Lett. 1999; 314:141–51.
132. Tama F, Miyashita O, Brooks CL 3rd. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. J. Struct. Biol. 2004; 147:315–26. [PubMed: 15450300]
133. 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. [PubMed: 18275820]
134. Trabuco LG, Schreiner E, Eargle J, Cornish P, Ha T, et al. The role of L1 stalk-tRNA interaction in the ribosome elongation cycle. J. Mol. Biol. 2010; 402:741–60. [PubMed: 20691699]
135. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K. Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. Structure. 2008; 16:673–83. [PubMed: 18462672]
136. Trabuco LG, Villa E, Schreiner E, Harrison CB, Schulten K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods. 2009; 49:174–80. [PubMed: 19398010]
137. Valle M, Zavialov A, Sengupta J, Rawat U, Ehrenberg M, Frank J. Locking and unlocking of ribosomal motions. Cell. 2003; 114:123–34. [PubMed: 12859903]
138. Vashisth H, Skiniotis G, Brooks CL 3rd. Using enhanced sampling and structural restraints to refine atomic structures into low-resolution electron microscopy maps. Structure. 2012; 20:1453–62. [PubMed: 22958641]

139. Venditti V, Egner TK, Clore GM. Hybrid approaches to structural characterization of conformational ensembles of complex macromolecular systems combining NMR residual dipolar couplings and solution X-ray scattering. Chem. Rev. 2016; doi: 10.1021/acs.chemrev.5b00592
140. Villa E, Sengupta J, Trabuco LG, LeBarron J, Baxter WT, et al. Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. PNAS. 2009; 106:1063–68. [PubMed: 19122150]
141. Wadhams GH, Armitage JP. Making sense of it all: bacterial chemotaxis. Nat. Rev. Mol. Cell Biol. 2004; 5:1024–37. [PubMed: 15573139]
142. Waheed AA, Freed EO. HIV type 1 Gag as a target for antiviral therapy. AIDS Res. Hum. Retrovir. 2012; 28:54–75. [PubMed: 21848364]
143. Wang RYR, Kudryashev M, Li X, Egelman EH, Basler M, et al. De novo protein structure determination from near-atomic-resolution cryo-EM maps. Nat. Methods. 2015; 12:335–38. [PubMed: 25707029]
144. Wang X, Xu F, Liu J, Gao B, Liu Y, et al. Atomic model of rabbit hemorrhagic disease virus by cryo-electron microscopy and crystallography. PLOS Pathog. 2013; 9:e1003132. [PubMed: 23341770]
145. Whitford PC, Ahmed A, Yu Y, Hennelly SP, Tama F, et al. Excited states of ribosome translocation revealed through integrative molecular modeling. PNAS. 2011; 108:18943–48. [PubMed: 22080606]
146. Wickles S, Singharoy A, Andreani J, Seemayer S, Bischoff L, et al. A structural model of the active ribosome-bound membrane protein insertase YidC. eLife. 2014; 3:e03035. [PubMed: 25012291]
147. Wriggers W, Chacón P. Modeling tricks and fitting techniques for multiresolution structures. Structure. 2001; 9:779–88. [PubMed: 11566128]
148. Wright ER, Schooler JB, Ding HJ, Kieffer C, Fillmore C, et al. Electron cryotomography of immature HIV-1 virions reveals the structure of the CA and SP1 Gag shells. EMBO J. 2007; 26:2218–26. [PubMed: 17396149]
149. Wu X, Subramaniam S, Case DA, Wu KW, Brooks BR. Targeted conformational search with map-restrained self-guided Langevin dynamics: application to flexible fitting into electron microscopic density maps. J. Struct. Biol. 2013; 183:429–40. [PubMed: 23876978]
150. Xue Y, Skrynnikov NR. Ensemble MD simulations restrained via crystallographic data: Accurate structure leads to accurate dynamics. Protein Sci. 2014; 23:488–507. [PubMed: 24452989]
151. Yu F, Joshi SM, Ma YM, Kingston RL, Simon MN, Vogt VM. Characterization of Rous sarcoma virus Gag particles assembled in vitro. J. Virol. 2001; 75:2753–64. [PubMed: 11222698]
152. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, et al. Crystal structure of the ribosome at 5.5Å resolution. Science. 2001; 292:883–96. [PubMed: 11283358]
153. Yusupova GZ, Yusupov MM, Cate JHD, Noller HF. The path of messenger RNA through the ribosome. Cell. 2001; 106:233–41. [PubMed: 11511350]
154. Zhang L, Ren G. IPET and FETR: experimental approach for studying molecular structure dynamics by cryo-electron tomography of a single-molecule structure. PLOS ONE. 2012; 7:e30249. [PubMed: 22291925]
155. Zhang X, Zhang L, Tong H, Peng B, Rames MJ, et al. 3D structural fluctuation of IgG1 antibody revealed by individual particle electron tomography. Sci. Rep. 2015; 5:9803. [PubMed: 25940394]
156. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature. 2013; 497:643–6. [PubMed: 23719463]
157. Zhao J, Benlekbir S, Rubinstein JL. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. Nature. 2015; 521:241–45. [PubMed: 25971514]

