

1 **CryoEM-based Hybrid Modeling Approaches for Structure Determination**

2 C. Keith Cassidy¹, Benjamin A. Himes², Zaida Luthey-Schulten³ and Peijun Zhang^{2,4,5*}

3
4 ¹Department of Physics, Beckman Institute, University of Illinois at Urbana-Champaign,
5 Urbana, IL, USA

6 ²Department of Structural Biology, University of Pittsburgh School of Medicine,
7 Pittsburgh, PA, USA

8 ³Department of Chemistry, Center for the Physics of Living Cells, University of Illinois
9 at Urbana-Champaign, Urbana, IL, USA

10 ⁴Department of Structural Biology, University of Oxford, Oxford, UK

11 ⁵Electron Bio-Imaging Centre, Diamond Light Sources, Harwell Science and Innovation
12 Campus, Didcot OX11 0DE, UK.

13
14 *Correspondence to: **peijun@strubi.ox.ac.uk** (P.Z.)

1 **Abstract**

2 Recent advances in cryo-electron microscopy (cryoEM) have dramatically improved the
3 resolutions at which vitrified biological specimens can be studied, revealing new
4 structural and mechanistic insights over a broad range of spatial scales. Bolstered by
5 these advances, much effort has been directed towards the development of hybrid
6 modeling methodologies for the construction and refinement of high-fidelity atomistic
7 models from cryoEM data. In this brief review, we will survey the key elements of
8 cryoEM-based hybrid modeling, providing an overview of available computational tools
9 and strategies as well as several recent applications.

10

11 **Highlights**

- 12 • CryoEM now provides near-native structural information for a broad
13 spectrum of biomolecules and their complexes.
- 14 • Hybrid modeling approaches enable the construction, refinement, and
15 validation of atomic models using cryoEM data over a wide range of
16 resolutions.
- 17 • CryoEM-based hybrid modeling is a powerful and rapidly evolving approach
18 for gaining unique, atomistic insight into biomolecular structure and
19 function.

20

1 **Introduction**

2 The application of cryoEM to biomolecular structure determination has become a
3 widespread and lucrative approach. Enhanced by new cameras that directly detect
4 electrons [1], improved microscopes with more stable optics [2], and advances in image
5 processing software [3], cryoEM is now used to determine high-resolution
6 macromolecular structures of broad biological significance in their native and functional
7 contexts [4••]. At the same time, the applicability of cryoEM to biological systems
8 spanning a wide range of length scales has helped to give rise to a more cohesive picture
9 of microbiological processes, one that now ranges from the single molecule all the way to
10 the whole cell [5].

11 CryoEM may be used in several modalities depending on the nature of the specimen
12 under investigation; including its stability, size, and conformational flexibility. At low
13 resolution (20-100 Å), three-dimensional (3D) images called tomograms may be
14 reconstructed using cryo-electron tomography (cryoET) to reveal structures of
15 pleomorphic objects, including small bacterial cells, isolated viruses, and elements of
16 cellular ultrastructure such as mitochondria and other organelles. In cases where multiple
17 instances of a molecule are present, sub-tomograms may be computationally extracted
18 from these reconstructions for further alignment and classification (cryoSTAC) [6]. Upon
19 averaging, these aligned sub-populations lead to intermediate resolution (5-20 Å)
20 structures of biological assemblies such as virus capsids and microtubules as well as
21 molecular motors and other macromolecular machines. Particularly exciting are
22 biomolecular structures solved to high resolution (2-5 Å) via single-particle-analysis
23 (SPA) [7•] or electron crystallography [8], which now routinely rival the resolutions
24 obtained through X-ray crystallography (XRC) and NMR spectroscopy [9•]. Moreover,
25 SPA and cryoSTAC, in addition to their increased resolution capabilities, greatly relieve
26 constraints on specimen size and heterogeneity imposed by more traditional high-
27 resolution techniques that rely on ensemble-based measurements, permitting the study of
28 many previously intractable systems, including membrane-bound proteins and large,
29 dynamic biomolecular complexes [10].

Accordingly, hybrid modeling methodologies, which seek to extend the chemical interpretability of cryoEM data through the derivation of high-fidelity atomistic models, have become an area of fervent development [11,12]. How suitable a particular hybrid modeling protocol is for solving a given structure, depends largely on the quality and resolution of the cryoEM data used as input, assumed here to take the form of a cryoEM density map: a 3D image in which the Coulomb potential distribution of the specimen is represented by linearly proportional gray-scale intensities. The Electron Microscopy Data Bank (EMDB) [13•], created to publicly archive cryoEM maps, now boasts over 5000 entries, with new maps being deposited at an ever-increasing rate. In the following sections, we will survey computational strategies for constructing, refining, and validating atomic models based on cryoEM maps, highlighting specific tools and procedures for both the intermediate-resolution and high-resolution regimes (summarized in Figure 1 and Table 1). It should be noted at the outset of our discussion that the division of hybrid modeling tools into the specific resolution ranges presented here is, of course, not strictly inherent to the tools themselves, but rather should be thought of as a fuzzy demarcation centered on resolutions where the tool may be applied with greatest confidence.

Hybrid Modeling at Intermediate Resolution

At intermediate resolutions (5-20 Å), which are now routinely reached by both SPA and cryoSTAC, hybrid modeling strategies range from the rough placement of rigid models within a cryoEM map to their all-atom, density-driven conformational refinement. As the information present at these resolutions is, in general, not sufficient to unambiguously discern a structure's full topology or, in lower resolution cases, even its tertiary organization, intermediate-resolution approaches typically require the user to provide a complete, atomistic model as input. Molecular coordinates may be taken directly from one or more of the 100,000-plus structures in the Protein Data Bank (PDB) [14] or derived from the solved structure of a suitably close homologue using comparative modeling suites [15,16] or web-based homology tools [17–20]. A particularly comprehensive suite of tools for comparative modeling, Modeller [15] takes as input a sequence alignment and automatically constructs, taking into account multiple structural

1 templates as well as user-provided structural restraints, an ensemble of complete atomic
2 models ranked according to DOPE score [21]. For proteins with less than 150 residues, a
3 model may be constructed using *ab initio* structure prediction [22–24], although it is
4 currently difficult to ensure model reliability [25]. When possible, the prudent modeler
5 will compare multiple preliminary models constructed using several different tools. For
6 the purpose of model selection, a number of scoring functions are available to quantify
7 the stereochemical quality and energetics of a given model [26–30], the most popular of
8 which, MolProbity [26], is discussed in more detail below.

9 Whether the goal is to assemble known components into an unknown complex (i.e.,
10 quaternary assembly) or to derive a model of a new functional state, the initial model(s)
11 must be fit to the map to facilitate direct comparison. Fitting strategies fall into two
12 classes: rigid fitting (commonly called “docking”) and flexible fitting. Rigid-fitting
13 algorithms search for the optimal positioning of one or more high-resolution models
14 within a map without allowing changes in model conformation. Many programs exist for
15 single and multi-body rigid fitting, which utilize different searching and scoring strategies
16 to increase efficiency or accuracy at different map resolutions [31–36]. To reduce the
17 complexity of the fitting problem, cryoEM maps are often segmented into separate
18 densities corresponding to individual biomolecules or molecular domains. UCSF-
19 Chimera [37] a feature-rich suite for the visualization and general manipulation of
20 cryoEM data, enables both rapid rigid fitting through its “Fit in Map” routine as well as
21 high-precision segmentation using the Segger-based [38,39] “Segment Map” routine.

22
23 While useful for the analysis of low resolution maps or construction of preliminary
24 models, rigid fitting is often not sufficient for obtaining the optimal fit to a map owing to
25 the inherently dynamic nature of many biomolecules and their complexes. To improve
26 model-map overlap, flexible-fitting methods allow a rigidly docked model to be
27 plastically deformed while (ideally) maintaining proper stereochemistry. Several methods
28 have been developed that utilize reduced representations of biomolecular mobility or
29 structure to assess flexibility and increase fitting efficiency, including those based on
30 elastic network models [40], normal modes analysis [41–43], constrained geometries
31 [44], and Bayesian statistics [45]. More accurate, albeit computationally more expensive,

1 methods employ detailed force fields combined with Monte Carlo (MC) [46,47] or
2 molecular dynamics (MD) [48,49•,50,51] simulation. Of these, Molecular Dynamics
3 Flexible Fitting (MDFF) [48,52] has become one of the most widely used, assisting in the
4 structural determination of the HIV-1 virus capsid [53,54], Ribosome [55–57], and many
5 others [58–61,62••,63••].

