Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Structural Biology

Computational modeling of protein assemblies

Neelesh Soni[1] and MS Madhusudhan[1,2]

Computational methods to predict the 3D structures of protein interactions fall into 3 categories—template based modeling, protein–protein docking and hybrid/integrative modeling. The two most important considerations for modeling methods are sampling and scoring conformations. Sampling has benefitted from techniques such as fast Fourier transforms (FFT), spherical harmonics and higher order manifolds. Scoring complexes to determine binding free energy is still a challenging problem. Rapid advances have been made in hybrid modeling where experimental data are amalgamated with computations. These methods have received a boost from the popularity of experimental methods such as electron microscopy (EM). While increasingly larger and complicated complexes are now getting elucidated by integrative methods, modeling conformational flexibility remains a challenge. Ongoing improvements to these techniques portend a future where organelles or even cells could be accurately modeled at a molecular level.

Addresses
[1] Indian Institute of Science Education and Research, Pune. Dr. Homi Bhabha Road, Pashan, Pune 411008, India
[2] Bioinformatics Institute, A*STAR, 30 Biopolis Street, #07-01 Matrix, 138671, Singapore

Corresponding author: Madhusudhan, M.S. (madhusudhan@iiserpune.ac.in)

Current Opinion in Structural Biology 2017, **44**:179–189

This review comes from a themed issue on **Sequences and topology**

Edited by **Ramanathan Sowdhamini** and **Kenji Mizuguchi**

http://dx.doi.org/10.1016/j.sbi.2017.04.006

Introduction

Increasingly, the attention in molecular structural biology is being focused on molecular interactions and molecular machines/complexes. The frontiers are being pushed by the solution of larger and more complex assemblies of biomolecules [1,2•,3–5]. These assemblies not only elucidate the functional roles of individual molecules but also uncover complex molecular and cellular phenomena. In the last 2–3 years there has been a large influx of near atomic resolution structures of protein complexes using electron microscopy (EM). This has changed the landscape of structural biology [6,7], and is conceivably only the beginning of a rapidly increasing trend (Figure 1).

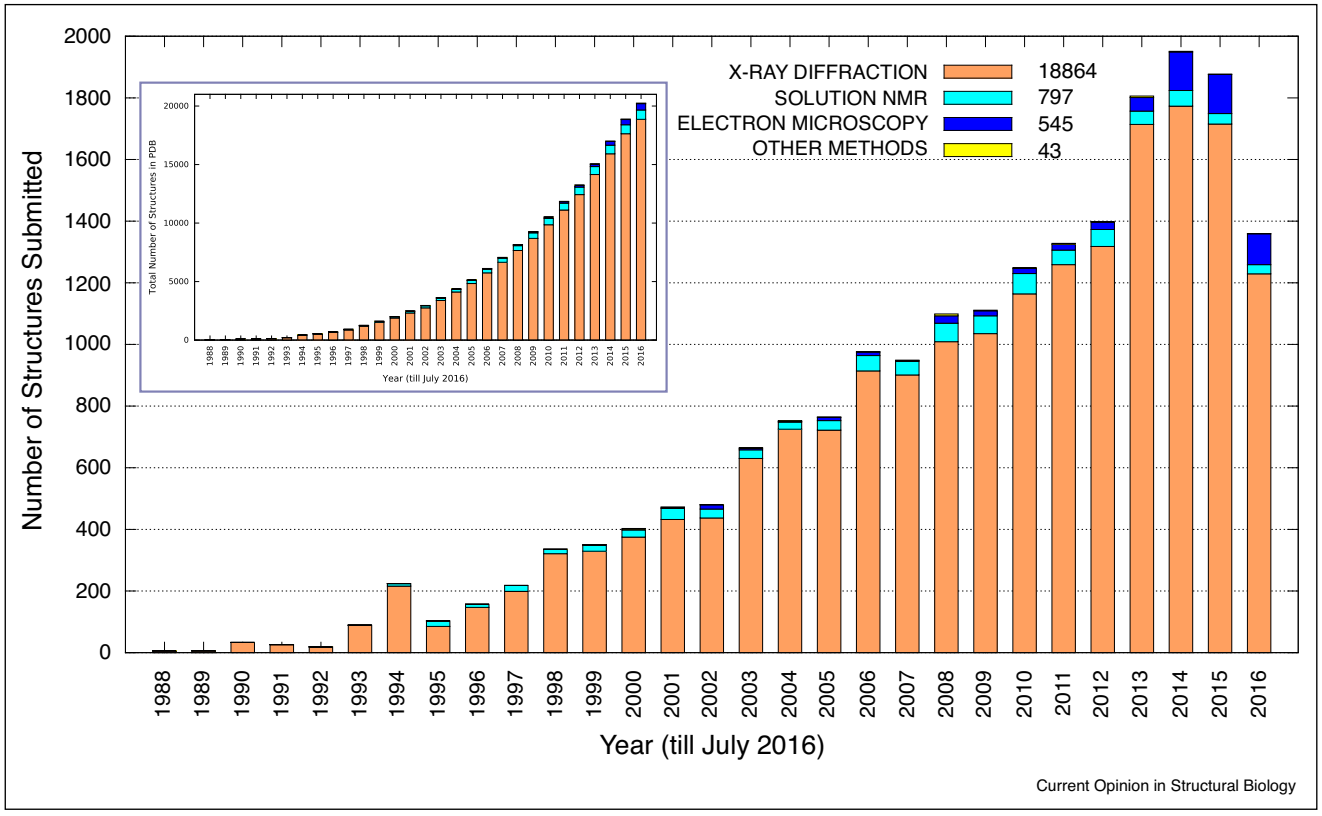
Despite these recent advances vast numbers of interacting complexes are not annotated by experimentally determined 3D structures [8]. In this review we restrict the discussion to only proteins complexes. By various estimates the number of such protein complexes could be in the hundreds of thousands, but limited, just as there are only a few thousands of different protein folds [9,10]. These complexes encompass a wide variety of associations including obligate interactions, transient interactions, interactions restricted to sub-cellular compartments, host–pathogen interactions [11] *etc.* Protein complexes could be homo- or hetero-oligomeric with up to 70% being the former [12]. Though, we now have deeper insights into the dynamics and evolution of the quaternary associations [12,13,14•] in these protein complexes, a significant challenge of modeling them still remains. A large number of curated repositories and databases house information on known protein complexes [15–25]. These data are potential training grounds for computational methods to make predictions of protein–protein interactions. This review deals with the different computational techniques that are used to predict/construct 3D structures of protein complexes.

Modeling protein complexes

Protein complexes could be as simple as a stable symmetric homo-dimer or as complicated as a multi-component hetero-oligomer whose constituents assemble/disassemble dynamically. *In silico* modeling of protein assemblies can be broadly categorized into 3 types—(a) template based docking [26••,27,28], (b) template-free computational docking and (c) Docking guided by restraints obtained from experimental procedures or Hybrid modeling (see section on hybrid modeling and Figure 2). This section and the ones immediately following it deal primarily with template based and template free modeling methods.

Template based methods could, in principle, be used to model multi-component complexes [29–31]. A pre-requisite for this would be multimeric templates. These are often in short supply and hence the coverage of template-based methods is also mostly limited to pair wise interactions [29,32,33]. Template-based methods have also evolved to not rely on homologous templates alone but to also cover cases where there is interface similarity without sequence homology [26••,34,35]. Despite the fact that interface templates are possibly present to model many interactions [36], template based methods when used as is, generally underperform in comparison to other

Figure 1



Growth of multimeric complexes in PDB.
The number of protein oligomers (with more than 2 subunits) deposited in the PDB [107], as a function of time (spanning the years 1988–2016 (July)). The cumulative growth in these numbers is shown in the inset. Structures solved by X-ray crystallography, NMR spectroscopy, electron microscopy (EM) and other methods are shown in orange, cyan, blue and yellow respectively. The number of 3D structures of the complexes solved by the different methods is shown in the legend. While X-ray crystallography continues to be the largest contributor, the last 2–3 years has seen a spurt in the number of such structures solved by EM. Of the 545 EM structures, 352 (65%) have been solved in the last 2.5 years. Given that EM is a technique amenable to solving the structures of large assemblies with relatively simpler sample preparation, it is set to elucidate the 3D structures of a large number of complexes, both big and small [7,108]. Computational modeling and simulations are integral to the solution and interpretation of the many of the EM structures. It is expected that the accuracy and applicability of these methods would improve with the increase in the number of EM resolved structures.

docking methods [27]. However, with special consideration, these methods could perform on par with the best available docking tools [27,31].