SUMMARY POINTS

1. Computational hybrid methods feature a unique ability to build missing structural features from experimentally determined structures, as well as integrate data from diverse experimental sources to produce unified, fully atomistic structures of large macromolecular complexes in native-state conformations.
2. There is no one-size-fits-all recipe for applying computational hybrid methods to refine a biomolecular structure. The inherent structural heterogeneity across different macro-molecular complexes presents distinct modeling challenges that must be addressed with respect to the nuances of each system and to the available experimental data.
3. Moving forward, a set of standardized metrics should be developed to objectively characterize the accuracy, validity, and reproducibility of computationally refined structural models.
4. The role of computational hybrid methods in the large-scale structural refinement field will become increasingly prominent as we enter the exciting era of super-resolution experiments, facilitating within the next decade the solution of billion-atom structures and even the atomistic description of the smallest cellular organisms.

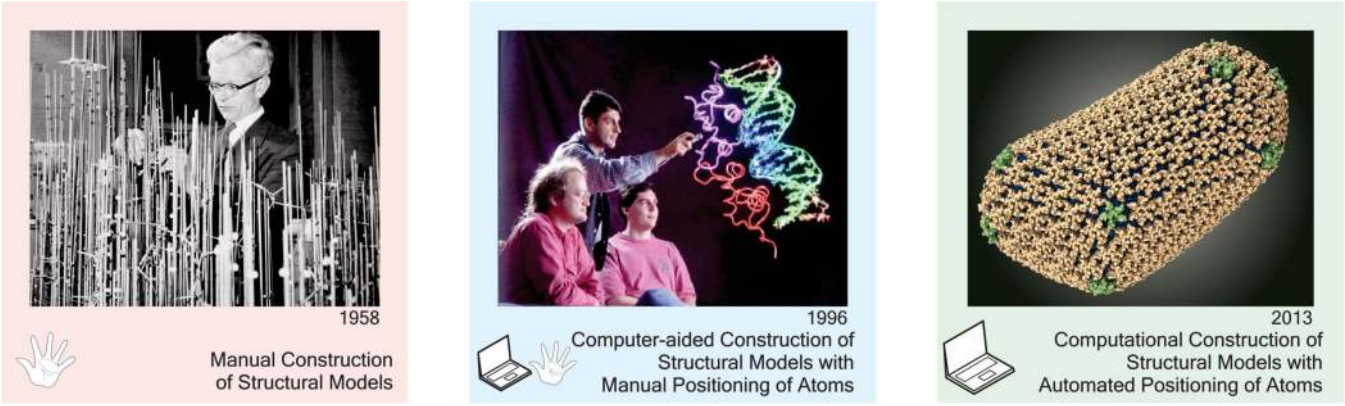

Figure 1.

Evolution of modeling tools used for structure determination. From manual construction of physical models, such as the forest of rods, in the 1950s (left), to computer-aided model building using visualization software in the 1980s and 1990s (middle), to today's state-of-the-art computational structural refinement approaches such as MDFF (right), technological advancements have enabled scientists to solve the atomistic structures of increasingly large biomolecular systems. The image of John Kendrew building the all-atom structure of myoglobin (left) is reproduced here with permission from MRC Laboratory of Molecular Biology, (c) 1958 by MRC Laboratory of Molecular Biology.

