Docking approaches for modeling multi-molecular assemblies

Mireia Rosell[1,2] and Juan Fernández-Recio[1,2]

Computational docking approaches aim to overcome the limited availability of experimental structural data on protein–protein interactions, which are key in biology. The field is rapidly moving from the traditional docking methodologies for modeling of binary complexes to more integrative approaches using template-based, data-driven modeling of multi-molecular assemblies. We will review here the predictive capabilities of current docking methods in blind conditions, based on the results from the most recent community-wide blind experiments. Integration of template-based and *ab initio* docking approaches is emerging as the optimal strategy for modeling protein complexes and multimolecular assemblies. We will also review the new methodological advances on *ab initio* docking and integrative modeling.

Addresses
[1] Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain
[2] Instituto de Ciencias de la Vid y del Vino (ICVV), CSIC - Universidad de La Rioja - Gobierno de La Rioja, 26007 Logroño, Spain

Corresponding author:
Fernández-Recio, Juan (juan.fernandezrecio@icvv.es)

Introduction

Protein–protein interactions are key for the majority of biological functions. Proteins can form highly specific transient or permanent complexes that range from binary pairs to multi-molecular assemblies, often involving other biomolecules. A detailed structural knowledge of such complexes at atomic level would improve our understanding of biological processes and facilitate intervention for biomedical and biotechnological purposes. For example, recently reported structural data on the dynamic assembly formed by the SARS-CoV-2 trimeric spike protein and the cell receptor ACE2 are key to understand the molecular mechanisms of the virus infectivity and can be essential for the development of new vaccines and therapeutic candidates against COVID-19 [1••,2••,3]. However, structural data is available for only a small fraction of the protein interactome. For instance, the total number of protein–protein interactions in human is estimated to range from 130 000 [4] to 650 000 [5], but less than 7000 of these interactions have available 3D structure (Interactome3D, 2019_1 version) [6]. In this context, computational docking approaches aim to overcome the limited availability of experimental structural data. Since the first reported protein-protein docking algorithms in the early 90's, based on Fast Fourier Transform (FFT) sampling [7], the methodological developments have mostly focused on *ab initio* docking of binary complexes, starting from the structure of the unbound components. Some of the most popular methods are FTDock [8], ZDOCK [9] or MolFit [10]. The method HEX [11] and later FRODOCK [12] used polar Fourier correlations to accelerate docking calculations. Other different approaches using stochastic search based on global-energy optimization are ICM-DISCO [13,14], Rosetta-Dock [15], HADDOCK [16], or SwarmDock [17].

With the increasing availability of complex structures, in recent years attention is focused on template-based structural modeling of complexes, based the standard principles of homology-based modeling. The term template-based docking (as opposed to *ab initio* docking) is specifically used when a model is built by superimposing the structures (or models) of the unbound subunits onto the corresponding subunits of a template complex structure [18]. One advantage is that template-based modeling can be applied to multi-molecular complexes, not just to binary complexes as *ab initio* docking. In addition, it has been suggested that templates are available for the large majority of cases in which interacting subunits have structural information [19]. However, the general availability of good-quality templates that could be reliable used for template-based predictions seems much lower [20•]. Actually, for the majority of known interactions, only templates with remote homology are available [4], for which direct application of template-based methods leads to poor predictions [21]. Modeling multi-molecular assemblies implies additional challenges. For instance, some of the interfaces might not have available templates, in which case, we could model them by *ab initio* docking, in combination with restraints from evolutionary data or from available experimental information. Another challenge is to identify the relevant oligomerization state of the assembly when is different from that in the template

[22], in which case, alternative orientations provided by *ab initio* docking can be very helpful. Modeling the conformational variability of the assembly components imposes an additional difficulty. Indeed, directly taking the structure of a given subunit in another context (e.g. unbound state, different assembly or alternative oligomerization state) might lead to inaccurate models. For this, it can be useful the application of protein-protein docking and associated procedures, such as energy scoring, minimization, or flexible refinement.

We will review here the predictive capabilities of current protein-protein docking methods in blind conditions, based on the results from the most recent CASP [23••] and CAPRI [24•] experiments. These tests show that combination of template-based and *ab initio* docking approaches is emerging as the optimal strategy for modeling protein complexes and multimolecular assemblies. We will also review the most recent methodological novelties on *ab initio* docking, and new approaches for the inclusion of experimental information and integrative modeling.