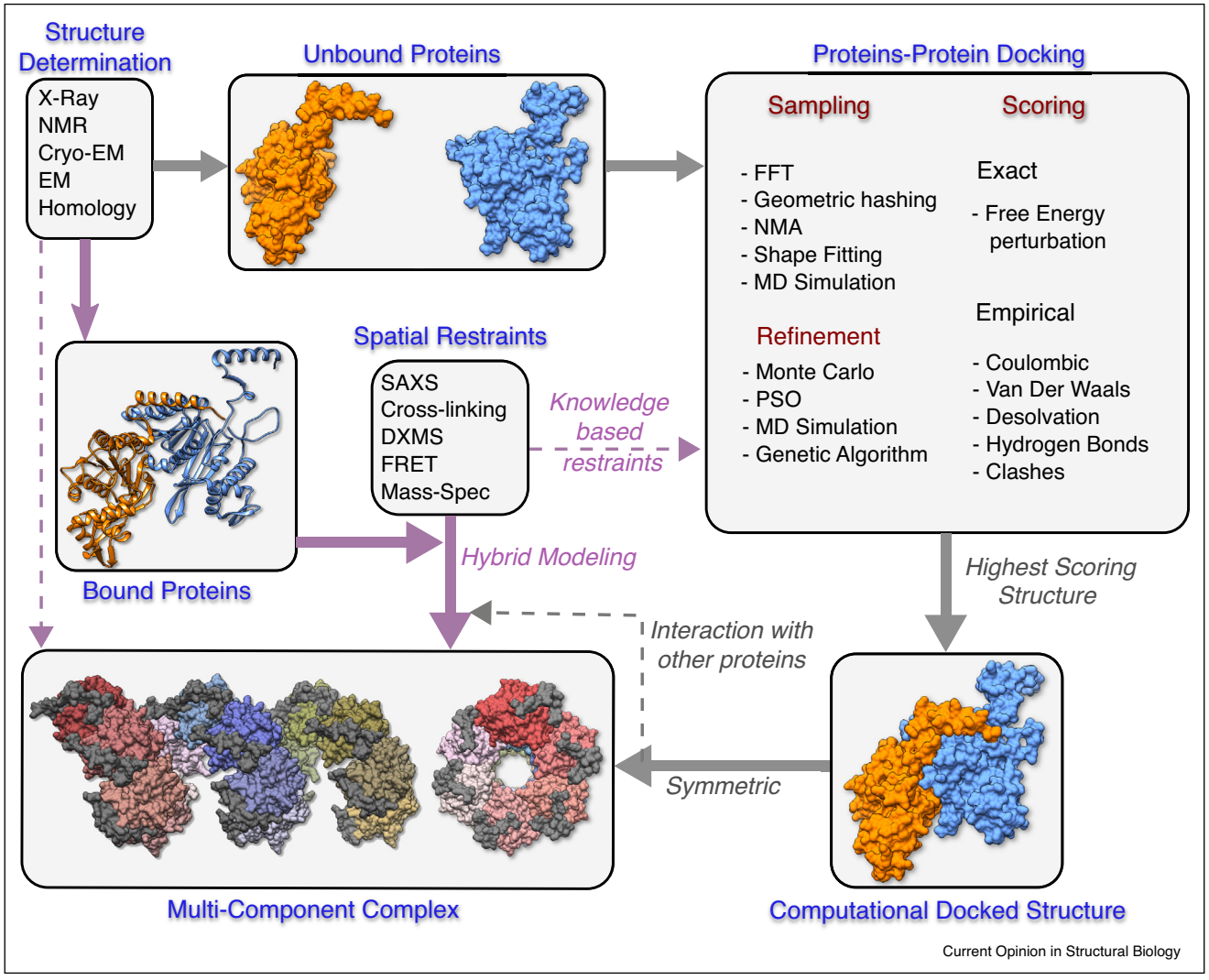
Most (all) protein docking methods assemble pairs of protein structures. A few of the docking software engineer their pair-wise assemblies to account for larger oligomeric complexes. Notable among such methods are RosettaDock [37], CombDock [38,39] and Multi-LZerd [40]

All docking programs face the two significant challenges of scoring and sampling. Template based docking programs do minimal sampling, while template free docking methods invest considerable effort at this stage. The docking software should sample a sufficient number of conformations such that near native conformations of the protein–protein complex are encountered. The programs should also be armed with a scoring scheme to discriminate between native and non-native conformations. Ideally, the scoring scheme should also be sensitive enough to discriminate between native and near native conformations. In the sections below we review recent advances in different sampling and scoring techniques. Though we have treated these two aspects separately, many of the methods integrate scoring and sampling during the docking process.

Sampling conformational space

An exhaustive search of all possible conformations of protein–protein interfaces including conformation flexibility is practically intractable. Prior knowledge of the (approximate) location of interactions sites could reduce sampling complexity and time. In the absence of such knowledge, many docking programs consider proteins as rigid bodies. Even with this approximation most methods do not exhaustively scan the entire molecular surface of the interactors. Methods such as the FFT correlation

Figure 2



A general schema for computationally assembling a multi-protein complex.
Computational techniques are necessary when the structures of multi-oligomeric complexes cannot be obtained experimentally. The individual components of an assembly could either be obtained from the PDB or computationally modeled. The whole complex or sub-complexes could be obtained by protein–protein docking. Mostly these docking exercises deal with pairs of proteins and employ a myriad of sampling and scoring techniques to produce an optimal docked structure. Whole or sub-complexes could also be obtained by assimilating data from different experiments that provide restraints to the docking exercise. The experiments could be any one of many (a partial list is shown) whose readout directly or indirectly leads to some spatial inference. The 32-mer complex structure represented in the multi-component complex panel is a homology/template-based model of the inactive human Rad51 filament (16mer) complexed with the 16 BRC4 fragments. The modeled structure, together with the EM structure of the active filament, helps elucidate the mechanism of interaction between Rad51 and the tumor suppressor BRCA2 [104•,105].

approach [41,42•] and Geometric Hashing [43] are exceptions to this trend and systematically cover the entire accessible interacting surface. Most current methods initially perform rigid body docking and then select the top few conformations as filtered by their scoring functions. These conformations are further optimized by incorporating the surface flexibility of the interacting partners [44]. Numerous algorithms are employed to model the dynamics/flexibility of the interacting surfaces such as normal mode analysis [45], molecular dynamic simulations [46,47] and sampling the torsion angle space from available rotamer libraries [48–51].

To efficiently sample near native conformations, sampling algorithms have evolved by combining different features and techniques. The programs CLUSPRO [52], GRAMM-X [53], DOT [54], HEX [55], FTDOCK [56], ZDOCK [57,58••,59], M-ZDOCK [60], PIPER [61] and MEGADOCK4.0 [62] all use FFT based sampling that implement shape complementarity along with other

improvements. For instance, DOT and HEX incorporate electrostatic complementarity whereas FTDOCK, ZDOCK, MEGADOCK 4.0 and M-ZDOCK combine electrostatic and desolvation free energy. PIPER incorporates pairwise knowledge-based interaction potentials. Among the methods that do not take recourse to FFT, PATCHDOCK [63,64] and SYMMDOCK [63,64] work by dividing the surface into concave, convex and flat patches. This ensures better shape complementarity and fewer atomic clashes while maximizing interface area. Another non-FFT method, RosettaDock [65,66•], attempts to achieve atomic level accuracy by searching backbone and side chain conformational space using Monte Carlo in combination with various biochemical and pairwise residue propensities. FRODOCK2.0, an improved version of FRODOCK [67••,68], uses a combination of grid based potentials, spherical harmonics (to deal with rotational degrees of freedom) and a knowledge based potential for docking.