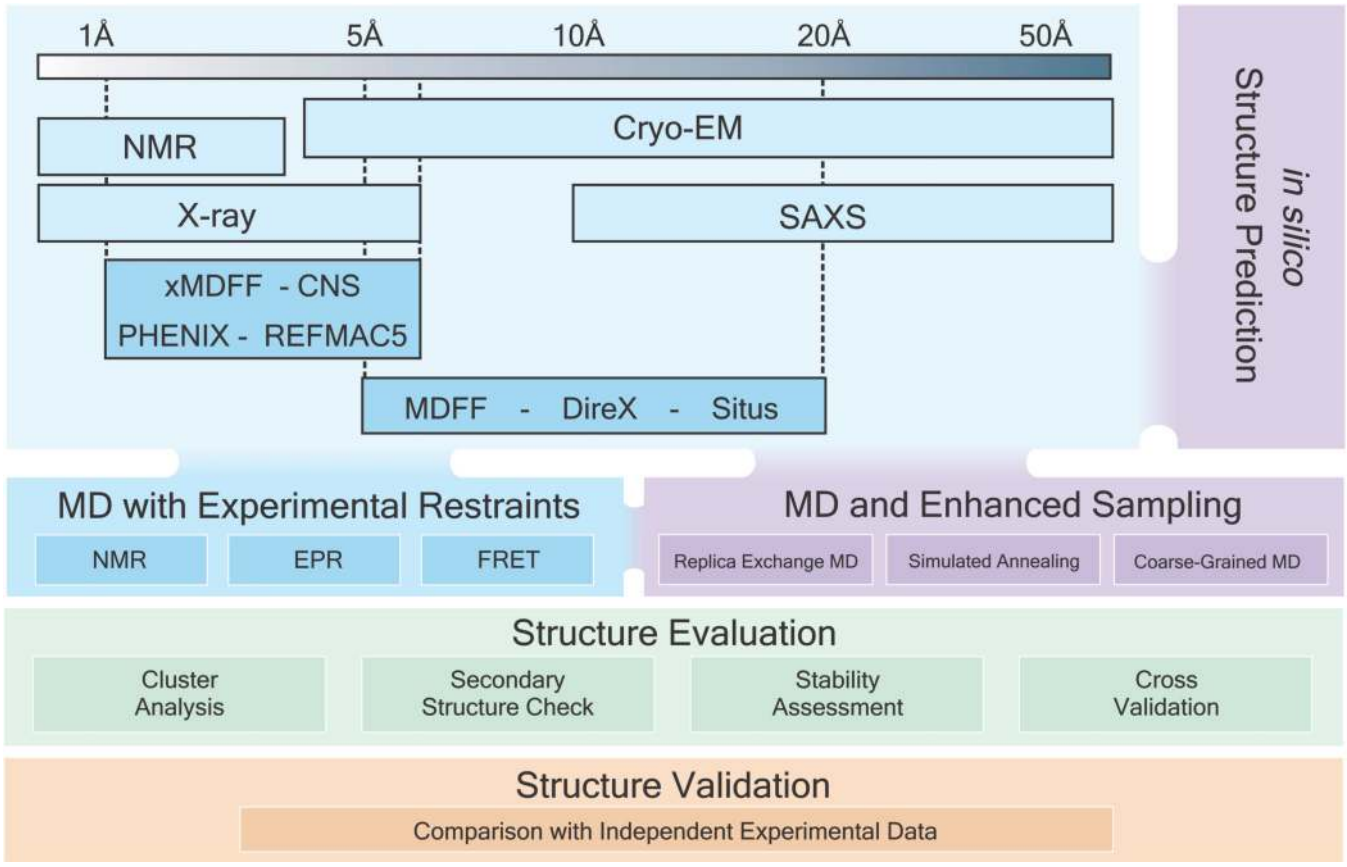

Figure 2.