6 MDFF approaches the flexible fitting problem through the addition of a density-derived
7 term to the MD potential energy function, thereby generating forces that explicitly drive
8 the conformations of the substrate biomolecules towards regions of high density within
9 the map while taking into account all-atom electrostatic and solvent effects [52].

10 Furthermore, MDFF's tight integration with the molecular visualization program VMD
11 [64] and high-performance MD package NAMD [65] provides increased functionality,
12 including tools for setting up MDFF simulations [52,66] and efficiently analyzing
13 quality-of-fit [67] as well as combining structural refinement with enhanced sampling
14 methods [68,69••] or additional restraints based on symmetry [70] or complimentary
15 experimental data [71•,72]. Importantly, NAMD's scalability to very large system sizes,
16 having already enabled the application of MDFF for the determination of structures tens-
17 of-millions of atoms in size [11,53], provides a promising platform for the future
18 development of methods to tackle even larger sub-cellular complexes [73–75]. Figure 2
19 highlights a recent application in which Cassidy *et al.* used MDFF and other hybrid
20 modeling tools to derive an atomic model of the bacterial chemosensory array 20 million
21 atoms in size based on an 11.7 Å map determined via cryoSTAC [63••]. All-atom MD
22 simulations with NAMD were subsequently performed to further refine a portion of the
23 model and study its dynamical behavior, allowing for the identification of novel
24 interfaces and tertiary motions in a signaling protein ubiquitous throughout bacterial
25 motility.

26 **Hybrid Modeling at High Resolution**

27 At high resolution (2-5 Å), the degree of structural detail present within a given map
28 typically permits *de novo* modeling techniques to construct a primary sequence directly
29 into the density without the use of a structural template [76]. If the resolution is

1 sufficiently high, usually better than 3.5 Å, structure topology will be unambiguous,
2 secondary structure features will be clearly resolved, and the majority of side chains will
3 be chemically identifiable. In such cases, standard manual-modeling tools from software
4 suites originally developed for XRC such as Coot [77], Phenix [78], and others [79–81]
5 may be utilized to obtain a refined atomic model “from scratch” while also correcting for
6 small discrepancies in model-map overlap [82,83]. Though well established, the use of
7 XRC tools to interpret near-atomic resolution maps can be labor intensive and care must
8 still be taken to avoid misinterpretations of the density.

9 As map resolution decreases below 3.5 Å, precise sequence registration and overall
10 topology become more ambiguous, and *de novo* modeling requires the addition of
11 structural information. If a homologous structure is not known, most *de novo* workflows
12 build up protein structure hierarchically, beginning with the identification of pre-defined
13 secondary structural elements within the density [84–87] and from these constructing an
14 alpha carbon trace using pathwalking [88,89•] or other pattern-recognition strategies
15 [90,91]. Given an alpha carbon trace, a fully atomistic model, including unresolved loop
16 segments and side chains, can be constructed and optimized using routines from the
17 Rosetta software suite [46,92,93,95•]. A broad source of tools for general *de novo*
18 modeling, Rosetta uses an iterative, fragment-assembly approach, choosing small
19 fragments from the PDB based on local sequence homology and employing MC sampling
20 with an empirical force field [30] to conformationally optimize and assemble fragments.

21 In many cases, a mostly complete structural template will be available, which can be
22 conformationally refined using a high-resolution cryoEM map in lieu of large-scale *de*
23 *novo* modeling. The Rosetta approach has been particularly successful in the use of 3.5-5
24 Å and higher resolution cryoEM maps for structural refinement [94,95•,96]. Above the 5
25 Å threshold, however, the structural information within the map is usually not enough to
26 correctly direct fragment positioning for large protein segments [96]. To help relieve this
27 constraint, RosettaES, a sampling protocol based on the above fragment-assembly
28 approach, but using fragment ensembles pruned in a “beam search” fashion, was recently
29 shown to efficiently and automatically construct, without a template, accurate models of
30 missing segments up to 100 residues using high-resolution maps while also allowing for

the modeling of multiple interacting segments [97••]. In addition, improvements to the MDFF protocol, namely Cascade MDFF and Resolution Exchange MDFF (RE-MDFF), have been shown to produce models of comparable quality to Rosetta using sub-5 Å maps [69••], though, it should be noted that methods based on MDFF alone do not currently provide functionality for *de novo* modeling.

In general, when interpreting the structural information present in high-resolution cryoEM maps, a combination of several hybrid modeling tools, each with their unique strengths, will be required to arrive at an intact and well-fitting atomic model. A case in point, Wehmer *et al.* recently used an iterative combination of Rosetta and MDFF (similar to that described in [98]) as well as homology modeling and XRC tools to construct atomic models of the 26S proteasome using maps of four distinct conformers, including the inactive and active complex at 4.1 and 4.5 Å resolution respectively [62••]. In another recent example, shown in Figure 3, Chen *et al.* applied a wide range of hybrid modeling tools and MD simulation to derive atomic models of the AMPA receptor-TARP complex in its closed, active, and desensitized states, revealing neurotransmitter-induced conformational changes in the receptor's ligand-binding and transmembrane regions [99,100••]. Though beyond the scope of this review, these applications also emphasize the importance of advanced computational tools for image processing, which enable the isolation of distinct conformational states within SPA data [101–104].

Model Validation

The final and most important step in any hybrid-modeling pipeline is a critical assessment of the accuracy of the resulting atomic model. Numerous methods have been developed to quantify either the stereochemical quality of a model or its goodness-of-fit to a cryoEM map. MolProbity [26], a widely used tool for judging the integrity of biomolecular structures, provides a detailed breakdown of model quality by comparing certain geometric quantities (e.g., backbone bond lengths and dihedral angles, side chain rotamers, etc.) to corresponding statistics derived from the PDB, allowing the user to diagnose structural “outliers” within a model. Meanwhile, the global and local goodness-of-fit between a model and corresponding density can be measured in real-space by

1 metrics based on Pearson's cross-correlation coefficient (CCC) [105,106] (although, a
2 number of other measures exist that may perform better in certain situations [107,108]) or
3 in reciprocal space using measures such as the integrated Fourier shell correlation (iFSC)
4 or the estimated phase error [96]. Complimentary to the above tools, EMRinger [109•]
5 can be utilized to assess both the stereochemistry and quality-of-fit of high-resolution
6 features in cryoEM maps, in particular, side chain conformation.

7 Regardless of map resolution or the hybrid modeling protocol used, special care must be
8 taken not to over-fit the cryoEM data [110,111]. Over-fitting occurs when the density is
9 allowed to determine features of the model for which an appropriate amount of
10 information is not present, for instance, if the map is insufficiently resolved or if sample
11 heterogeneity (e.g., in a flexible loop or dynamic domain region) has led to local regions
12 of reduced resolution or missing density. Tools such as ResMap [112•] can be used for
13 detecting spatially variable resolution within a map, helping to pinpoint local regions of
14 reduced information that may potentially affect model accuracy [113,114]. Globally,
15 over-fitting must be minimized by reducing the number of refined parameters (e.g., by
16 the use of secondary-structure constraints or selective exclusion of certain atoms) or
17 through the addition of supplemental structural information (e.g. by employing a physics-
18 based force field) [115]. In addition, several cross-validation procedures, in which the
19 original cryoEM data are split into multiple, ideally independent sets that are separately
20 used for model construction/refinement and accuracy assessment, have been proposed to
21 identify over-fitting to high-resolution maps [82,96,116,117•,118].

22 Though, the above validation tools cannot yet attest to the absolute accuracy of a model
23 (i.e., how far it deviates from the "real" biomolecular structure), such considerations can
24 help to evaluate how consistent a model is with other known structures as well as whether
25 or not it represents the cryoEM data well. Towards this end, it is critical that the modeler
26 clearly highlight the degree of confidence with which a feature has been determined, for
27 example, by annotating a single model [119•] or reporting an ensemble of models that
28 capture feature variability [120]. Moreover, the results of multiple fitting or model
29 construction protocols, using different parameters, can be compared to help lend
30 confidence to a given solution [121,122]. Finally, the subsequent application of unbiased

1 MD simulation provides powerful means by which to gauge the stability of certain model
2 features while simultaneously exploring the post-refinement free energy landscape [123].