Scoring schemes for evaluating protein–protein interfaces

Increasing computational power should alleviate some difficulties associated with conformational sampling. The main issue would then be to distinguish the native, or near-native structure from among the sampled decoys. Many aspects of protein interactions confound scoring schemes. For instance, analyses of protein–protein interfaces show that native conformations do not necessarily have the most hydrogen bonds or the largest buried solvent accessible surface area [69–71]. Induced fit conformations on the interface and the role of molecular crowding [1,72] and solvent (water and ions) in mediating interactions further complicate scoring schemas.

Just as in sampling techniques, scoring techniques have evolved by combining various features and properties. Methods such as GRAMM-X [53], ATTRACT [73,74], 3D-DOCK [50,75], LZERD [76] *etc.* use different combinations of solvation energy, dielectric constants, electrostatics, van der Waals interaction, hydrogen bonds, clashes, clustering using pairwise RMSD *etc.* Of these methods, HADDOCK [77–79], CLUSPRO [52] and SWARMDOCK [72,80] have demonstrably outperformed other methods [27]. HADDOCK scoring works synergistically with sampling to refine docked structures using explicit water, buried surface area, interaction restraint energy and desolvation energy. Similarly, CLUSPRO uses electrostatics and desolvation free energy for filtering models. SWARMDOCK uses particle swarm optimization along with van der Waals, electrostatics, hydrogen bonding and desolvation energy terms.

While this review mainly concerns itself with methods that bundle model building and assessment into a single package, one must be aware of the development of stand-alone assessment methods. These assessments make use of interaction energy and sequence conservation to deduce binding affinity [81]. Methods such as KFC [82], MINERVA [83], HotPoint [84], *etc.* (not an exhaustive list), also make predictions of hotspot residues. For a recent comparison of the performance of these methods refer to [85]. These methods could be used in conjunction with any of the packages mentioned above as an independent assessment of the predicted complexes. Recent advances have also seen the growth of machine learning methods to predict interaction interfaces (for a review see [86]).

Challenges for sampling and scoring techniques

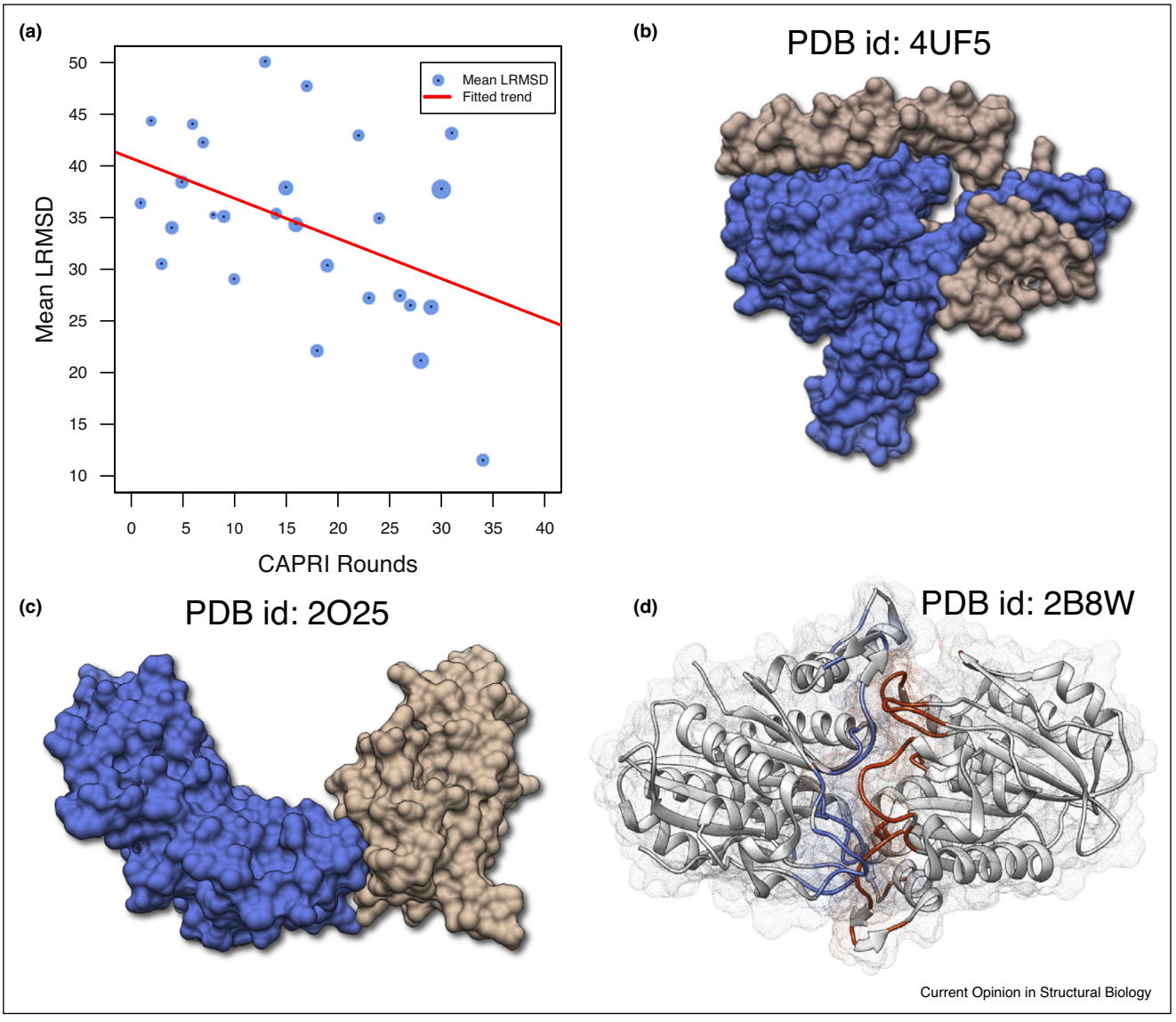
There are two major sampling challenges to protein–protein docking—speed and conformational flexibility. Docking methods should be able to sift through billions of possible configurations. Many methods hence use FFT based sampling as it is swifter than Monte Carlo and geometric fitting based methods. Recent advances such as the fast manifold Fourier transform [42•] and GPU implementations (as in MEGADOCK4.0) have significantly enhanced FFT sampling speed even further. Programs such as FRODOCK and HADDOCK have further improved sampling speed with parallel implementation of their algorithms. In all these methods, the increase in speed does not compromise on accuracy. The problem with conformational flexibility, however, still remains. Methods such as HADDOCK, that take some flexibility into consideration, albeit at a refinement stage, tend to perform better than methods that treat interactors as rigid bodies. Note that in methods that account for flexibility, only small conformational changes are sampled. Extensive conformational variations, such as those observed in some induced fit interactions remain computationally illusive.

To identify a binding event, a fast and accurate scoring function should complement an exhaustive sampling algorithm. Ideally, the scoring scheme should compute the free energy of binding. Such computations are hard to accomplish and none of the current methods consistently make accurate estimates of the binding free energy [87]. Most scoring functions discriminate between binding and non-binding events, while being unable to consistently rank order binding partners by affinity [88].

Another aspect of scoring that needs improving, in the face of the increasing number of multi-component complexes being modeled, is determining the relative stoichiometry of binding partners. In other words we need to determine how many concurrent binding partners a particular protein is likely to have. The scoring scheme should be able to determine if the free energy of the complex would decrease with/without additional interactors. In general, the environment of the interaction could influence the strength of binding. A recent method makes use of explicit water in the docking protocol to account for this environment [89•]

The CAPRI experiment serves as an ideal testing platform for new docking methods. It is also a barometer of the progress we have made in protein–protein docking over the years. Prediction difficulties are usually associated with interfaces that are made of more than 1 surface patch, or with interfaces that are flexible (Figure 3).