General strategy for structural refinement of biomolecules using modern hybrid methods. Primary sources of structural information include experimental methods, such as X-ray crystallography, NMR spectroscopy, cryo-EM, and SAXS, as well as in silico structure prediction tools. All-atom structures are generated by integrating experimental data from a range of accessible resolutions using computational approaches, such as those implemented in MDFF and xMDFF. Complete structural models are further refined through MD simulations, employing restraints based on experimental data or enhanced sampling techniques. Final models are evaluated with theoretical checking tools and validated on the basis of data from additional experimental studies.

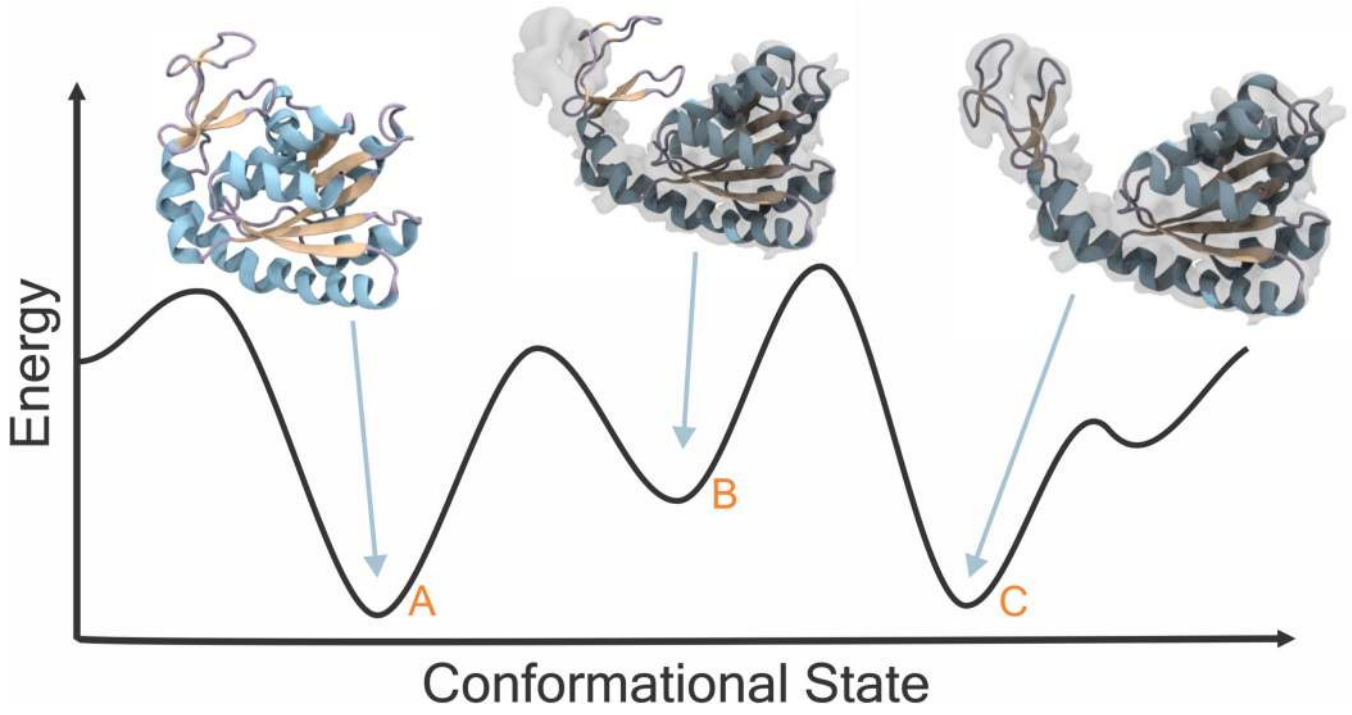

Figure 3.

The energy landscape of a protein in different conformations. During MD simulations, including MDFF, molecules can become trapped in local energy minima representing nonrelevant conformations. For example, the X-ray structure corresponding to energy minimum (*a*) must visit the conformational state corresponding to energy minimum (*b*) before reaching the conformation given by the cryo-EM density, which corresponds to energy minimum (*c*). In MD simulations, a molecule may spend a long time sampling intermediate energy minima like (*b*) before overcoming the barriers to arrive at relevant conformations. The sampling can be performed more efficiently by employing enhanced sampling methods, such as temperature-accelerated MDFF (138), that facilitate the crossing of energy barriers.