3 **Outlook**

4 Rapid advances in the high-resolution capabilities of cryoEM have opened the door to the
5 study of previously inaccessible biomolecular systems, ranging from membrane-bound
6 proteins to flexible and heterogeneous macromolecular complexes. At the same time,
7 steady progress at lower to intermediate resolutions continues to produce quality images
8 of numerous biomolecular structures central to diverse microbiological processes. Here,
9 we have surveyed the state-of-the-art in computational methods for constructing, refining,
10 and validating atomic models using intermediate and high-resolution cryoEM maps.
11 Nevertheless, while the development of new and improved methodologies has been key
12 to the success of the hybrid modeling community highlighted in this review, the current
13 state of the field is such that its most pressing issues involve an improvement in the
14 general usability of already existing software, tools, and protocols. In particular, many
15 hybrid modeling tools currently exist in the wild, scattered over different sources and
16 computing platforms with user interfaces and documentation of widely varying quality,
17 reducing their value to non-experts or those not specializing in scientific computing. We
18 suggest that efforts be made to ahistorically centralize hybrid modeling tools and
19 software, including source code, files for conducting clear, positive-control test cases
20 (called unit tests), and documentation detailing basic user-procedures for each tool,
21 thereby greatly reducing their “black box” nature. Apart from increased visibility, such a
22 tool bank would facilitate direct method comparison and protocol customization as well
23 as provide a space for cataloging invaluable community feedback regarding the
24 application of tools to novel biomolecules and extensions to code functionality, helping
25 to ensure that tools can be made compatible with the new trends and standards developed
26 within the rapidly evolving cryoEM community. Moreover, general usability may be
27 increased by minimizing the technical overhead or computational cost required to
28 conduct a given protocol, for instance, through the use of affordable, cloud-based
29 infrastructures [69••], which allow wider community access to specialized computational
30 resources (e.g., clusters with high-memory or GPU-accelerated nodes). Finally, to ensure

1 that tools are used correctly and in a way that can be reproduced, standard validation and
2 annotation procedures should be adopted and, where possible, made a prerequisite for
3 model deposition in public databases [111,124•]. Community-wide challenges issued by
4 the EMDB (<http://challenges.emdatabank.org>) have contributed greatly to a raised
5 awareness of these and other issues [13•] and will become increasingly important as both
6 the cryoEM and hybrid-modeling communities continue to pursue the development of
7 advanced methods for determining the structures of increasingly complex biological
8 assemblies.

9 **Acknowledgements**

10 We are thankful to Yuhang Wang and Emad Tajkhorshid (University of Illinois) for
11 providing the images used in Figure 3. CKC and ZLS gratefully acknowledge support
12 from National Institutes of Health (NIH-2P41GM10460128) and the National Science
13 Foundation (NSF-PHY1430124). BH and PZ acknowledge support from National
14 Institutes of Health (GM085043), and PZ acknowledges support from the Wellcome
15 Trust Investigator Award (206422/Z/17/Z).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Grigorieff N: **Direct detection pays off for electron cryo-microscopy.** *Elife* 2013, **2**:e00573.
2. Elmlund D, Le SN, Elmlund H: **High-resolution cryo-EM: the nuts and bolts.** *Curr Opin Struct Biol* 2017, **46**:1–6.
3. Vinothkumar KR, Henderson R: **Single particle electron cryomicroscopy: trends, issues and future perspective.** *Q Rev Biophys* 2016, **49**:e13.
4. Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JLS, et al.: **Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery.** *Cell* 2016, **165**:1698–1707.
 - Presents a series of ground-breaking cryoEM structures, including the first sub-2 Å resolution structure as well as the first sub-100 kDa structure, which extend the perceived scope of high-resolution cryoEM to even smaller systems with implications for the study of novel drug targets.
5. Oikonomou CM, Jensen GJ: **Cellular Electron Cryotomography: Toward Structural Biology In Situ TEM: transmission electron microscopy.** *Annu Rev Biochem* 2017, **86**:873–896.
6. Wan W, Briggs JAG: **Cryo-Electron Tomography and Subtomogram Averaging.** In *Methods in Enzymology.* . Elsevier Inc.; 2016:329–367.
7. Cheng Y: **Single-particle Cryo-EM at crystallographic resolution.** *Cell* 2015, **161**:450–457.
 - Provides a broad discussion of recent technological advances that have allowed SPA to determine structures to near-atomic resolution.
8. de la Cruz MJ, Hattne J, Shi D, Seidler P, Rodriguez J, Reyes FE, Sawaya MR, Cascio D, Weiss SC, Kim SK, et al.: **Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED.** *Nat Methods* 2017, **14**:399–402.
9. Bartesaghi A, Merk A, Banerjee S, Matthies D, Wu X, Milne JLS, Subramaniam S: **2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor.** *Science* 2015, **348**:1147–51.
 - Reports a 2.2 Å structure of beta-galactosidase in which key active site features, including a bound inhibitor and catalytic water molecules, are clearly resolved, demonstrating that SPA can obtain the level of chemical detail needed for applications such as rational drug design.
10. Fernandez-Leiro R, Scheres SHW: **Unravelling biological macromolecules with cryo-electron microscopy.** *Nature* 2016, **537**:339–346.
11. Goh BC, Hadden JA, Bernardi RC, Singharoy A, McGreevy R, Rudack T, Cassidy CK, Schulten K: **Computational Methodologies for Real-Space Structural Refinement of Large Macromolecular Complexes.** *Annu Rev Biophys* 2016, **45**:253–278.
12. Villa E, Lasker K: **Finding the right fit: Chiseling structures out of cryo-**

- 1 **electron microscopy maps.** *Curr Opin Struct Biol* 2014, **25**:118–125.
- 2 13. Lawson CL, Patwardhan A, Baker ML, Hryc C, Garcia ES, Hudson BP, Lagerstedt
3 I, Ludtke SJ, Pintilie G, Sala R, et al.: **EMDataBank unified data resource for**
4 **3DEM.** *Nucleic Acids Res* 2016, **44**:D396–D403.
- 5 • Provides an overview of the EMDB archive, including new features for map validation
6 as well as EMDB-sponsored community challenges for assessing different cryoEM
7 reconstruction and modeling protocols.
- 8 14. Berman HM, Kleywegt GJ, Nakamura H, Markley JL: **The Protein Data Bank**
9 **archive as an open data resource.** *J Comput Aided Mol Des* 2014, **28**:1009–
10 1014.
- 11 15. Webb B, Sali A: **Protein Structure Modeling with MODELLER.** In *Protein*
12 *Structure Prediction.* . Springer; 2014:1–15.
- 13 16. Song Y, Dimaio F, Wang RYR, Kim D, Miles C, Brunette T, Thompson J, Baker
14 D: **High-resolution comparative modeling with RosettaCM.** *Structure* 2013,
15 **21**:1735–1742.
- 16 17. Arnold K, Bordoli L, Kopp J, Schwede T: **The SWISS-MODEL workspace: a**
17 **web-based environment for protein structure homology modelling.**
18 *Bioinformatics* 2006, **22**:195–201.
- 19 18. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J: **Template-based**
20 **protein structure modeling using the RaptorX web server.** *Nat Protoc* 2012,
21 **7**:1511–22.
- 22 19. Roy A, Kucukural A, Zhang Y: **I-TASSER: a unified platform for automated**
23 **protein structure and function prediction.** *Nat Protoc* 2010, **5**:725–738.
- 24 20. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE: **The Phyre2 web**
25 **portal for protein modeling , prediction and analysis.** *Nat Protoc* 2015, **10**:845–
26 858.
- 27 21. Shen M, Sali A: **Statistical potential for assessment and prediction of protein**
28 **structures.** *Protein Sci* 2006, **15**:2507–2524.
- 29 22. MacCallum JL, Perez A, Dill KA: **Determining protein structures by combining**
30 **semireliable data with atomistic physical models by Bayesian inference.** *Proc*
31 *Natl Acad Sci U S A* 2015, **112**:6985–90.
- 32 23. Bradley P, Misura KMS, Baker D: **Toward high-resolution de novo structure**
33 **prediction for small proteins.** *Science (80-)* 2005, **309**:1868–1871.
- 34 24. Wu S, Skolnick J, Zhang Y: **Ab initio modeling of small proteins by iterative**
35 **TASSER simulations.** *BMC Biol* 2007, **5**:1–10.
- 36 25. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A: **Critical**
37 **assessment of methods of protein structure prediction: Progress and new**
38 **directions in round XI.** *Proteins Struct Funct Bioinforma* 2016, **84**:4–14.
- 39 26. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW,
40 Arendall III WB, Snoeyink J, Richardson JS, et al.: **MolProbity: all-atom**
41 **contacts and structure validation for proteins and nucleic acids.** *Nucleic Acids*
42 *Res* 2007, **35**:W375–W383.
- 43 27. Laskowski RA, MacArthur MW, Moss DS, Thornton JM: **PROCHECK: a**
44 **program to check the stereochemical quality of protein structures.** *J Appl*
45 *Crystallogr* 1993, **26**:283–291.
- 46 28. Uziela K, Wallner B: **ProQ2: estimation of model accuracy implemented in**

Rosetta. *Bioinformatics* 2016, **32**:1411–1413.