Predictive capabilities of computational docking: the state-of-the-art

Ab initio computational docking can provide acceptable models within the top 10 predictions in up to 40% of the cases, according to reported evaluation studies of different methodologies in current protein–protein docking benchmark version 5.0 [20•,25,26].

Traditionally CASP has been focused on the prediction of the structure of individual proteins. However, very often proteins are found as oligomeric assemblies, which adds complexity to the modeling effort. To evaluate the applicability of docking methodologies for the prediction of protein oligomeric assemblies, the last three CASP editions included a CASP-CAPRI joint experiment focused on multimeric assemblies, which are independently evaluated by CASP and CAPRI communities. The recent CASP13-CAPRI challenge comprised a total of 20 protein oligomeric assemblies, including 14 homo-complexes and 6 hetero-complexes, which could be classified into 15 dimers and 5 multimeric assemblies [23••]. In the 9 'easy' targets, there were good structural templates for the (partial or full) assembly, while for some of the remaining 11 'difficult' targets, it was possible to find remote templates for part of the assembly. The availability of templates in each case is critical to explain the predictive success of the groups. Focusing on the results for the top 10 predictions (to facilitate comparison with the reported performances of different docking methods in the literature), the best-performing group submitted acceptable (or better) models for 13 targets (65% of the cases) (Figure 1). In the 'easy' targets, the best-performing group submitted acceptable models for all these cases, while in the 'difficult' targets, the

best-performing group submitted acceptable models for only 4 of such targets (36% of the cases). Regarding the quality of the models, high-quality models [23••] were submitted by any group in 78% of the 'easy' targets (with template), but only in 9% of the 'difficult' targets (no template).

On the other side, the recent 7th CAPRI edition showed more heterogeneity in its targets, comprising 8 protein-protein, 3 protein-peptide, and 5 protein-oligosaccharide complexes, all hetero-oligomers (except for a homo-decamer), which could be classified in 10 dimers and 6 multimeric assemblies [24•]. The actual number of evaluated targets was 19, because some of the interfaces in these multimeric assemblies were considered as independent targets. There were structural templates for a total of 13 target interfaces (6 protein–protein, 2 protein–peptide, and 5 protein–saccharide). This was determinant for the overall predictive success of the groups as well as for the quality of the predicted models. Overall, the maximum number of target interfaces successfully predicted by a single group was 13 (i.e. success in 68% of the cases) (Figure 1). But in cases with no available template, the best-performing groups submitted acceptable models for only 2 target interfaces (i.e. success in 33% of the cases). Regarding the quality of the models, high-quality models [24•] were submitted by any group in 31% of the 'easy' targets (with template) and in 17% of the 'difficult' targets (no template). The 7th CAPRI edition showed that *ab initio* docking in cases for which there is no available template is still highly challenging, and progress is actually coming from the efficient procedures to combine template-based modeling and other docking methodologies.

Combination of template-based and *ab initio* docking

The CASP and CAPRI experiments show that template-based modeling approaches are clearly the tools of choice when one can use templates of sufficient quality. However, very often only remote templates are available, which might not be good enough to provide reliable models, as above discussed [21]. In unclear situations, a relevant question is which method to choose, or how to efficiently combine these protein-protein docking approaches depending on each specific case [20•]. This is even more relevant when modeling multimeric complexes, in which some interfaces might be modelled based on homologous structures, while others would need *ab initio* docking, as above mentioned. An updated version of the InterEvDock2 server [27••] can perform template-based docking or *ab initio* docking with evolutionary constraints, depending on the case. But the question is still open about how to efficiently combine template-based and *ab initio* docking when reliability of the template is unclear. We can obtain some hints from the recent CASP and CAPRI experiments.