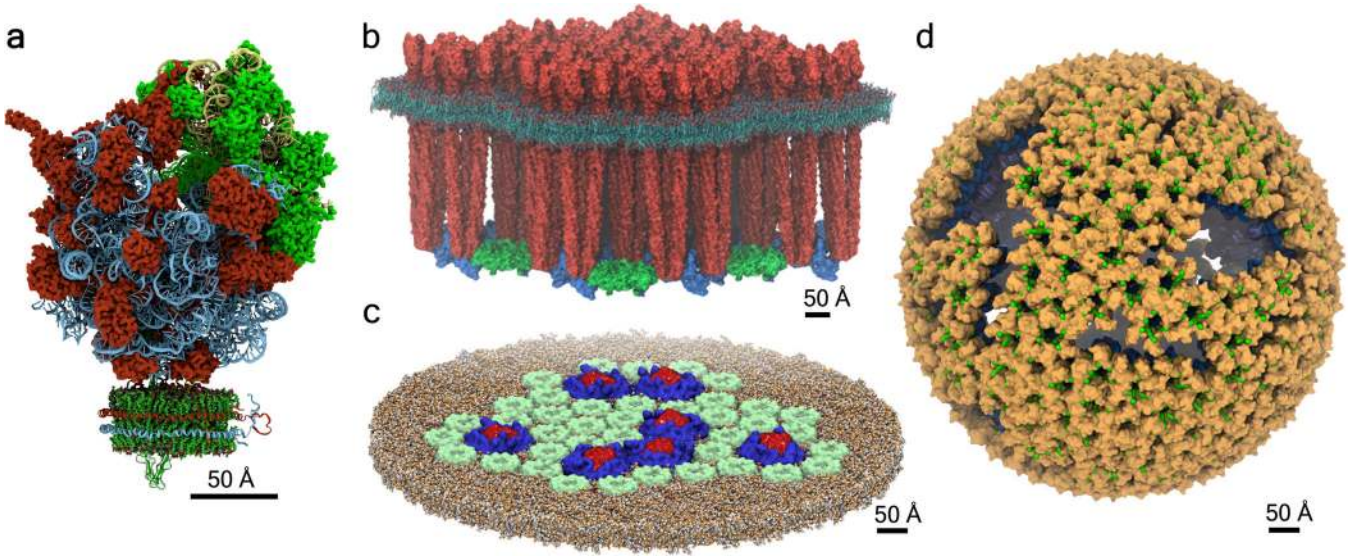

Figure 4.

Complexity and structural distinctiveness of biomolecular systems. (*a*) The ribosome with SecY translocon lipoprotein complex, (*b*) bacterial chemosensory array, and (*c*) lamellar chromatophore patch represent examples of complex systems composed of multiple classes of biomolecular components, including various proteins, RNA, and lipid membranes. Contrarily, the immature capsid of Rous sarcoma virus (*d*) is made of multiple copies of a single protein. Although the structure of the bacterial chemosensory array is characterized by a regular, repeating arrangement, the ribosome is inherently asymmetric, the chromatophore is based on a heterogenous composition of subcomponents, and the virus capsid exhibits an irregular spherical shell lattice.