Figure 3



The progress made in protein assembly modeling as gauged from the CAPRI experiments.
LRMSD is the root mean square deviation (RMSD) of the backbone atoms of amino acids in protein complexes when modeled structures are compared to the native. The smaller the LRMSD, the better is the prediction, with the most accurate prediction having a value of 0. Starting from July 2001, 35 rounds of the CAPRI experiment has elicited predictions for 107 different target complexes from several different research groups [27,109–115]. The average LRMSD is computed for all qualified models of a particular CAPRI round. The size of the points are indicative of number of structures deposited in a particular round. The average LRMSD is decreasing over time indicating that computational techniques are improving with time (Panel A). However, the absolute value of the average LRMSD remains high (currently ~20 Å). The LRMSD improvement is not monotonic, as for different rounds the criteria for acceptability of models differ. Another contributing factor could be that, unlike the homology modeling part of CASP experiments, there is no measure such as sequence identity or e-value that gauges the difficulty level of targets. The CAPRI predictors, automated servers and humans alike, faced difficulties when sub units contacted one another in several non-contiguous patches (Panel B, CAPRI target 101, Complex of ubiquitin carboxyl-terminal hydrolase isozyme L5 (blue) and nuclear factor related to Kappa-B binding protein (grey); pdb code: 4uf5 [116]), or when the interface regions were small (crystal contact like) (Panel C, CAPRI target 27, Ubiquitin-conjugating enzyme E2 (blue) and SUMO-1-conjugating enzyme UBC9 (grey); pdb code: 2o25) or when the interface residue were predominantly from non-secondary structure regions (Panel D, CAPRI target 23, homodimer of Inteferon induced guanylate binding protein; pdb code: 2b8w [117]).

However, on average, the overall accuracy with which the 3D structures of protein complexes have been modeled has continuously improved over the years.

Hybrid modeling

From the state of the art in protein–protein docking, it is clear that the accuracy of predicted/modeled 3D structure improves with some prior knowledge. For instance, knowledge of the binding site regions vastly reduces the search space. A more generic implementation of this idea is to complement computational docking/modeling efforts with experimental data [90]. These data should be of the kind that could directly or indirectly lead to inference of spatial features such as distances, angles, torsion angles, volumes *etc.* [91]. Advances in technology, such as EM and Mass spectrometry, make it possible to attempt to solve 3D structures of large and complex molecular machines [6,7,92–94]. However, a key aspect of these new technologies is the flexibility to combine experimental data with computations [94,95]. For instance, knowing the overall shape (obtained from EM studies) does not necessarily inform one about sub unit packing [96]. This leads to the requirement of new and improved algorithms to 'dock' subunits with restraints [97,98]. Depending on the restraints, different types of docking solutions are possible, including fitting/refining structures and/or models to EM density maps. For flexible fitting and refinement, molecular dynamics simulations and other methods are utilized [99–101].

There are many protein complexes that have been solved using experimental data integrated with computations (Table 1, and schematically represented by the purple arrowhead path in Figure 2). Sometimes, none of the experimental techniques are able to determine the

Table 1

Examples of structures elucidated by hybrid modeling. This list includes only those complexes that have used at least two different experimental methods in addition to computational technique(s) to construct the 3D structure

Model system	Structural data used	Reference
Phosphoinositide 3-Kinase p85α Homodimer	SAXS, Cross-linking	[118]
Yeast Mediator complex	Cross-linking, Mass spectrometry, X-ray crystallography, homology modeling, Cryo-EM	[119]
Chromatin	Mass spectrometry, Cryo-EM	[120–123]
Human and Yeast TFIIH	Cross-linking, Mass spectrometry, biochemical analysis, Electron microscopy	[124]
Bilitranslocase Transmembrane Domain	Bioinformatics, NMR, FRET	[125]
Yeast eIF3:eIF5 complex	Cross-linking, Mass spectrometry	[126]
DNA Repair Complex of Human Rad9-Hus1-Rad1/FEN1/DN	X-ray crystallography, Single particle EM	[127,128]
Ubiquitin-modified PCNA	SAXS	[127,129]
Photoreceptor phosphodiesterase PDE6 catalytic dimer (αβ)	Cross-linking and Mass spectrometry	[130]
Actin-scruin complex	X-ray crystallography, Cross-linking	[131]
SEA complex	Affinity purification, Cross-linking	[132]
40S·eIF1·eIF3 Translation Initiation Complex	Electron microscopy, Cross-linking, Mass spectrometry	[133]
26S proteosome	X-ray crystallography, Cross-linking, Mass spectrometry, Cryo-EM	[134]
Nucleo-pore Complex, Nup84 complex	Ultracentrifugation, Quantitative immunoblotting, Affinity purification, overlay assay, Electron microscopy, Immuno-EM, Bio-informatics, Membrane fractionation, Cross-linking, Mass spectrometry, X-ray crystallography, homology modeling	[135–137]
Proteasomal lid (back calculate)	Mass Spectrometry, Proteomics, ion mobility-MS, Cross-linking	[138]
Proteasome interacting proteins (PIPs)(back calculate)	Mass Spectrometry, Proteomics, ion mobility-MS, Cross-linking	[138]
Chaperone involved in formation proteasomal base complex (back calculate)	Mass Spectrometry, Proteomics, ion mobility-MS, Cross-linking	[138]
Collicin-immunity protein and dnase complex (back calculate)	X-ray crystallography, Mas spectrometry	[5]
Eukaryotic Ribosome	Cryo-EM, Homology Modeling	[139]
Mammalian Ribosome	Cryo-EM, Homology Modeling	[140]
Exosome on polyribonucleotide phosphorylase (PNPase)	Homology modeling, Electron microscopy	[33]
RNA polymerase II with RPB4/7	X-ray crystallography, Homology modeling, Electron microscopy	[33]
Chaperonin CCT and phophoducin 2(PLP2) and G protein homolog VID27	X-ray crystallography, Electron microscopy	[33]
Ski complex	Homology modeling, Electron microscopy	[33]
RNase P (POP) complex	Electron microscopy	[33]
Bacterial type II pilus	Cryo-EM, EM, NMR, Cross-linking	[141,142]
Ryanodine receptor channel	Cryo-EM, Homology Modeling	[143]
TriC/CCT chaperonin	Cryo-EM, Homology Modeling	[144]
γ-tubulin small complex	Cryo-EM, Cross-linking, Homology Modeling	[145]
Triprptidyl peptidase II	Cryo-EM, Homology Modeling, molecular dynamics fitting, Single particle analysis	[146]
Dam1 kinetochore complex	Cross-linking	[147]

spatial restraints enough for hybrid modeling (purple color dashed line in Figure 2). In such cases protein-protein docking and knowledge-based restraints contribute significantly in determining the 3D structure of the complex (gray arrow heads in Figure 2). Recent successes of hybrid modeling include the NUP82 complex of the Nucleopore [102••] and the Ferrodoxin-FusA complex [103••]. A more comprehensive list is given in Table 1.

What is in store for the future of computational modeling of complexes

Template (homology) based docking methods, despite their limitations, are still effective when dealing with protein complexes whose related homologues have been characterized experimentally. These methods are particularly powerful in furthering biological hypothesis, for instance, in determining the 3D structures of active and inactive states of complexes [104•,105]. Many large scale interactome studies have been carried out using such template based studies alone [1,29,32]. The lack of adequate templates, and the fact that 'off the shelf' homology solutions are unreliable, means that protein–protein docking is widely used. Over the years, the CAPRI experiments have shown that docking methods have continually improved. Docking sampling techniques have made many advances, mostly in processing speed, using methods such as FFT. Scoring the complexes remains a significant problem and is driving the evolution of new and improved methods. Our analyses of the results over several rounds of CAPRI show that pairwise binding when mediated by more than one surface patch leads to difficulties. Somewhat related to this is the fact that docking is mostly restricted to assembling pairs of proteins or sub-complexes. There appears to be no clear method for an *a priori* determination of the number of components of the complex during the docking procedure.