Figure 1
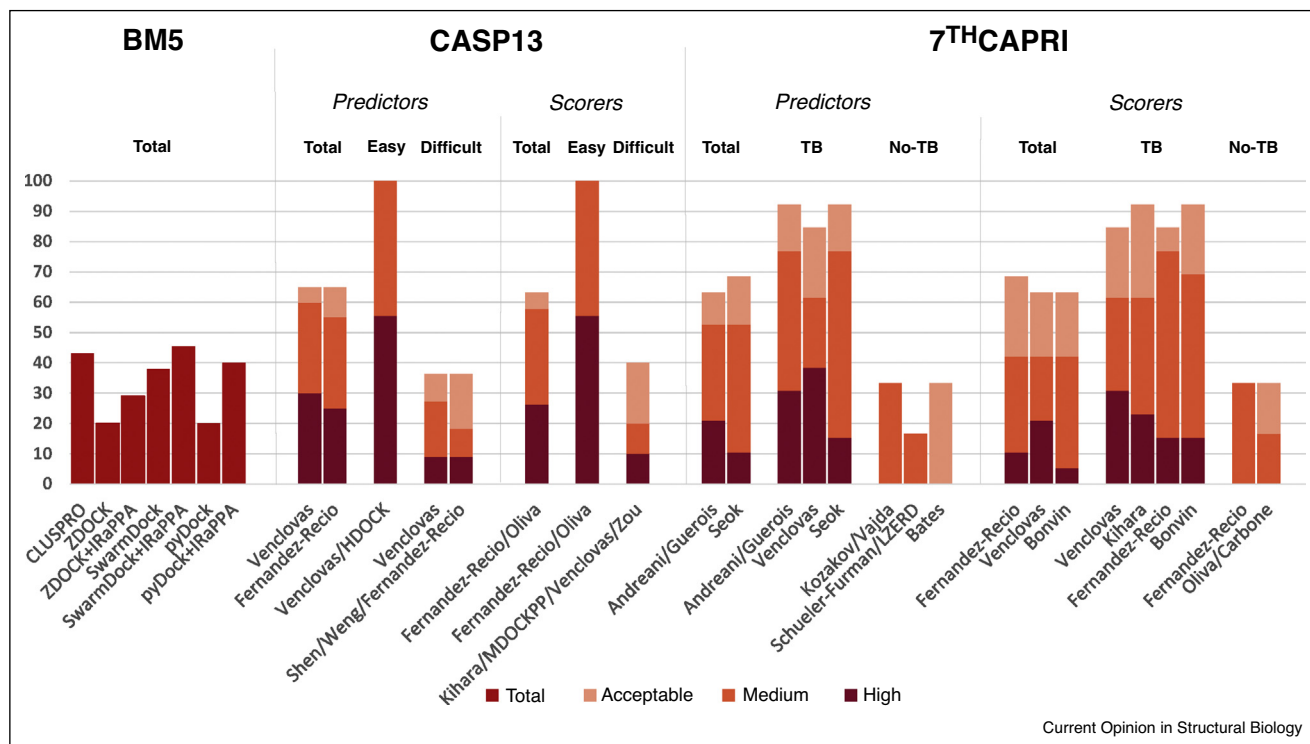
Predictive success rates of state-of-the-art docking approaches on different benchmark sets. ClusPro performance on Protein–Protein Docking Benchmark 5.0 (BM5) is taken from Ref. [20•]. Performance of other docking methods on BM5 is taken from Ref. [26]. The rest of results are taken from CASP13-CAPRI and 7th CAPRI blind experiments. Sampling and scoring strategies included, but were not limited to: FFT-based sampling (*ZDOCK, FTDock, ClusPro, Weng, Kozakov/Vajda, Shueler-Furman, Venclovas, pyDock, Fernandez-Recio, HDOCK, MDockPP, Zou, Shen, Seok*), geometric hashing (*Kihara, LZerD*), particle swarm optimization (*Bates*), NMA-based sampling (*Shen, Bates*), information-driven sampling (*Bonvin*), energy-based scoring (*pyDock, Fernandez-Recio*), machine learning-based scoring (*IRaPPA, Shen*), statistical potentials (*Kihara, MDockPP, Zou*), evolutionary-based scoring (*Andreani/Guerois*), Voronoi-based scoring (*Venclovas*), shape-based scoring (*HDOCK*), docking-based contact consensus and residue propensities (*Oliva, Carbone*), and flexible refinement (*Shueler-Furman, Seok*).

In the recent CASP13-CAPRI joint assembly prediction experiment, one of the most efficient approaches was that of Fernández-Recio, based on a combination of template-based and *ab initio* docking followed by pyDock scoring [23••], which ranked 2nd and 1st among all the CAPRI predictors and scorers groups, respectively. Models for the subunits were built by CASP-hosted servers. Then, *ab initio* docking was applied in all cases, using appropriate symmetry constraints or interface restraints from literature. Additionally, when reliable templates were found, template-based models were built by superimposing all possible models of the monomers onto them. After sorting all built models by pyDock scoring, the proportion of template-based and *ab initio* docking models in the final set of submitted models depended on the reliability of the templates (Figure 2). The difference with other methodologies was more evident on the 'difficult' cases for which no clear template was available. For instance, in T154 *ab initio* docking by pyDock produced the only acceptable models among all participants. In T157, pyDock also produced some of the few successful models of all groups.