29. Wiederstein M, Sippl MJ: **ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins.** *Nucleic Acids Res* 2007, **35**:W407–W410.
30. Alford RF, Leaver-Fay A, Jeliaskov JR, O’Meara MJ, DiMaio FP, Park H, Shapovalov M V, Renfrew PD, Mulligan VK, Kappel K, et al.: **The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design.** *J Chem Theory Comput* 2017, **13**:3031–3048.
31. Wriggers W: **Conventions and workflows for using Situs.** *Acta Crystallogr Sect D Biol Crystallogr* 2012, **68**:344–351.
32. Van Zundert GCP, Melquiond ASJ, Bonvin AMJJ: **Integrative modeling of biomolecular complexes: HADDOCKing with Cryo-electron microscopy data.** *Structure* 2015, **23**:949–960.
33. de Vries SJ, Zacharias M: **ATTRACT-EM: A New Method for the Computational Assembly of Large Molecular Machines Using Cryo-EM Maps.** *PLoS One* 2012, **7**:e49733.
34. de Vries SJ, van Dijk M, Bonvin AMJJ: **The HADDOCK web server for data-driven biomolecular docking.** *Nat Protoc* 2010, **5**:883–897.
35. Esquivel-Rodríguez J, Kihara D: **Fitting multimeric protein complexes into electron microscopy maps using 3D zernike descriptors.** *J Phys Chem B* 2012, **116**:6854–6861.
36. Lasker K, Topf M, Sali A, Wolfson HJ: **Inferential optimization for simultaneous fitting of multiple components into a CryoEM map of their assembly.** *J Mol Biol* 2009, **388**:180–194.
37. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE: **UCSF Chimera—A Visualization System for Exploratory Research and Analysis.** *J Comput Chem* 2004, **25**:1605–1612.
38. Pintilie G, Chiu W: **Comparison of Segger and other methods for segmentation and rigid-body docking of molecular components in Cryo-EM density maps.** *Biopolymers* 2012, **97**:742–760.
39. Pintilie GD, Zhang J, Goddard TD, Chiu W, Gossard DC: **Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions.** *J Struct Biol* 2010, **170**:427–438.
40. Schröder GF, Brunger AT, Levitt M: **Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution.** *Structure* 2007, **15**:1630–1641.
41. Lopéz-Blanco JR, Chacón P: **IMODFIT: Efficient and robust flexible fitting based on vibrational analysis in internal coordinates.** *J Struct Biol* 2013, **184**:261–270.
42. Tama F, Miyashita O, Brooks CL: **Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis.** *J Mol Biol* 2004, **337**:985–999.
43. Suhre K, Navaza J, Sanejouand Y-H: **NORMA: a tool for flexible fitting of high-resolution protein structures into low-resolution electron-microscopy-derived density maps.** *Acta Crystallogr Sect D Biol Crystallogr* 2006, **62**:1098–1100.

- 1 44. Jolley CC, Wells SA, Fromme P, Thorpe MF: **Fitting low-resolution cryo-EM**
2 **maps of proteins using constrained geometric simulations.** *Biophys J* 2008,
3 **94**:1613–1621.
- 4 45. Habeck M: **Bayesian Modeling of Biomolecular Assemblies with Cryo-EM**
5 **Maps.** *Front Mol Biosci* 2017, **4**:1416.
- 6 46. Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K,
7 Renfrew PD, Smith CA, Sheffler W, et al.: **Rosetta3: An object-oriented**
8 **software suite for the simulation and design of macromolecules.** *Methods*
9 *Enzymol* 2011, **487**:545–574.
- 10 47. Lindert S, Alexander N, Wötzel N, Karakaş M, Stewart PL, Meiler J: **EM-Fold:**
11 **De novo atomic-detail protein structure determination from medium-**
12 **resolution density maps.** *Structure* 2012, **20**:464–478.
- 13 48. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K: **Flexible fitting of atomic**
14 **structures into electron microscopy maps using molecular dynamics.** *Structure*
15 2008, **16**:673–683.
- 16 49. Joseph AP, Malhotra S, Burnley T, Wood C, Clare DK, Winn M, Topf M:
17 **Refinement of atomic models in high resolution EM reconstructions using**
18 **Flex-EM and local assessment.** *Methods* 2016, **100**:42–49.
- 19 • Demonstrates an extension of the Flex-EM flexible fitting method for structure
20 refinement using high resolution cryoEM maps.
- 21 50. Orzechowski M, Tama F: **Flexible Fitting of High-Resolution X-Ray Structures**
22 **into Cryoelectron Microscopy Maps Using Biased Molecular Dynamics**
23 **Simulations.** *Biophys J* 2008, **95**:5692–5705.
- 24 51. Miyashita O, Kobayashi C, Mori T, Sugita Y, Tama F: **Flexible fitting to cryo-**
25 **EM density map using ensemble molecular dynamics simulations.** *J Comput*
26 *Chem* 2017, **38**:1447–1461.
- 27 52. McGreevy R, Teo I, Singharoy A, Schulten K: **Advances in the molecular**
28 **dynamics flexible fitting method for cryo-EM modeling.** *Methods* 2016,
29 **100**:50–60.
- 30 53. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, Ahn J, Gronenborn
31 AM, Schulten K, Aiken C, et al.: **Mature HIV-1 capsid structure by cryo-**
32 **electron microscopy and all-atom molecular dynamics.** *Nature* 2013, **497**:643–
33 646.
- 34 54. Schur FKM, Hagen WJH, Rumlová M, Ruml T, Müller B, Kräusslich H-G, Briggs
35 JAG: **Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å**
36 **resolution.** *Nature* 2015, **517**:505–508.
- 37 55. Villa E, Sengupta J, G.~Trabuco L, LeBarron J, T.~Baxter W, R.~Shaikh T, A.
38 Grassucci R, Nissen P, Ehrenberg M, Schulten K, et al.: **Ribosome-induced**
39 **Changes in Elongation Factor Tu Conformation Control GTP Hydrolysis.**
40 *PNAS* 2009, **106**:1063–1068.
- 41 56. Frauenfeld J, Gumbart J, Sluis EO Van Der, Funes S, Gartmann M, Beatrix B,
42 Mielke T, Berninghausen O, Becker T, Schulten K, et al.: **Cryo-EM structure of**
43 **the ribosome-SecYE complex in the membrane environment.** *Nat Struct Mol*
44 *Biol* 2011, **18**:614–21.
- 45 57. Wickles S, Singharoy A, Andreani J, Seemayer S, Bischoff L, Berninghausen O,
46 Soeding J, Schulten K, van der Sluis EO, Beckmann R: **A structural model of the**