Docking, despite some limitations, is instrumental in providing experimentally testable hypothesis. Successful experimental confirmations of the predictions not only validate the hypothesis but also validate the 3D models. In this light, the last few years have seen the emergence of integrated modeling or hybrid modeling. Philosophically, experiments and computations are treated as complementary approaches. Hybrid modeling techniques use many experimental and computational methods to resolve the structure of interacting complexes. The advent of EM has been a fillip to the development of such methods. Conformational flexibility is taken care of by modeling and molecular dynamics [106] and soon ambitiously large systems of complexes now modeled by packages such as CellPack [2•] could be within the reach of such methods.

Acknowledgements

MSM would like to acknowledge the Wellcome trust-DBT India alliance for a senior fellowship. Neeladri Sen is thanked for useful discussions.

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. Im W, Liang J, Olson A, Zhou HX, Vajda S, Vakser IA: **Challenges in structural approaches to cell modeling**. *J. Mol. Biol.* 2016, **428**:2943-2964.

2. • Johnson GT, Autin L, Al-Alusi M, Goodsell DS, Sanner MF, Olson AJ: **cellPACK: a virtual mesoscope to model and visualize structural systems biology**. *Nat. Methods* 2014, **12**:85-91.

This visualization/modeling tool offers a window view into the future of hybrid modeling where models would be built not just for individual complexes but for systems of complexes such as organelles or entire cells.

3. Wang RY-R, Song Y, Barad B, Cheng Y, Fraser J, DiMaio F: *Automated Structure Refinement of Macromolecular Assemblies from Cryo-EM maps Using Rosetta*. 2016.

4. Robinson CV, Sali A, Baumeister W: **The molecular sociology of the cell**. *Nature* 2007, **450**:973-982.

5. Karaca E, Bonvin AMJJ: **Advances in integrative modeling of biomolecular complexes**. *Methods* 2013, **59**:372-381.

6. Beck M, Baumeister W, Alberts B, Asano S, Robinson CV, Schur FK, Pfeffer S *et al.*: **Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail?** *Trends Cell Biol.* 2016, **0**:291-294.

7. Glaeser RM: **How good can cryo-EM become?** *Nat. Methods* 2016, **13**:28-32.

8. Mosca R, Pons T, Céol A, Valencia A, Aloy P: **Towards a detailed atlas of protein–protein interactions**. *Curr. Opin. Struct. Biol.* 2013, **23**:929-940.

9. Hart GT, Ramani AK, Marcotte EM: **How complete are current yeast and human protein-interaction networks?** *Genome Biol.* 2006, **7**:120.

10. Hakes L, Pinney JW, Robertson DL, Lovell SC: **Protein–protein interaction networks and biology—what's the connection?** *Nat. Biotechnol.* 2008, **26**:69-72.

11. Davis FP, Barkan DT, Eswar N, McKerrow JH, Sali A: **Host pathogen protein interactions predicted by comparative modeling**. *Protein Sci.* 2007, **16**:2585-2596.

12. Levy ED, Teichmann S: **Structural, evolutionary, and assembly principles of protein oligomerization**. *Prog. Mol. Biol. Transl. Sci.* 2013, **117**:25-51.

13. Levy ED, Boeri Erba E, Robinson CV, Teichmann SA: **Assembly reflects evolution of protein complexes**. *Nature* 2008, **453**:1262-1265.

14. • Ahnert SE, Marsh JA, Hernandez H, Robinson CV, Teichmann SA: **Principles of assembly reveal a periodic table of protein complexes**. *Science* 2015, **350**:aaa2245.

A compact description of the quaternary associations in proteins and the patterns they follow.

15. Davis FP, Sali A: **PIBASE: a comprehensive database of structurally defined protein interfaces**. *Bioinformatics* 2005, **21**:1901-1907.

16. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D: **The database of interacting proteins: 2004 update**. *Nucleic Acids Res.* 2004, **32**:D449-D451.

17. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes HW *et al.*: **The MIPS mammalian protein–protein interaction database**. *Bioinformatics* 2005, **21**:832-834.

18. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Von Mering C *et al.*: **STRING v9.1: protein–protein interaction networks, with increased coverage and integration**. *Nucleic Acids Res.* 2013, **41**:D808-D815.

19. Veres DV, Gyurkó DM, Thaler B, Szalay KZ, Fazekas D, Korcsmáros T, Csermely P: **ComPPI: a cellular compartment-specific database for protein–protein interaction network analysis**. *Nucleic Acids Res.* 2015, **43**:D485-D493.

20. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M: **BioGRID: a general repository for interaction datasets**. *Nucleic Acids Res.* 2006, **34**:D535-9.

21. Prieto C, De Las Rivas J: **APID: agile protein interaction DataAnalyzer**. *Nucleic Acids Res.* 2006, **34**:W298-W302.

22. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, Galeota E, Sacco F, Palma A, Nardozza AP, Santonico E *et al.*: **MINT, the molecular interaction database: 2012 Update**. *Nucleic Acids Res* 2012, **40**.

23. Bader GD, Betel D, Hogue CWV: **BIND: the biomolecular interaction network database**. *Nucleic Acids Res.* 2003, **31**:248-250.

24. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, Orchard S, Vingron M, Roechert B, Roepstorff P, Valencia A *et al.*: **IntAct: an open source molecular interaction database**. *Nucleic Acids Res.* 2004, **32**:D452-D455.

25. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, Surendranath V, Niranjan V, Muthusamy B, Gandhi TKB, Gronborg M *et al.*: **Development of human protein reference database as an initial platform for approaching systems biology in humans**. *Genome Res.* 2003, **13**:2363-2371.

26. •• Baspinar A, Cukuroglu E, Nussinov R, Keskin O, Gursoy A: **PRISM: a web server and repository for prediction of protein-protein interactions and modeling their 3D complexes**. *Nucleic Acids Res.* 2014, **42**.

PRISM pushes the envelope with template based modeling requiring only the interface region to be similar while the rest of the protein could be different.

27. Lensink MF, Velankar S, Kryshtafovych A, Huang SY, Schneidman-Duhovny D, Sali A, Segura J, Fernandez-Fuentes N, Viswanath S, Elber R *et al.*: **Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: a CASP-CAPRI experiment**. *Proteins Struct. Funct. Bioinform.* 2016, **84(Suppl. 1)**:323-348.

28. Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A: **Critical assessment of methods of protein structure prediction (CASP)—round x**. *Proteins Struct. Funct. Bioinform.* 2014, **82**:1-6.

29. Davis FP, Braberg H, Shen MY, Pieper U, Sali A, Madhusudhan MS: **Protein complex compositions predicted by structural similarity**. *Nucleic Acids Res.* 2006, **34**:2943-2952.

30. Marti-Renom MA, Pieper U, Madhusudhan MS, Rossi A, Eswar N, Davis FP, Al-Shahrour F, Dopazo J, Sali A: **DBAli tools: mining the protein structure space**. *Nucleic Acids Res.* 2007, **35**.

31. Szilagyi A, Zhang Y: **Template-based structure modeling of protein–protein interactions**. *Curr. Opin. Struct. Biol.* 2014, **24**:10-23.

32. Zhang QC, Petrey D, Deng L, Qiang L, Shi Y, Thu CA, Bisikirska B, Lefebvre C, Accili D, Hunter T *et al.*: **Structure-based prediction of protein–protein interactions on a genome-wide scale**. *Nature* 2012, **490**:1-6.

33. Aloy P, Böttcher B, Ceulemans H, Leutwein C, Mellwig C, Fischer S, Gavin A-C, Bork P, Superti-Furga G, Serrano L *et al.*: **Structure-based assembly of protein complexes in yeast**. *Science* 2004, **303**:2026-2029.

34. Tuncbag N, Gursoy A, Nussinov R, Keskin O: **Predicting protein–protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM**. *Nat Protoc.* 2011, **6**:1341-1354.

35. Teyra J, Hawkins J, Zhu H, Pisabarro MT: **Studies on the inference of protein binding regions across fold space based on structural similarities**. *Proteins Struct. Funct. Bioinform.* 2011, **79**:499-508.