For scorers, pyDock was used to evaluate all the proposed models, and in case of reliable templates, consistency between energy-based scoring and template-based data was sought.

In 7th CAPRI, predictions using template information were in general successful. Indeed, failing to use available templates, as Fernández-Recio did in T122, T125 interface 1/4, and T133 targets, led to much worse predictions (although interestingly, this group was successful in the latter target, using only *ab initio* docking). This shows that it is critical to choose the optimal docking approach for each case, depending on the template availability. In the rest of targets, templates were used indirectly. In the two protein-peptide targets with good templates (T134, T135), *ab initio* docking with pyDock with restraints from the available templates was successful. In the six protein-saccharide targets (T126-130), *ab initio* docking on the cavity identified from the available templates was also successful. These represent alternative strategies to combine *ab initio* docking with template information.

Figure 2
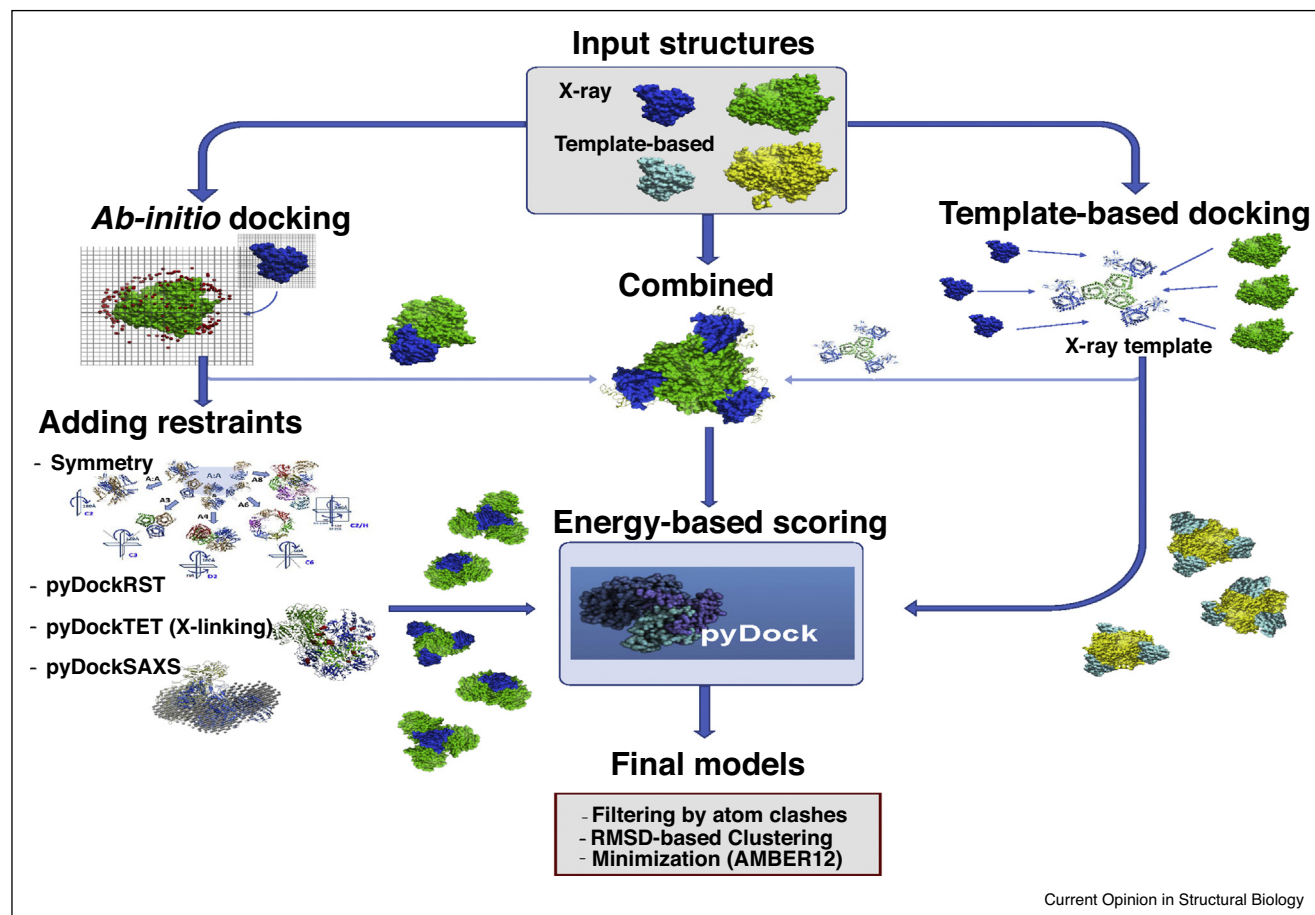
An example of the combination of template-based, *ab initio* docking and external data for integrative modeling of complexes. The scheme is based on the strategy followed by our group (*Fernandez-Recio*) as predictors in the recent CASP13-CAPRI and 7th CAPRI experiments.