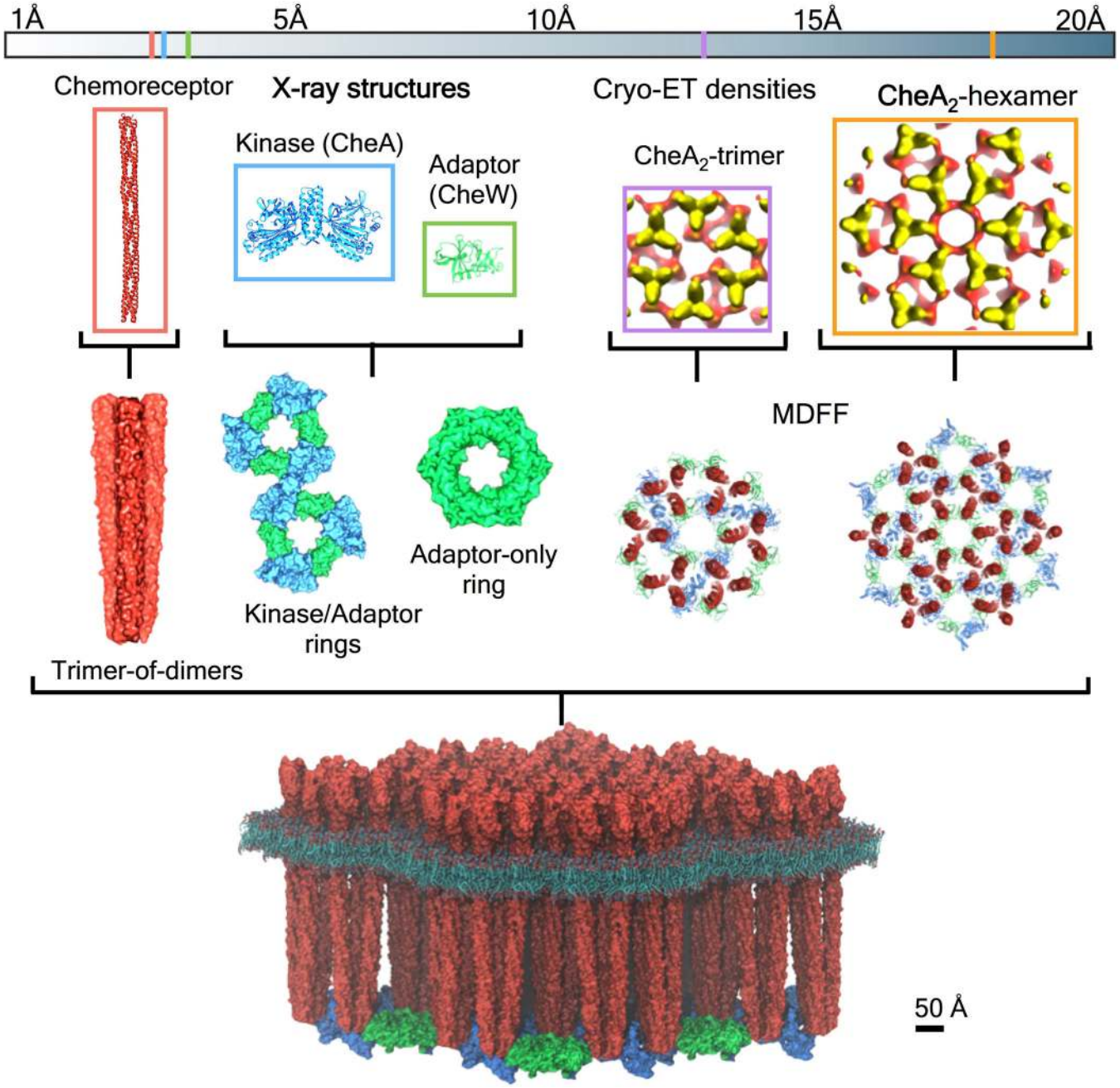

Figure 5.

Construction and refinement of an atomistic model of the bacterial chemosensory array. X-ray structures of the kinase CheA (17), adaptor protein CheW (78), and chemoreceptor (102) from the thermophilic bacterium *Thermotoga maritima* were used to model key array substructures, which were arranged according to cryo-ET densities to produce models of the extended array architecture, namely the CheA2-trimer and CheA2-hexamer assemblies. Subsequently, molecular dynamics flexible fitting (MDFF) simulations with symmetry restraints (28) were carried out to refine the structures of the component models to their array-bound conformations (27).