- active ribosome-bound membrane protein insertase YidC. *Elife* 2014, 3:e03035.
58. Khoshouei M, Radjainia M, Baumeister W, Danev R: **Cryo-EM structure of haemoglobin at 3.2 Å determined with the Volta phase plate.** *Nat Commun* 2017, 8:16099.
59. Zhao J, Benlekbir S, Rubinstein J: **Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase.** *Nature* 2015, 521:241–245.
60. Liu J-J, Niu C-Y, Wu Y, Tan D, Wang Y, Ye M-D, Liu Y, Zhao W, Zhou K, Liu Q-S, et al.: **CryoEM structure of yeast cytoplasmic exosome complex.** *Cell Res* 2016, 26:822.
61. He Y, Yan C, Fang J, Inouye C, Tjian R, Ivanov I, Nogales E: **Near-atomic resolution visualization of human transcription promoter opening.** *Nature* 2016, 533:359–365.
62. Wehmer M, Rudack T, Beck F, Aufderheide A, Pfeifer G, Plitzko JM, Förster F, Schulten K, Baumeister W, Sakata E: **Structural insights into the functional cycle of the ATPase module of the 26S proteasome.** *Proc Natl Acad Sci* 2017.
- Reports the use of an iterative combination of MDFF and Rosetta as well as XRC refinement tools to construct atomic models of the 26S proteasome using cryoEM maps of several distinct conformers ranging from ~4-8 Å resolution. The resulting models enable functional insight into the process of protein degradation.
63. Cassidy CK, Himes BA, Alvarez FJ, Ma J, Zhao G, Perilla JR, Schulten K, Zhang P: **CryoEM and Computer Simulations Reveal a Novel Kinase Conformational Switch in Bacterial Chemotaxis Signaling.** *Elife* 2015, 4:e08419.
- Reports the first atomic model of the extended bacterial chemosensory array, derived using cryoSTAC and MDFF-based hybrid modeling, as well as additional large-scale MD simulations investigating array dynamics.
64. Humphrey W, Dalke A, Schulten K: **VMD: visual molecular dynamics.** *J Mol Graph* 1996, 14:33–38.
65. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K: **Scalable molecular dynamics with NAMD.** *J Comput Chem* 2005, 26:1781–1802.
66. Ribeiro J V, Bernardi RC, Rudack T, Stone JE, Phillips JC, Freddolino PL, Schulten K: **QwikMD — Integrative Molecular Dynamics Toolkit for Novices and Experts.** *Sci Rep* 2016, 6:26536.
67. 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, 169:265–283.
68. Bernardi RC, Melo MCR, Schulten K: **Enhanced sampling techniques in molecular dynamics simulations of biological systems.** *Biochim Biophys Acta - Gen Subj* 2015, 1850:872–877.
69. Singharoy A, Teo I, McGreevy R, Stone JE, Zhao J, Schulten K: **Molecular dynamics-based refinement and validation for sub-5 Å cryo-electron microscopy maps.** *Elife* 2016, 5:1–33.
- Presents improvements to the MDFF protocol that extend its applicability for structure refinement to high resolution cryoEM maps. In addition, several map-model validation

criteria unique to MD-based refinement methods are discussed as well as new cloud-computing MDFF capabilities.

70. Chan K-Y, Gumbart J, McGreevy R, M.~Watermeyer J, Sewell B ~Trevo., Schulten K: **Symmetry-restrained flexible fitting for symmetric EM maps.** *STR* 2011, **19**:1211–1218.

71. Perilla JR, Zhao G, Lu M, Ning J, Hou G, Byeon IL, Gronenborn AM, Polenova T, Zhang P: **CryoEM Structure Refinement by Integrating NMR Chemical Shifts with Molecular Dynamics Simulations.** *J Phys Chem B* 2017, **121**:3853–3863.

• Presents an approach for biasing MD simulations using NMR chemical shift information and demonstrates its use for the further refinement of atomic models based on cryoEM data.

72. Shen R, Han W, Fiorin G, Islam SM, Schulten K, Roux B: **Structural refinement of proteins by restrained molecular dynamics simulations with non-interacting molecular fragments.** *PLoS Comput Biol* 2015, **11**:e1004368.

73. Grange M, Vasishtan D, Grünewald K: **Cellular electron cryo tomography and in situ sub-volume averaging reveal the context of microtubule-based processes.** *J Struct Biol* 2017, **197**:181–190.

74. von Appen A, Kosinski J, Sparks L, Ori A, DiGuilio AL, Vollmer B, Mackmull M-T, Banterle N, Parca L, Kastiris P, et al.: **In situ structural analysis of the human nuclear pore complex.** *Nature* 2015, **526**:140–143.

75. Hu B, Lara-Tejero M, Kong Q, Galán JE, Liu J: **In situ molecular architecture of the Salmonella type III secretion machine.** *Cell* 2017, **168**:1065–1074.

76. DiMaio F, Chiu W: **Tools for Model Building and Optimization into Near-Atomic Resolution Electron Cryo-Microscopy Density Maps.** In *Methods in enzymology*. . Elsevier Inc.; 2016:255–276.

77. Emsley P, Cowtan K: **Coot: model-building tools for molecular graphics.** *Acta Crystallogr Sect D Biol Crystallogr* 2004, **60**:2126–2132.

78. Adams PD, Afonine P V, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al.: **PHENIX: a comprehensive Python-based system for macromolecular structure solution.** *Acta Crystallogr Sect D Biol Crystallogr* 2010, **66**:213–221.

79. Cowtan K: **The Buccaneer software for automated model building.** *Acta Crystallogr Sect D Biol Crystallogr* 2006, **62**:1002–1011.

80. Langer GG, Cohen SX, Lamzin VS, Perrakis A: **Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7.** *Nat Protoc* 2008, **3**:1171.

81. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA: **REFMAC5 for the refinement of macromolecular crystal structures.** *Acta Crystallogr Sect D Biol Crystallogr* 2011, **67**:355–367.

82. Brown A, Long F, Nicholls RA, Toots J, Emsley P, Murshudov G: **Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions.** *Acta Crystallogr Sect D Biol Crystallogr* 2015, **71**:136–153.

83. Afonine P V, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW,

- Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD: **Towards automated crystallographic structure refinement with phenix. refine.** *Acta Crystallogr Sect D Biol Crystallogr* 2012, **68**:352–367.
84. Rusu M, Wriggers W: **Evolutionary bidirectional expansion for the tracing of alpha helices in cryo-electron microscopy reconstructions.** *J Struct Biol* 2012, **177**:410–419.
85. Yu Z, Bajaj C: **Computational approaches for automatic structural analysis of large biomolecular complexes.** *IEEE/ACM Trans Comput Biol Bioinforma* 2008, **5**:568–582.
86. Baker ML, Ju T, Chiu W: **Identification of Secondary Structure Elements in Intermediate-Resolution Density Maps.** *Structure* 2007, **15**:7–19.
87. Si D, He J: **Beta-sheet Detection and Representation from Medium Resolution Cryo-EM Density Maps.** In *Proceedings of the International Conference on Bioinformatics*. . 2007:764–770.
88. Baker ML, Abeysinghe SS, Schuh S, Coleman RA, Abrams A, Marsh MP, Hryc CF, Ruths T, Chiu W, Ju T: **Modeling protein structure at near atomic resolutions with Gorgon.** *J Struct Biol* 2011, **174**:360–373.
89. Chen M, Baldwin PR, Ludtke SJ, Baker ML: **De Novo modeling in cryo-EM density maps with Pathwalking.** *J Struct Biol* 2016, **196**:289–298.
- Reports enhancements to the Pathwalking approach, including improved automation and SSE detection, for the *de novo* construction of backbone traces into high resolution cryoEM maps.
90. Ioerger TR, Sacchettini JC: **Automatic modeling of protein backbones in electron-density maps via prediction of Ca coordinates.** *Acta Crystallogr Sect D Biol Crystallogr* 2002, **58**:2043–2054.
91. Zhou N, Wang H, Wang J: **EMBuilder: A Template Matching-based Automatic Model-building Program for High-resolution Cryo-Electron Microscopy Maps.** *Sci Rep* 2017, **7**:2664.
92. DiMaio F, Kondrashov DA, Bitto E, Soni A, Bingman CA, Phillips Jr GN, Shavlik JW: **Creating protein models from electron-density maps using particle-filtering methods.** *Bioinformatics* 2007, **23**:2851–2858.
93. Wang RY, Kudryashev M, Li X, Egelman EH, Basler M, Cheng Y, Baker D, DiMaio F: **De novo protein structure determination from near-atomic-resolution cryo-EM maps.** *Nat Methods* 2015, **12**:335–8.
94. DiMaio F, Tyka MD, Baker ML, Chiu W, Baker D: **Refinement of protein structures into low-resolution density maps using rosetta.** *J Mol Biol* 2009, **392**:181–190.
95. Wang RY-R, Song Y, Barad BA, Cheng Y, Fraser JS, DiMaio F: **Automated structure refinement of macromolecular assemblies from cryo-EM maps using Rosetta.** *Elife* 2016, **5**:e17219.
- Reports an automated approach for structure refinement based on previously published Rosetta-based protocols (See also [96]) as well as a number of improvements for better interpretation of 3-5 Å cryoEM maps. The efficacy of the approach is demonstrated using several diverse biomolecular complexes.
96. DiMaio F, Song Y, Li X, Brunner MJ, Xu C, Conticello V, Egelman E, Marlovits TC, Cheng Y, Baker D: **Atomic-accuracy models from 4.5-Å cryo-electron**