Finally, in the scorers experiment, pyDock got the best performance when considering top 10 predictions, which shows its capabilities to evaluate complex models derived from combined approaches (template-based, *ab initio*, refinement) [24•] (Figure 1).

Novel methodological developments in protein docking

The most successful approach as predictor in CASP13-CAPRI was that of Venclovas group. They basically used template-based models when reliable templates were found, and free docking with HEX [11] otherwise. One of the reasons of their success could be the use of VoroMQA [28] for the evaluation and selection of the final models. However, they were less efficient in the scorers experiment (rank 7th), which might indicate that this function seems mostly optimized for their own pipeline for template-based and docking generation, while its application to models generated by other sources represents a challenge to be solved. Other successful

approach was the use of CONSRANK [29,30] for the ranking of docking models. CONSRANK is based on the most frequent inter-residue contacts in the ensemble of decoys, and has been updated to Clust-CONSRANK with the addition of a recently developed clustering procedure [31]. The best-performing server in CASP13-CAPRI was HDOCK [32], from Huang's group, who developed a new pairwise shape-based scoring function (LSC) for protein–protein docking to take into account long-range interactions between protein atoms [33•].

Other recent new developments in protein docking are RosettaDock 4.0, which shows improved predictions for flexible cases [34•], LightDock, using glowworm swarm optimization with NMA-based flexible search [35], or CIPS, a new scoring procedure [36•] based on interface propensities from docking calculations. Docking interface propensities have interesting applications, such as interface prediction [37], and more recently, characterization

Use of external information for integrative docking

The identification of correct docking poses often fails due to intrinsic errors in current scoring functions, incorrect consideration of oligomerization states, or because of multiple interfaces that are not usually included in docking calculations. For all these reasons, the use of external information on a given complex is often critical for successful docking predictions. The pioneering HADDOCK [16], as well as other protein–protein docking methods, such as pyDock [39], ZDOCK [40] or LightDock [41] have developed procedures to include distance restraints to improve the docking calculations. In this line, evolutionary information can be a relevant source of information for docking [42]. Indeed, the most successful docking approach in the recent 7th CAPRI edition was that of the Andreani and Guerois group. The challenging cases of this CAPRI edition encouraged them to go beyond their traditional rigid-body and InterEvScore approach, so they applied different strategies for the inclusion of evolutionary constraints, such as template-based modeling with RosettaCM-based protocol [43], identification of conserved anchoring interface motifs when only remote homologs were available, and covariation-based modeling of interacting subunits in cases in which traditional homology-based modeling would fail [44••].

In a broader sense, integrative computational approaches that aim to efficiently use experimental structural data and additional information from a variety of sources for the structural modeling of complexes are becoming increasingly popular [45]. One example is the integration of Small-Angle X-ray Scattering (SAXS) experimental data in *ab initio* docking methods such as pyDock [46–48], HADDOCK [49], PatchDock [50,51], ATTRACT [52] or ClusPro [53]. And chemical cross-linking data has also been integrated in protein docking methods such as ZDOCK [54]. In the 7th CAPRI experiment, the use of integrative modeling approaches was blindly evaluated. Targets T150 and T151 were the same complex as T149, a challenging multi-domain dimer, for which SAXS and chemical cross-linking data were provided, respectively. Interestingly, the inclusion of restraints from SAXS data improved the models submitted by pyDock for the original target (with few successful groups), and the cross-linking data further improved pyDock submissions [55].

Conclusions

The most recent community-wide blind tests on the structural prediction of multi-molecular assemblies and heteromeric protein complexes (including interaction with peptides and saccharides) clearly showed that template availability, as well as any additional information on the complex, are critical for the modeling success. Several groups are focusing their efforts on developing new procedures for efficient integration of template-based and evolutionary information with *ab initio* docking methods, which are producing more accurate and realistic models. Additional methodological developments on protein docking include improvement of scoring functions, and better treatment of conformational flexibility during docking search, but the field is clearly moving towards an integrative analysis and modeling of protein complexes.