36. Kundrotas PJ, Zhu Z, Janin J, Vakser IA: **Templates are available to model nearly all complexes of structurally characterized proteins**. *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:9438-9441.

37. André I, Bradley P, Wang C, Baker D: **Prediction of the structure of symmetrical protein assemblies**. *Proc. Natl. Acad. Sci. U. S. A.* 2007, **104**:17656-17661.

38. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ: **Protein structure prediction via combinatorial assembly of sub-structural units**. *Bioinformatics*. 2003.

39. Inbar Y, Benyamini H, Nussinov R, Wolfson HJ, Honig B: **Prediction of multimolecular assemblies by multiple docking**. *J. Mol. Biol.* 2005, **349**:435-447.

40. Esquivel-Rodríguez J, Yang YD, Kihara D: **Multi-LZerD: multiple protein docking for asymmetric complexes**. *Proteins Struct. Funct. Bioinform.* 2012, **80**:1818-1833.

41. Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem AA, Aflalo C, Vakser IA: **Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques**. *Proc. Natl. Acad. Sci. U. S. A.* 1992, **89**:2195-2199.

42. • Padhorny D, Kazennov A, Zerbe BS, Porter KA, Xia B, Mottarella SE, Kholodov Y, Ritchie DW, Vajda S, Kozakov D: **Protein–protein docking by fast generalized Fourier transforms on 5D rotational manifolds**. *Proc. Natl. Acad. Sci.* 2016, **113**:E4286-E4293.

This method has improved the speed of FFT sampling without sacrificing on accuracy. The 5D rotational manifold problem has been attempted several times in the past with little success until now.

43. Norel R, Fischer D, Wolfson HJ, Nussinov R: **Molecular surface recognition by a computer vision-based technique**. *Protein Eng. Des. Sel.* 1994, **7**:39-46.

44. Bonvin AM: **Flexible protein–protein docking**. *Curr. Opin. Struct. Biol.* 2006, **16**:194-200.

45. Mashiach E, Nussinov R, Wolfson HJ: **FiberDock: flexible induced-fit backbone refinement in molecular docking**. *Proteins Struct. Funct. Bioinform.* 2010, **78**:1503-1519.

46. Rajamani D, Thiel S, Vajda S, Camacho CJ: **Anchor residues in protein–protein interactions**. *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**:11287-11292.

47. Smith GR, Sternberg MJE, Bates PA: **The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein–protein docking**. *J. Mol. Biol.* 2005, **347**:1077-1101.

48. Shapovalov MV, Dunbrack RL: **A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions**. *Structure* 2011, **19**:844-858.

49. Holm L, Sander C: **Fast and simple Monte Carlo algorithm for side chain optimization in proteins: application to model building by homology**. *Proteins* 1992, **14**:213-223.

50. Carter P, Lesk VI, Islam SA, Sternberg MJE: **Protein–protein docking using 3D-Dock in rounds 3, 4, and 5 of CAPRI**. *Proteins: Structure, Function and Genetics*. 2005:281-288.

51. Gray JJ, Moughon SE, Kortemme T, Schueler-Furman O, Misura K, Morozov AV, Baker D: **Protein–protein docking predictions for the CAPRI experiment**. *Proteins Struct. Funct. Genet.* 2003, **52**:118-122.

52. Comeau SR, Gatchell DW, Vajda S, Camacho CJ: **ClusPro: an automated docking and discrimination method for the prediction of protein complexes**. *Bioinformatics* 2004, **20**:45-50.

53. Tovchigrechko A, Vakser IA: **GRAMM-X public web server for protein–protein docking**. *Nucleic Acids Res.* 2006, **34**:W310-4.

54. Moont G, Gabb HA, Sternberg MJE: **Use of pair potentials across protein interfaces in screening predicted docked complexes**. *Proteins Struct. Funct. Genet.* 1999, **35**:364-373.

55. Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW: **HexServer: an FFT-based protein docking server powered by graphics processors**. *Nucleic Acids Res.* 2010, **38**: W445-W449.

56. Gabb HA, Jackson RM, Sternberg MJE: **Modelling protein docking using shape complementarity, electrostatics and biochemical information**. *J. Mol. Biol.* 1997, **272**:106-120.

57. Chen R, Li L, Weng Z: **ZDOCK: an initial-stage protein-docking algorithm**. *Proteins Struct. Funct. Genet.* 2003, **52**:80-87.

58. •• Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z: **ZDOCK server: interactive docking prediction of protein–protein complexes and symmetric multimers**. *Bioinformatics* 2014, **30**:1771-1773.

A consistently well performing method in many CAPRI experiments, now incorporating symmetry to account for multimers.

59. Vreven T, Pierce BG, Hwang H, Weng Z: **Performance of ZDOCK in CAPRI rounds 20–26**. *Proteins Struct. Funct. Bioinform.* 2013, **81**:2175-2182.

60. Pierce B, Tong W, Weng Z: **M-ZDOCK: a grid-based approach for Cn symmetric multimer docking**. *Bioinformatics* 2005, **21**:1472-1478.

61. Kozakov D, Brenke R, Comeau SR, Vajda S: **PIPER: an FFT-based protein docking program with pairwise potentials**. *Proteins Struct. Funct. Genet.* 2006, **65**:392-406.

62. Ohue M, Shimoda T, Suzuki S, Matsuzaki Y, Ishida T, Akiyama Y: **MEGADOCK 4.0: an ultra-high-performance protein–protein docking software for heterogeneous supercomputers**. *Bioinformatics* 2014, **30**:3281-3283.

63. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ: **PatchDock and SymmDock: servers for rigid and symmetric docking**. *Nucleic Acids Res.* 2005, **33**:W363-7.

64. Duhovny D, Nussinov R, Wolfson HJ: **Efficient unbound docking of rigid molecules**. In *Algorithms in Bioinformatics. WABI 2002. Lecture Notes in Computer Science*, 2452. Edited by Guigó R, Gusfield D . Berlin, Heidelberg: Springer; 2017:185-200.

65. Lyskov S, Gray JJ: **The RosettaDock server for local protein–protein docking**. *Nucleic Acids Res* 2008:36.

66. • Zhang Z, Schindler CEM, Lange OF, Zacharias M: **Application of enhanced sampling Monte Carlo methods for high-resolution protein–protein docking in Rosetta**. *PLoS One* 2015, **10**: e0125941.

This is a non-FFT based sampling method and uses a replica exchange Monte Carlo approach to efficiently dock pairs of proteins.

67. •• Ramírez-Aportela E, López-Blanco JR, Chacón P: **FRODOCK 2.0: fast protein–protein docking server**. *Bioinformatics* 2016: btw141.

FRODOCK's sampling uses spherical harmonics and is a faster alternative to ZDOCK and Swarmdock.

68. Garzon JI, Lopéz-Blanco JR, Pons C, Kovacs J, Abagyan R, Fernandez-Recio J, Chacon P: **FRODOCK: a new approach for fast rotational protein–protein docking**. *Bioinformatics* 2009, **25**:2544-2551.

69. Tsai CJ, Lin SL, Wolfson HJ, Nussinov R: **Studies of protein–protein interfaces: a statistical analysis of the hydrophobic effect**. *Protein Sci.* 1997, **6**:53-64.

70. Xu D, Lin SL, Nussinov R: **Protein binding versus protein folding: the role of hydrophilic bridges in protein associations**. *J. Mol. Biol.* 1997, **265**:68-84.

71. Norel R, Petrey D, Wolfson HJ, Nussinov R: **Examination of shape complementarity in docking of Unbound proteins**. *Proteins Struct. Funct. Genet.* 1999, **36**:307-317.

72. Li X, Moal IH, Bates PA: **Detection and refinement of encounter complexes for protein–protein docking: taking account of macromolecular crowding**. *Proteins Struct. Funct. Bioinform.* 2010, **78**:3189-3196.

73. May A, Zacharias M: **Protein–protein docking in CAPRI using ATTRACT to account for global and local flexibility**. *Proteins: Structure, Function and Genetics*. 2007:774-780.