- 1 **microscopy data with density-guided iterative local refinement.** *Nat Methods*
2 2015, **12**:361–365.
- 3 97. Frenz B, Walls AC, Egelman EH, Veisler D, DiMaio F: **RosettaES: a sampling**
4 **strategy enabling automated interpretation of difficult cryo-EM maps.** *Nat*
5 *Methods* 2017, **14**:797–800.
- 6 •• Presents RosettaES, an automated protocol for *de novo* construction of large protein
7 segments into 3-5 Å cryoEM maps. Using several benchmark systems, RosettaES was
8 shown on to produce accurate models for missing segments up to 100 residues in length
9 and also permitted the modeling of multiple, interacting segments.
- 10 98. Lindert S, McCammon JA: **Improved cryoEM-guided iterative molecular**
11 **dynamics-rosetta protein structure refinement protocol for high precision**
12 **protein structure prediction.** *J Chem Theory Comput* 2015, **11**:1337–1346.
- 13 99. Zhao Y, Chen S, Yoshioka C, Bacongus I, Gouaux E: **Architecture of fully**
14 **occupied GluA2 AMPA receptor–TARP complex elucidated by cryo-EM.**
15 *Nature* 2016, **536**:108–111.
- 16 100. Chen S, Zhao Y, Wang Y, Shekhar M, Tajkhorshid E, Gouaux E, Chen S, Zhao Y,
17 Wang Y, Shekhar M, et al.: **Activation and Desensitization Mechanism of**
18 **AMPA Article Activation and Desensitization Mechanism of AMPA**
19 **Receptor-TARP Complex by Cryo-EM.** *Cell* 2017, **170**:1–13.
- 20 •• Reports the use of a wide range of hybrid modeling tools, including MDFF and XRC
21 tools, to construct atomic models of the AMPA-receptor-TARP complex in its activated
22 and desensitized states using high-to-intermediate resolution cryoEM maps. The resulting
23 models provide residue-specific insight into important neurotransmitter-induced
24 conformational changes.
- 25 101. Scheres SHW: **RELION: implementation of a Bayesian approach to cryo-EM**
26 **structure determination.** *J Struct Biol* 2012, **180**:519–530.
- 27 102. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA: **cryoSPARC: algorithms for**
28 **rapid unsupervised cryo-EM structure determination.** *Nat Methods* 2017,
29 **14**:290–296.
- 30 103. Mindell JA, Grigorieff N: **Accurate determination of local defocus and**
31 **specimen tilt in electron microscopy.** *J Struct Biol* 2003, **142**:334–347.
- 32 104. Jonić S: **Computational methods for analyzing conformational variability of**
33 **macromolecular complexes from cryo-electron microscopy images.** *Curr Opin*
34 *Struct Biol* 2017, **43**:114–121.
- 35 105. Roseman AM: **Docking structures of domains into maps from cryo-electron**
36 **microscopy using local correlation.** *Acta Crystallogr Sect D Biol Crystallogr*
37 2000, **56**:1332–1340.
- 38 106. Wriggers W, Chacon P: **Modeling tricks and fitting techniques for**
39 **multiresolution structures.** *Structure* 2001, **9**:779–788.
- 40 107. Farabella I, Vasishtan D, Joseph AP, Prasad A, Sahota H, Topf M: **TEMPy : a**
41 **Python library for assessment of three-dimensional electron microscopy**
42 **density fits.** *J Appl Crystallogr* 2015, **48**:1314–1323.
- 43 108. Vasishtan D, Topf M: **Scoring functions for cryoEM density fitting.** *J Struct*
44 *Biol* 2011, **174**:333–343.
- 45 109. Barad BA, Echols N, Wang RY, Cheng Y, DiMaio F, Adams PD, Fraser JS:
46 **EMRinger: side chain-directed model and map validation for 3D cryo-**

- electron microscopy.** *Nat Methods* 2015, **12**:943–6.
- Describes EMRinger, a tool that analyzes the statistical signature of high resolution features in cryoEM maps, especially side chain conformation, for the purpose of model refinement and validation.
110. Schneidman-Duhovny D, Pellarin R, Sali A: **Uncertainty in integrative structural modeling.** *Curr Opin Struct Biol* 2014, **28**:96–104.
111. Henderson R, Sali A, Baker ML, Carragher B, Devkota B, Downing KH, Egelman EH, Feng Z, Frank J, Grigorieff N, et al.: **Outcome of the first electron microscopy validation task force meeting.** *Structure* 2012, **20**:205–214.
112. Kucukelbir A, Sigworth FJ, Tagare HD: **Quantifying the local resolution of cryo-EM density maps.** *Nat Methods* 2014, **11**:63–5.
- Presents ResMap, a user-friendly tool for assessing and visualizing local variations in cryoEM map resolution.
113. Cardone G, Heymann JB, Steven AC: **One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions.** *J Struct Biol* 2013, **184**:226–236.
114. Lyman Monroe GT, Daisuke Kihara: **Variability of Protein Structure Models from Electron Microscopy.** *Structure* 2017, **25**:592–602.
115. Volkman N: **The joys and perils of flexible fitting.** In *Advances in Experimental Medicine and Biology*. . 2014:137–155.
116. Chen S, McMullan G, Faruqi AR, Murshudov GN, Short JM, Scheres SHW, Henderson R: **High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy.** *Ultramicroscopy* 2013, **135**:24–35.
117. Falkner B, Schröder GF: **Cross-validation in cryo-EM-based structural modeling.** *Proc Natl Acad Sci U S A* 2013, **110**:8930–5.
- Presents a procedure for cross-validation of cryoEM structures in which high frequency data, excluded from the reconstruction used for modeling, are subsequently applied to detect over-fitting.
118. Dimaio F, Zhang J, Chiu W, Baker D: **Cryo-EM model validation using independent map reconstructions.** 2013, **22**:865–868.
119. Hryc CF, Chen D-H, Afonine P V, Jakana J, Wang Z, Haase-Pettingell C, Jiang W, Adams PD, King JA, Schmid MF, et al.: **Accurate model annotation of a near-atomic resolution cryo-EM map.** *Proc Natl Acad Sci* 2017, **114**:3103–3108.
- Describes a procedure for annotating cryoEM structures using β -factors and local density values to produce density-weighted models that describe the uncertainty in positioning of each atom.
120. Herzik MA, Fraser J, Lander GC: **A multi-model approach to assessing local and global cryo-EM map quality.** *bioRxiv* 2017, doi:10.1101/128561.
121. Ahmed A, Tama F: **Consensus among multiple approaches as a reliability measure for flexible fitting into cryo-EM data.** *J Struct Biol* 2013, **182**:67–77.
122. Wriggers W, He J: **Numerical geometry of map and model assessment.** *J Struct Biol* 2015, **192**:255–261.
123. Perilla JR, Goh BC, Cassidy CK, Liu B, Bernardi RC, Rudack T, Yu H, Wu Z, Schulten K: **Molecular dynamics simulations of large macromolecular**

- 1 **complexes**. *Curr Opin Struct Biol* 2015, **31**:64–74.
- 2 124. Sali A, Berman HM, Schwede T, Trewhella J, Kleywegt G, Burley SK, Markley J,
- 3 Nakamura H, Adams P, Bonvin AMJJ, et al.: **Outcome of the first wwPDB**
- 4 **hybrid/integrative methods task force workshop**. *Structure* 2015, **23**:1156–
- 5 1167.
- 6 • Reports the outcome of discussions, involving a diverse set of experts, regarding the
- 7 interpretation and publication of integrative and hybrid models as well as associated data.
- 8
- 9

1 **Table 1.** List of commonly used methods and tool suites for various hybrid modeling
2 tasks.

	Task	Method/Tool Suite**
Input	Comparative Modeling	Modeller [15], RosettaCM [16]
	<i>Ab initio</i> Structure Prediction	MELD [22], I-TASSER [19]
	Model Selection/Scoring	MolProbity [26], DOPE [21]
	Map Segmentation	Chimera [37], Segger [38]
EM-based Modeling	Rigid Fitting	Chimera [37], Situs [31], HADDOCK [32]
	Flexible Fitting	MDFF [52], Flex-EM [49], DireX [40]
	<i>De novo</i> Modeling (<3.5 Å)	Coot [77], Phenix [78], REFMAC [81]
	<i>De novo</i> Modeling (3.5-5 Å)	Gorgon [88], Rosetta <i>de novo</i> [95, 96], RosettaES [97]
	Structural Refinement (<5 Å)	Rosetta <i>de novo</i> [95, 96], RE-MDFF [69]
Validation	Stereochemical Assessment	MolProbity [26], EMRinger [109]
	Goodness-of-fit Measure	CCC [108], iFSC [96], ResMap [112]
	Feature Stability	All-atom MD [123]
	Over-fitting Detection	Cross-validation [117,118]

3

4 ** The authors emphasize that the cited examples are based on their personal user
5 preference and represent but a fraction of tools available for any given task.