Conflict of interest statement

Nothing declared.

Acknowledgements

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. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q,
•• Zhang L, Wang X: **Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor**. *Nature* 2020, **581**:215-220
This manuscript reveals the 3D crystal structure of the SARS-CoV-2 spike receptor-binding domain (RBD) bound to the cell receptor ACE2. The structural comparison with SARS-CoV RBD-ACE2 complex helps to identify the critical residues for ACE2 binding, and bring new insights into convergent evolution between the SARSCoV-2 and SARS-CoV RBDs for improved binding to ACE2. They also structurally analyze the epitopes of two SARS-CoV antibodies targeting the RBD, providing insights into the future identification of cross-reactive antibodies.

2. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D:
•• **Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein**. *Cell* 2020, **181**:281-292
This manuscript describes the cryo-EM structure of SARS-CoV-2 spike ectodomain trimer in two different conformational states, closed and partially open (one S^B domain open), providing new insights for the design of vaccines and inhibitors of viral entry. They found that the SARSCoV-2 S glycoprotein harbors a furin cleavage site at the boundary between the S1/S2 subunits, which is not present in SARS-CoV and other SARS-related CoVs, and can be relevant to understand the virus virulence.

3. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS: **Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation**. *Science* 2020, **367**:1260-1263.

4. Venkatesan K, Rual J-F, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh K-I *et al.*: **An empirical framework for binary interactome mapping**. *Nat Methods* 2009, **6**:83-90.

5. Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C: **Estimating the size of the human interactome**. *Proc Natl Acad Sci U S A* 2008, **105**:6959-6964.

6. Mosca R, Ceol A, Aloy P: **Interactome3D: adding structural details to protein networks**. *Nat Methods* 2013, **10**:47-53.

7. 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.

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

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

10. Redington PK: **MOLFIT: a computer program for molecular superposition**. *Comput Chem* 1992, **16**:217-222.

11. Ritchie DW, Kemp GJL: **Protein docking using spherical polar Fourier correlations**. *Proteins* 2000, **39**:178-194.

12. 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.

13. Fernández-Recio J, Totrov M, Abagyan R: **Soft protein-protein docking in internal coordinates**. *Protein Sci* 2002, **11**:280-291.

14. Fernandez-Recio J, Totrov M, Abagyan R: **ICM-DISCO docking by global energy optimization with fully flexible side-chains**. *Proteins* 2003, **52**.

15. Gray JJ, Moughon S, Wang C, Schueler-Furman O, Kuhlman B, Rohl CA, Baker D: **Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations**. *J Mol Biol* 2003, **331**.

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

17. Moal IH, Bates PA: **SwarmDock and the use of normal modes in protein-protein docking**. *Int J Mol Sci* 2010, **11**:3623-3648.

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

19. 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.

20. Porter KA, Desta I, Kozakov D, Vajda S: **What method to use for**
• **protein–protein docking?** *Curr Opin Struct Biol* 2019, **55**:1-7
This review reflects on the predictive capabilities of the different existing docking approaches in view of their performance on Protein-Protein Docking Benchmark 5.0 and the CASP11-CAPRI, CASP12-CAPRI and 6th CAPRI blind experiments. On the basis of these results, they conclude that template-based methods yield more accurate predictions if good templates can be found, but generally fail without such templates. The argue that template-based docking for targets with good templates and free docking for targets with worse templates is likely to increase the success rates beyond 40%, which can be further improved by additional restraints from experimental information.

21. Negroni J, Mosca R, Aloy P: **Assessing the applicability of template-based protein docking in the twilight zone**. *Structure* 2014, **22**:1356-1362.

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

23. Lensink MF, Brysbaert G, Nadzirin N, Velankar S, Chaleil RAG,
•• Gerguri T, Bates PA, Laine E, Carbone A, Grudinin S et al.: **Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment**. *Proteins* 2019, **87**:1200-1221
This manuscript describes the results for CAPRI Round 46, the third joint CASP-CAPRI protein assembly prediction challenge. The Round comprised a total of 20 targets including 14 homo-oligomers and 6 hetero-complexes. Eight of the homo-oligomer targets and one heterodimer comprised proteins that could be readily modeled using templates from the Protein Data Bank, often available for the full assembly. The remaining 11 targets comprised 5 homodimers, 3 heterodimers, and two higher-order assemblies. These were more difficult to model, as their prediction mainly involved 'ab-initio' docking of subunit models derived from distantly related templates. A total of ~30 CAPRI groups, including 9 automatic servers, submitted on average ~2000 models per target. About 17 groups participated in the CAPRI scoring rounds, offered for most targets, submitting ~170 models per target. The prediction performance, measured by the fraction of models of acceptable quality or higher submitted across all predictors groups, was very good to excellent for the nine easy targets. Poorer performance was achieved by predictors for

the 11 difficult targets, with medium and high quality models submitted for only 3 of these targets. This experiment highlights yet again the unmet challenge of modeling the conformational changes of the protein components that occur upon binding or that must be accounted for in template-based modeling.