74. Zacharias M: **Protein–protein docking with a reduced protein model accounting for side-chain flexibility**. *Protein Sci. Publ. Protein Soc.* 2003, **12**:1271-1282.

75. Smith GR, Sternberg MJE: **Evaluation of the 3D-Dock protein docking suite in rounds 1 and 2 of the CAPRI blind trial**. *Proteins Struct. Funct. Genet.* 2003, **52**:74-79.

76. Venkatraman V, Yang YD, Sael L, Kihara D: **Protein–protein docking using region-based 3D Zernike descriptors**. *BMC Bioinform.* 2009, **10**:407.

77. De Vries SJ, Van Dijk ADJ, Krzeminski M, Van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AMJJ: **HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets**. *Proteins: Structure, Function and Genetics*. 2007:726-733.

78. Dominguez C, Boelens R, Bonvin AMJJ: **HADDOCK: a protein–protein docking approach based on biochemical or biophysical information**. *J. Am. Chem. Soc.* 2003, **125**:1731-1737.

79. de Vries SJ, van Dijk M, Bonvin AMJJ: **The HADDOCK web server for data-driven biomolecular docking**. *Nat. Protoc.* 2010, **5**:883-897.

80. Torchala M, Moal IH, Chaleil RAG, Fernandez-Recio J, Bates PA: **SwarmDock: a server for flexible protein–protein docking**. *Bioinformatics* 2013, **29**:807-809.

81. Malhotra S, Mathew OK, Sowdhamini R: **DOCKSCORE: a webserver for ranking protein–protein docked poses**. *BMC Bioinform.* 2015, **16**:127.

82. Zhu X, Mitchell JC: **KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features**. *Proteins Struct. Funct. Bioinform.* 2011, **79**:2671-2683.

83. Cho K-i, Kim D, Lee D: **A feature-based approach to modeling protein–protein interaction hot spots**. *Nucleic Acids Res.* 2009, **37**:2672-2687.

84. Tuncbag N, Keskin O, Gursoy A: **HotPoint: hot spot prediction server for protein interfaces**. *Nucleic Acids Res.* 2010, **38**:W402-W406.

85. Sukhwal A, Sowdhamini R: **PPcheck: a webserver for the quantitative analysis of protein–protein interfaces and prediction of residue hotspots**. *Bioinform. Biol. Insights* 2015, **9**:141-151.

86. Aumentado-Armstrong TT, Istrate B, Murgita RA, Nussinov R, Ponty Y, Carbone A, Langer T, Wolber G: **Algorithmic approaches to protein–protein interaction site prediction**. *Algorithms Mol. Biol.* 2015, **10**:7.

87. Kastritis PL, Bonvin AMJJ: **Are scoring functions in protein–protein docking ready to predict interactomes? Clues from a novel binding affinity benchmark**. *J. Proteome Res.* 2010, **9**:2216-2225.

88. Vreven T, Hwang H, Pierce BG, Weng Z: **Prediction of protein–protein binding free energies**. *Protein Sci.* 2012 http://dx.doi.org/10.1002/pro.2027.

89. • Parikh HI, Kellogg GE: **Intuitive, but not simple: including explicit water molecules in protein–protein docking simulations improves model quality**. *Proteins Struct. Funct. Bioinform.* 2014, **82**:916-932.

The role of water in some interactions could be significant and here it is explictly considered during the docking process.

90. Steven AC, Baumeister W: **The future is hybrid**. *J. Struct. Biol.* 2008, **163**:186-195.

91. Lasker K, Phillips JL, Russel D, Velázquez-Muriel J, Schneidman-Duhovny D, Tjioe E, Webb B, Schlessinger A, Sali A: **Integrative structure modeling of macromolecular assemblies from proteomics data**. *Mol. Cell. Proteomics* 2010, **9**:1689-1702.

92. Schröder GF: **Hybrid methods for macromolecular structure determination: experiment with expectations**. *Curr. Opin. Struct. Biol.* 2015, **31**:20-27.

93. 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.

94. Alber F, Förster F, Korkin D, Topf M, Sali A: **Integrating diverse data for structure determination of macromolecular assemblies**. *Annu. Rev. Biochem.* 2008, **77**:443-477.

95. Topf M, Sali A: **Combining electron microscopy and comparative protein structure modeling**. *Curr. Opin. Struct. Biol.* 2005, **15**:578-585.

96. 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.

97. Lasker K, Sali A, Wolfson HJ: **Determining macromolecular assembly structures by molecular docking and fitting into an electron density map**. *Proteins Struct. Funct. Bioinform.* 2010, **78**:3205-3211.

98. Topf M, Baker ML, John B, Chiu W, Sali A: **Structural characterization of components of protein assemblies by comparative modeling and electron cryo-microscopy**. *J. Struct. Biol.* 2005, **149**:191-203.

99. Trabuco LG, Villa E, Mitra K, Frank J, Schulten K: **Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics**. *Structure* 2008, **16**:673-683.

100. Zhang S, Vasishtan D, Xu M, Topf M, Alber F: **A fast mathematical programming procedure for simultaneous fitting of assembly components into cryoEM density maps**. *Bioinformatics* 2010:26.

101. Kuzu G, Keskin O, Nussinov R, Gursoy A: **PRISM-EM: template interface-based modelling of multi-protein complexes guided by cryo-electron microscopy density maps**. *Acta Crystallogr. Sect. D Struct. Biol.* 2016, **72**:1137-1148.

102. •• Fernandez-Martinez J, Kim SJ, Shi Y, Upla P, Pellarin R, Gagnon M, Chemmama IE, Wang J, Nudelman I, Zhang W *et al.*: **Structure and function of the nuclear pore complex cytoplasmic mRNA export platform**. *Cell* 2016, **167** 1215–1228. e25.

An application of the IMP method for integrative modeling, assimilating data from chemical cross linking, mass spectrometry, EM, SAXS and computational modeling.

103. •• Grinter R, Josts I, Mosbahi K, Roszak AW, Cogdell RJ, Bonvin AMJJ, Milner JJ, Kelly SM, Byron O, Smith BO *et al.*: **Structure of the bacterial plant-ferredoxin receptor FusA**. *Nat. Commun.* 2016, **7**:13308.

HADDOCK, a consistently superior performer at CAPRI, used here in conjunction with X-ray crystallograpy and NMR.

104. • Short JM, Liu Y, Chen S, Soni N, Madhusudhan MS, Shivji MK, Venkitaraman AR: **High-resolution structure of the presynaptic RAD51 filament on single-stranded DNA by electron cryo-microscopy**. *Nucleic Acids Res.* 2016, **44**:9017-9030.

An example of how with template based modeling one could deduce alternate conformations of experimentally determined structures.

105. Manjunath GP, Soni N, Vaddavalli PL, Shewale DJ, Madhusudhan MS, Muniyappa K: **Molecular mechanism underlying ATP-induced conformational changes in the nucleoprotein filament of *Mycobacterium smegmatis* RecA**. *Biochemistry* 2016, **55**:1850-1862.

106. 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**:174-180.

107. Dutta S, Berman HM: **Large macromolecular complexes in the protein data bank: a status report**. *Structure* 2005, **13**:381-388.

108. Nogales E: **The development of cryo-EM into a mainstream structural biology technique**. *Nat. Meth.* 2016, **13**:24-27.

109. Janin J: **Protein–protein docking tested in blind predictions: the CAPRI experiment**. *Mol. Biosyst.* 2010, **6**:2351-2362.

110. Janin J, Henrick K, Moult J, Ten Eyck L, Sternberg MJE, Vajda S, Vakser I, Wodak SJ: **CAPRI: a critical assessment of PRedicted interactions**. *Proteins Struct. Funct. Genet.* 2003, **52**:2-9.

111. Janin J: **Assessing predictions of protein–protein interaction: the CAPRI experiment**. *Protein Sci.* 2005, **14**:278-283.

112. Lensink MF, Wodak SJ: **Docking, scoring, and affinity prediction in CAPRI**. *Proteins Struct. Funct. Bioinform.* 2013, **81**:2082-2095.