1 **Figure legends:**

2 **Figure 1.** General workflow for cryoEM-based hybrid modeling as a function of map
3 resolution. An initial atomic model may be obtained, for instance, from an existing,
4 rigidly-docked structure or through the use of *de novo* modeling techniques. Depending
5 on the resolution of the map, model refinement strategies range from broad
6 conformational refinement using flexible-fitting methodologies to more detailed
7 secondary structure and side chain refinement using, for example, Rosetta-based
8 approaches or XRC software suites. The accuracy of a refined model may be
9 subsequently assessed by way of stereochemical or goodness-of-fit measures as well as
10 cross-validation procedures to identify over-fitting.

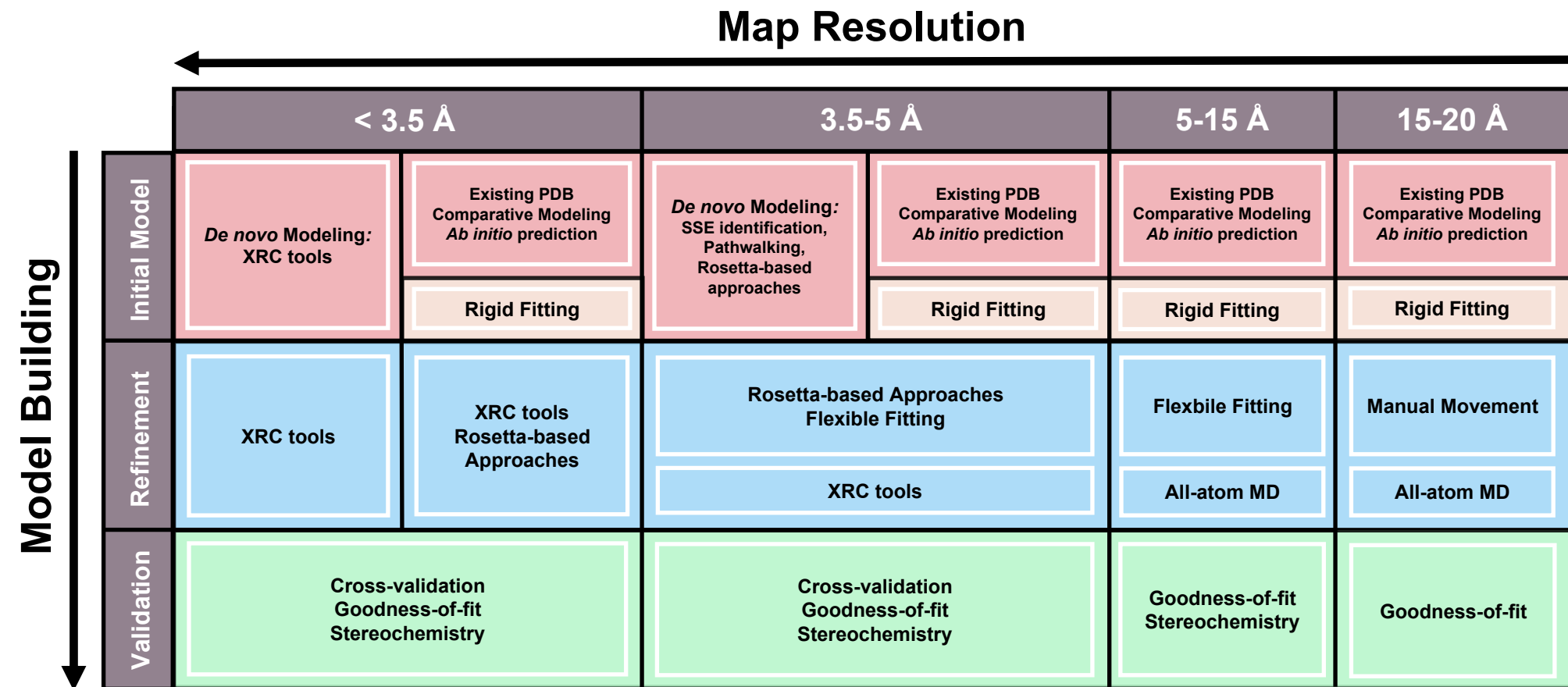
11

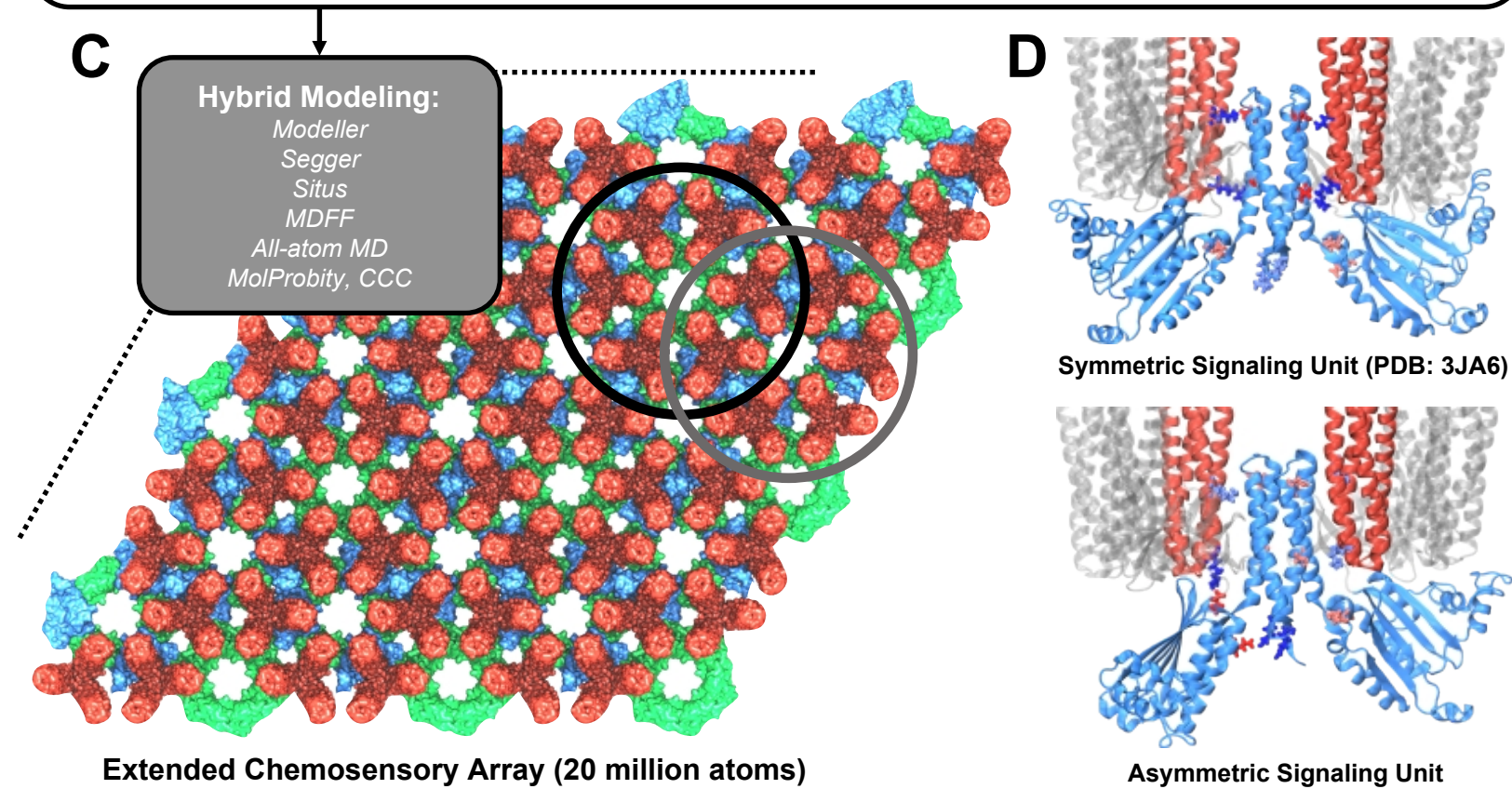
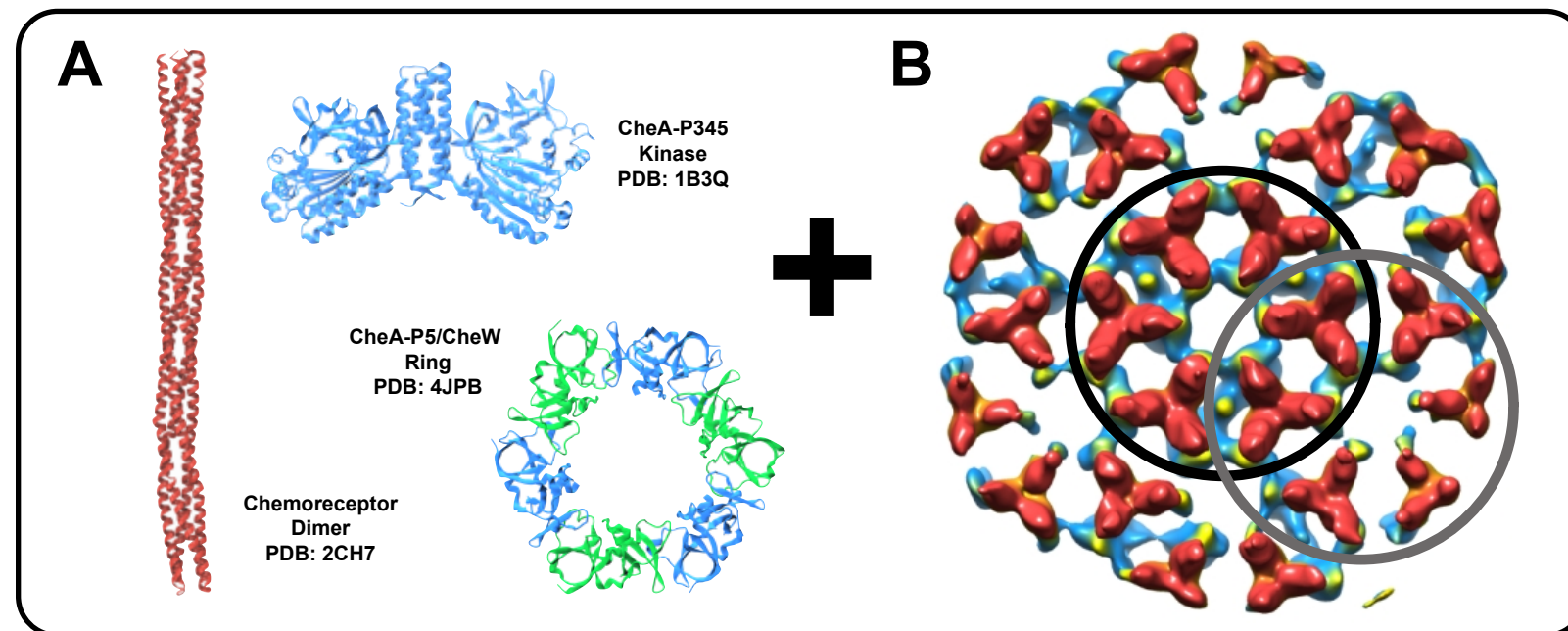
12 **Figure 2.** Quaternary structure of bacterial chemosensory array determined by hybrid
13 modeling and intermediate-resolution cryoSTAC. Combining high resolution, XRC-
14 derived component structures (A) with near-sub-nanometer resolution maps of
15 chemosensory array (B) Cassidy *et al.* used an assortment of hybrid modeling tools (grey
16 box) to construct an atomistic model of the extended signaling array 20-million-atoms in
17 size (C). Distinct symmetry centers used for sub-tomogram classification are circled in
18 black and grey in (B) along with their corresponding atomistic regions in (C). Further
19 investigation of a portion of the model using all-atom MD simulations confirmed its
20 stability and revealed tertiary conformational changes, along with stabilizing contacts
21 (shown in licorice representation), that impact kinase control within the fundamental
22 signaling unit (D). An atomic model of the MDFF-refined symmetric signaling unit was
23 deposited in the PDB under accession code 3JA6. Figure panels adapted from [63••].