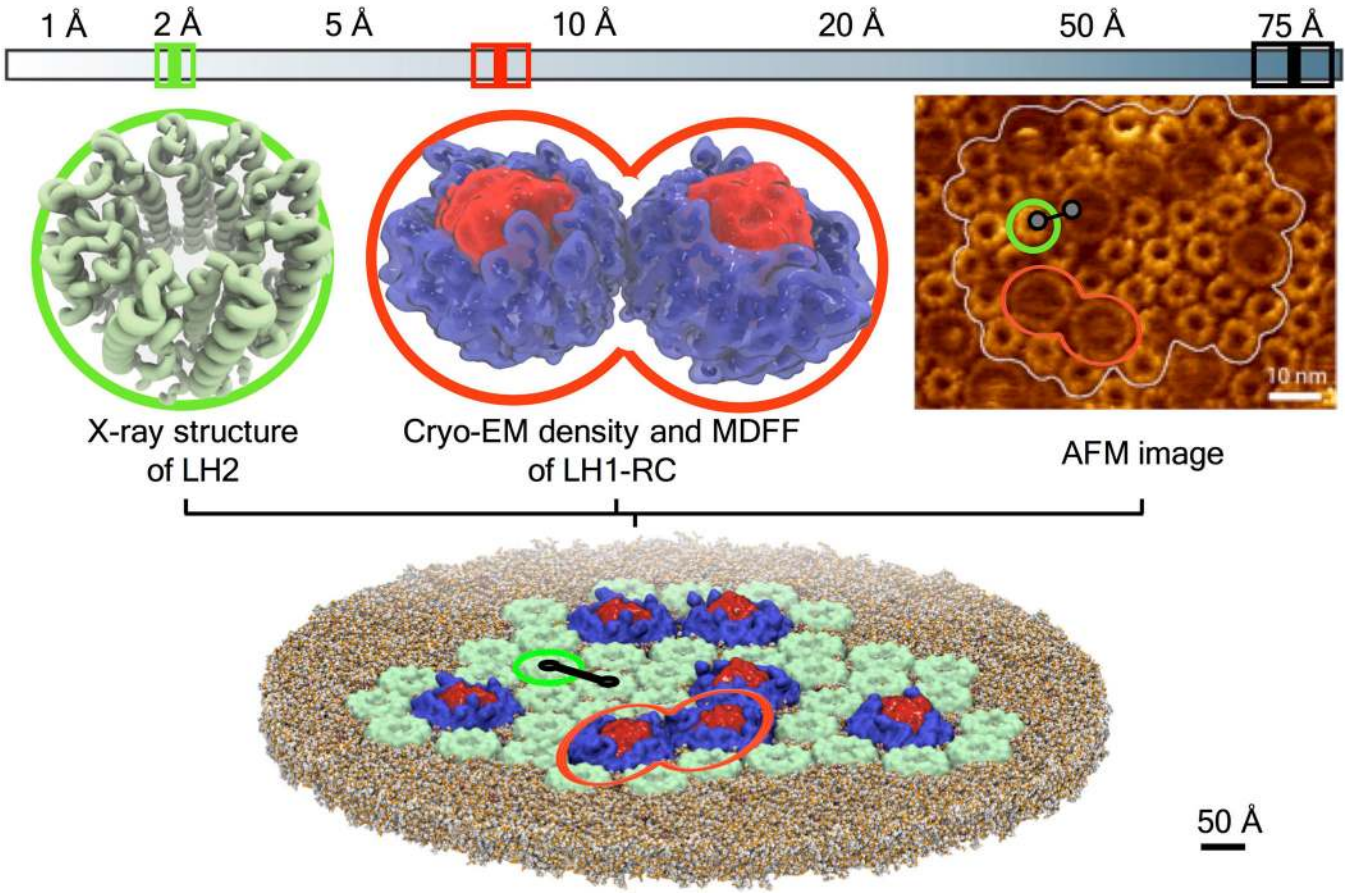

Figure 6.

Construction of an atomic-level model of the lamellar chromatophore patch in *Rhodospirillum photometricum*. The construction combined various experimental data of 2-Å (X-ray crystallography) to 75-Å (AFM) resolution, namely, X-ray structures of the light-harvesting protein LH2 (100), and models of the LH1 proteins derived through in silico structure prediction along with MDFF fitting to cryo-EM density maps. The stated structures were assembled according to AFM images within a realistic lipid membrane environment and refined through MD simulations (29). The AFM image shown was adapted from Reference 29 with permission.