24. Lensink MF, Nadzirin N, Velankar S, Wodak SJ: **Modeling protein-**
• **protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition**. *Proteins* 2019. (in press)
This manuscript describes a summary of the seventh Critical Assessment of Predicted Interactions (CAPRI) community-wide initiative. Performance was evaluated on the basis of 36 114 models of protein complexes submitted by 57 groups-including 13 automatic servers-in prediction rounds held during the years 2016–2019 for eight protein-protein, three protein-peptide, and five protein-oligosaccharide targets with different length ligands. Models of acceptable quality, or better, were obtained for a total of six protein-protein complexes, which included four of the challenging hetero-complexes and a homo-decamer. High accuracy models were obtained for two of the three protein–peptide targets, and for one of the protein–oligosaccharide targets. The remaining protein–sugar targets were predicted with medium accuracy. This analysis indicates that progress in predicting increasingly challenging and diverse types of targets is due to closer integration of template-based modeling techniques with docking, scoring, and model refinement procedures, and to significant incremental improvements in the underlying methodologies.

25. Vreven T, Moal IH, Vangone A, Pierce BG, Kastritis PL, Torchala M, Chaleil R, Jiménez-García B, Bates PA, Fernandez-Recio J et al.: **Updates to the integrated protein–protein interaction benchmarks: docking benchmark version 5 and affinity benchmark version 2**. *J Mol Biol* 2015, **427**:3031-3041.

26. Moal IH, Barradas-Bautista D, Jiménez-García B, Torchala M, van der Velde A, Vreven T, Weng Z, Bates PA, Fernández-Recio J: **IRaPPA: information retrieval based integration of biophysical models for protein assembly selection**. *Bioinformatics* 2017, **33**:1806-1813.

27. Quignot C, Rey J, Yu J, Tufféry P, Guerois R, Andreani J:
•• **InterEvDock2: an expanded server for protein docking using evolutionary and biological information from homology models and multimeric inputs**. *Nucleic Acids Res* 2018, **46**: W408-W416
This paper describes InterEvDock2 server, a major evolution of the previous InterEvDock server, which performs *ab initio* protein docking based on rigid-body sampling followed by consensus scoring using physics-based and statistical potentials, including the InterEvScore function specifically developed to incorporate co-evolutionary information in docking. InterEvDock2 includes automatic template search and comparative modeling of the input proteins. This new server has been benchmarked on 812 complexes in which structural models for the interacting partners can be built by homology and there is available co-evolutionary information in the PPI4DDOCK database. The server identifies a correct model among the top 10 consensus in 29% of the benchmark cases, a significant improvement with respect to the individual scoring functions.

28. Olechnovič K, Venclovas Č: **VoroMQA: assessment of protein structure quality using interatomic contact areas**. *Proteins* 2017, **85**:1131-1145.

29. Oliva R, Vangone A, Cavallo L: **Ranking multiple docking solutions based on the conservation of inter-residue contacts**. *Proteins* 2013, **81**:1571-1584.

30. Chermak E, Petta A, Serra L, Vangone A, Scarano V, Cavallo L, Oliva R: **CONSRANK: a server for the analysis, comparison and ranking of docking models based on inter-residue contacts**. *Bioinformatics* 2015, **31**:1481-1483.

31. Chermak E, De Donato R, Lensink MF, Petta A, Serra L, Scarano V, Cavallo L, Oliva R: **Introducing a clustering step in a consensus approach for the scoring of protein-protein docking models**. *PLoS One* 2016, **11**:e0166460.

32. Yan Y, Zhang D, Zhou P, Li B, Huang S-Y: **HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy**. *Nucleic Acids Res* 2017, **45**:W365-W373.