113. Méndez R, Leplae R, Lensink MF, Wodak SJ: **Assessment of CAPRI predictions in Rounds 3–5 shows progress in docking procedures**. *Proteins: Structure, Function and Genetics*. 2005:150-169.

114. Méndez R, Leplae R, De Maria L, Wodak SJ: **Assessment of blind predictions of protein–protein interactions: current status of docking methods**. *Proteins Struct. Funct. Genet.* 2003, **52**:51-67.

115. Lensink MF, Méndez R, Wodak SJ: **Docking and scoring protein complexes: CAPRI**. *Proteins: Structure, Function and Genetics*. 3rd Edition. 2007:704-718.

116. Sahtoe DD, vanDijk WJ, ElOualid F, Ekkebus R, Ovaa H, Sixma TK: **Mechanism of UCH-L5 activation and inhibition by DEUBAD domains in RPN13 and INO80G**. *Mol. Cell* 2015, **57**:887-900.

117. Ghosh A, Praefcke GJK, Renault L, Wittinghofer A, Herrmann C: **How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP**. *Nature* 2006, **440**:101-104.

118. LoPiccolo J, Kim SJ, Shi Y, Wu B, Wu H, Chait BT, Singer RH, Sali A, Brenowitz M, Bresnick AR *et al.*: **Assembly and molecular architecture of the phosphoinositide 3-kinase p85α homodimer**. *J. Biol. Chem.* 2015, **290**:30390-30405.

119. Robinson PJ, Trnka MJ, Pellarin R, Greenberg CH, Bushnell DA, Davis R, Burlingame AL, Sali A, Kornberg RD: **Molecular architecture of the yeast Mediator complex**. *Elife* 2015:4.

120. Baù D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, Dekker J, Marti-Renom MA: **The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules**. *Nat. Struct. Mol. Biol.* 2011, **18**:107-114.

121. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L: **Genome architectures revealed by tethered chromosome conformation capture and population-based modeling**. *Nat. Biotechnol.* 2012, **30**:90-98.

122. Thalassinos K, Pandurangan AP, Xu M, Alber F, Topf M: **Conformational states of macromolecular assemblies explored by integrative structure calculation**. *Structure* 2013, **21**:1500-1508.

123. Zhou XJ, Alber F: **Zooming in on genome organization**. *Nat. Methods* 2012, **9**:961-963.

124. Luo J, Cimermancic P, Viswanath S, Ebmeier CC, Kim B, Dehecq M, Raman V, Greenberg CH, Pellarin R, Sali A *et al.*: **Architecture of the human and yeast general transcription and DNA repair factor TFIIH**. *Mol. Cell* 2015, **59**:794-806.

125. Choudhury AR, Sikorska E, Van Den Boom J, Bayer P, Popenda , Szutkowski K, Jurga S, Bonomi M, Sali A, Zhukov I *et al.*: **Structural model of the bilitranslocase transmembrane domain supported by NMR and FRET data**. *PLoS One* 2015:10.

126. Politis A, Schmidt C, Tjioe E, Sandercock AM, Lasker K, Gordiyenko Y, Russel D, Sali A, Robinson CV: **Topological models of heteromeric protein assemblies from mass spectrometry: application to the yeast eIF3:eIF5 complex**. *Chem. Biol.* 2015, **22**:117-128.

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

128. Querol-Audí J, Yan C, Xu X, Tsutakawa SE, Tsai M-S, Tainer JA, Cooper PK, Nogales E, Ivanov I: **Repair complexes of FEN1 endonuclease, DNA, and Rad9-Hus1-Rad1 are distinguished from their PCNA counterparts by functionally important stability**. *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:8528-8533.

129. Tsutakawa SE, Van Wynsberghe AW, Freudenthal BD, Weinacht CP, Gakhar L, Washington MT, Zhuang Z, Tainer JA,

Ivanov I: **Solution X-ray scattering combined with computational modeling reveals multiple conformations of covalently bound ubiquitin on PCNA**. *Proc. Natl. Acad. Sci. U. S. A.* 2011, **108**:17672-17677.

130. Zeng-Elmore X, Gao XZ, Pellarin R, Schneidman-Duhovny D, Zhang XJ, Kozacka KA, Tang Y, Sali A, Chalkley RJ, Cote RH *et al.*: **Molecular architecture of photoreceptor phosphodiesterase elucidated by chemical cross-linking and integrative modeling**. *J. Mol. Biol.* 2014, **426**:3713-3728.

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

132. Algret R, Fernandez-Martinez J, Shi Y, Kim SJ, Pellarin R, Cimermancic P, Cochet E, Sali A, Chait BT, Rout MP *et al.*: **Molecular architecture and function of the SEA complex, a modulator of the TORC1 pathway**. *Mol. Cell. Proteomics* 2014, **13**:2855-2870.

133. Erzberger JP, Stengel F, Pellarin R, Zhang S, Schaefer T, Aylett CHS, Cimermančič P, Boehringer D, Sali A, Aebersold R *et al.*: **Molecular architecture of the 40S·eIF1·eIF3 translation initiation complex**. *Cell* 2014, **158**:1123-1135.

134. Lasker K, Forster F, Bohn S, Walzthoeni T, Villa E, Unverdorben P, Beck F, Aebersold R, Sali A, Baumeister W *et al.*: **Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach**. *Proc. Natl. Acad. Sci. U. S. A.* 2012, **109**:1380-1387.

135. Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprapto A, Karni-Schmidt O, Williams R, Chait BT *et al.*: **The molecular architecture of the nuclear pore complex**. *Nature* 2007, **450**:695-701.

136. Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, Schneidman-Duhovny D, Sali A, Rout MP, Chait BT: **Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex**. *Mol. Cell. Proteomics* 2014, **13**:2927-2943.

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

138. Politis A, Stengel F, Hall Z, Hernández H, Leitner A, Walzthoeni T, Robinson CV, Aebersold R: **A mass spectrometry-based hybrid method for structural modeling of protein complexes**. *Nat. Methods* 2014, **11**:403-406.

139. Taylor DJ, Devkota B, Huang AD, Topf M, Narayanan E, Sali A, Harvey SC, Frank J: **Comprehensive Molecular Structure of the Eukaryotic Ribosome**. *Structure* 2009, **17**:1591-1604.

140. Chandramouli P, Akey CW, Topf M, Ménétret J, Eswar N, Gutell RR, Sali A: **Structure of the mammalian 80S Ribosome at 8.7A resolution**. *Structure* 2008, **16**:535-548.

141. Campos M, Nilges M, Cisneros DA, Francetic O: **Detailed structural and assembly model of the type II secretion pilus from sparse data**. *Proc. Natl. Acad. Sci. U. S. A.* 2010, **107**:13081-13086.

142. Campos M, Francetic O, Nilges M: **Modeling pilus structures from sparse data**. *J. Struct. Biol.* 2011, **173**:436-444.

143. Serysheva II, Ludtke SJ, Baker ML, Cong Y, Topf M, Eramian D, Sali A, Hamilton SL, Chiu W: **Subnanometer-resolution electron cryomicroscopy-based domain models for the cytoplasmic region of skeletal muscle RyR channel**. *Proc. Natl. Acad. Sci. U. S. A.* 2008, **105**:9610-9615.

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

145. Greenberg CH, Kollman J, Zelter A, Johnson R, MacCoss MJ, Davis TN, Agard DA, Sali A: **Structure of γ-tubulin small complex based on a cryo-EM map, chemical cross-links, and a remotely related structure**. *J. Struct. Biol.* 2016, **194**:303-310.

146. Schönegge AM, Villa E, Förster F, Hegerl R, Peters J, Baumeister W, Rockel B: **The structure of human tripeptidyl peptidase II as determined by a hybrid approach**. *Structure* 2012, **20**:593-603.

147. Zelter A, Bonomi M, Kim ook J, Umbreit NT, Hoopmann MR, Johnson R, Riffle M, Jaschob D, MacCoss MJ, Moritz RL *et al.*: **The molecular architecture of the Dam1 kinetochore complex is defined by cross-linking based structural modelling**. *Nat. Commun.* 2015, **6**:8673.