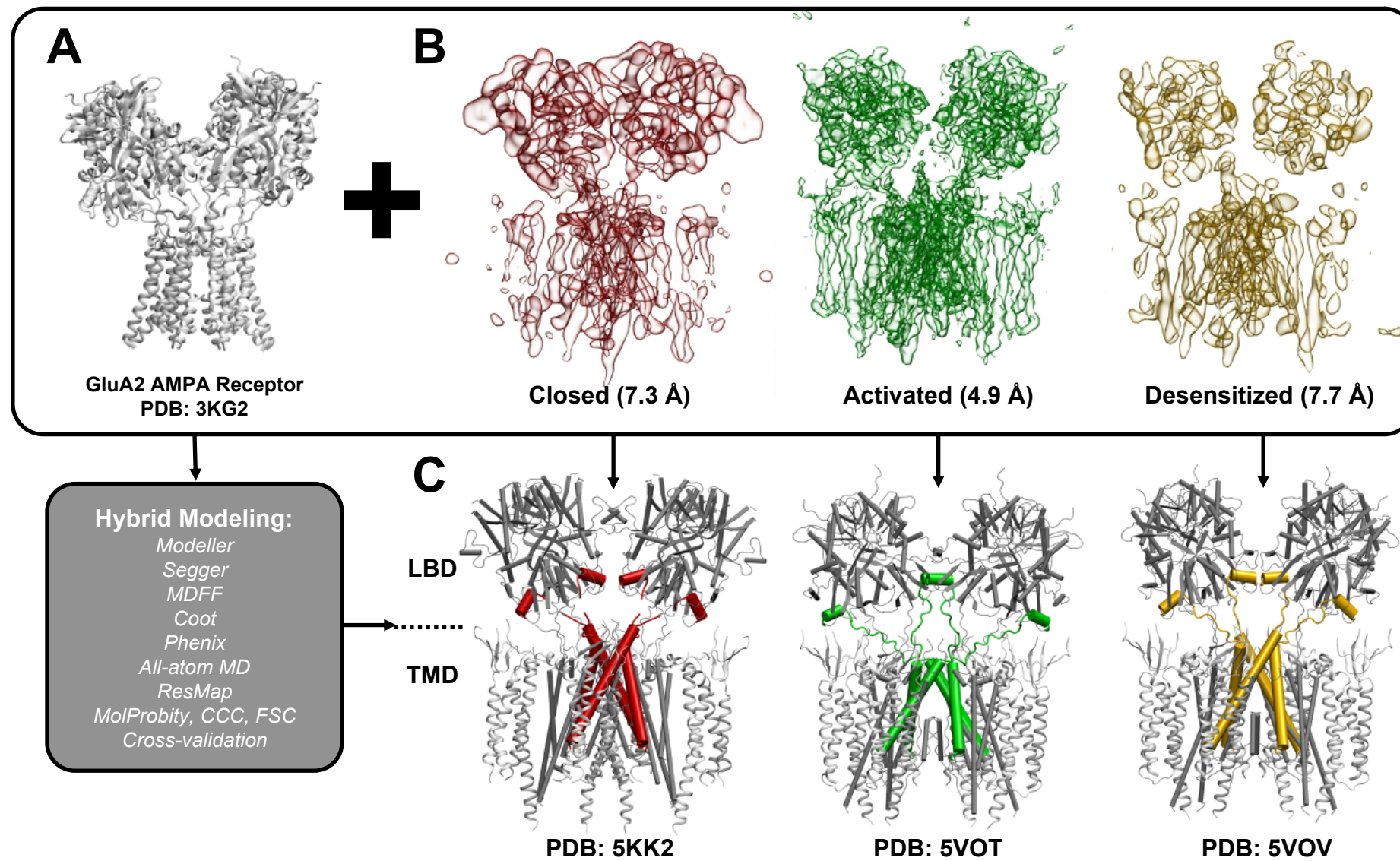
24

25 **Figure 3.** Hybrid modeling provides insight into the mechanisms of activation and
26 desensitization in AMPA receptor-TARP complex, a membrane-bound assembly
27 responsible for excitatory synaptic transmission in mammals. Using a wide range of
28 hybrid modeling tools (grey box) with a high-resolution XRC structure of AMPA

1 receptor (A), atomic models of the AMPA receptor-TARP complex corresponding to the
2 closed state (PDB: 5KK2), the activated (PDB: 5VOT) and desensitized (PDB: 5VOV)
3 states were constructed based on respective cryoEM density maps (B) [99,100••],
4 revealing the large-scale conformational changes undergone by the receptor in response
5 to neurotransmitter-binding as well as residue-level information suggesting potential
6 gating mechanisms (C). The central region of each model is shown using a pipe
7 representation to highlight the coupling of conformational changes in the ligand binding
8 (LBD) and transmembrane (TMD) domains (C). Reproduced with permission (Copyright
9 2017, Yuhang Wang).







No conflicts of interest.