33. Yan Y, Huang S-Y: **Pushing the accuracy limit of shape**
• **complementarity for protein-protein docking**. *BMC Bioinf* 2018, **20**:696
This manuscript describes a new pairwise shape-based scoring function (LSC) for protein-protein docking, with an exponential form to take into account long-range interactions. The function is incorporated in their FFT-

based docking program. They successfully compare the predictive performance with other FFT-based docking approaches. This scoring function is implemented in HDOCK, the best-performing server in CASP13-CAPRI, and one of the best ones in CAPRI 7th.

34. Marze NA, Roy Burman SS, Sheffler W, Gray JJ: **Efficient flexible**
 • **backbone flexible protein–protein docking for challenging targets**.
 Bioinformatics 2018, **34**:3461-3469
This paper describes the new RosettaDock 4.0, based on a new backbone sampling algorithm called Adaptive Conformer Selection (ACS) that efficiently uses conformational ensembles of proteins, and a new low resolution scoring with Motif Dock Score (MDS). The reported performance is significantly improved over previous version and other docking methods, especially for flexible cases. For highly flexible proteins, the docking procedure is successful when a suitable conformer generation method exists.

35. Jiménez-García B, Roel-Touris J, Romero-Durana M, Vidal M, Jiménez-González D, Fernández-Recio J: **LightDock: a new multi-scale approach to protein–protein docking**. *Bioinformatics* 2018, **34**:49-55.

36. Nadalin F, Carbone A: **Protein-protein interaction specificity is**
 • **captured by contact preferences and interface composition**. *Bioinformatics* 2018, **34**:459-468
This manuscript describes a new contact propensity matrix for scoring of protein–protein docking decoys called CIPS (Combined Interface Propensity for decoy Scoring), which combines interface composition with residue-residue contact preferences. They successfully compare it with other residue statistical potentials. They propose CIPS as a fast, accurate and robust method for selecting millions of docking decoys, and discuss the possibility of using it in docking-drive search, as well as to add constraints to the degrees of freedom of the quaternary structure.

37. Fernández-Recio J, Totrov M, Abagyan R: **Identification of protein–protein interaction sites from docking energy landscapes**. *J Mol Biol* 2004, **335**:843-865.

38. Dequeker C, Laine E, Carbone A: **Decrypting protein surfaces by combining evolution, geometry, and molecular docking**. *Proteins* 2019, **87**:952-965.

39. Chelliah V, Blundell TL, Fernández-Recio J: **Efficient restraints for protein–protein docking by comparison of observed amino acid substitution patterns with those predicted from local environment**. *J Mol Biol* 2006, **357**:1669-1682.

40. Mintseris J, Pierce B, Wiehe K, Anderson R, Chen R, Weng Z: **Integrating statistical pair potentials into protein complex prediction**. *Proteins* 2007, **69**:511-520.

41. Roel-Touris J, Bonvin AMJJ, Jiménez-García B: **LightDock goes information-driven**. *Bioinformatics* 2020, **36**:950-952.

42. Andreani J, Faure G, Guerois R: **InterEvScore: a novel coarse-grained interface scoring function using a multi-body statistical potential coupled to evolution**. *Bioinformatics* 2013, **29**:1742-1749.

43. Song Y, DiMaio F, Wang Ray Y-R, Kim D, Miles C, Brunette TJ, Thompson J, Baker D: **High-resolution comparative modeling with RosettaCM**. *Structure* 2013, **21**:1735-1742.

44. Nadaradjane AA, Quignot C, Traoré S, Andreani J, Guerois R:
•• **Docking proteins and peptides under evolutionary constraints in Critical Assessment of PRediction of Interactions rounds 38 to 45**. *Proteins* 2019. (in press)
This manuscript describes the participation of the Andreani/Guerois group in CAPRI 7th, where they had the best performance among all participants, using additional strategies to include evolutionary information in docking predictions beyond their standard InterEvDock pipeline. These strategies include template-based modeling with local adjustments for sequence identity templates above 30% and larger perturbations otherwise; covariation-based structure prediction for individual protein partners; and the identification of evolutionarily conserved and structurally recurrent anchoring interface motifs.

45. Koukos PI, Bonvin AMJJ: **Integrative modelling of biomolecular complexes**. *J Mol Biol* 2020, **432**:2861-2881.

46. Pons C, D'Abramo M, Svergun DI, Orozco M, Bernadó P, Fernández-Recio J: **Structural characterization of protein–protein complexes by integrating computational docking with small-angle scattering data**. *J Mol Biol* 2010, **403**:217-230.

47. Jimenez-Garcia B, Bernado P, Fernandez-Recio J: **Stuctural characterization of protein-protein interactions with pyDock SAXS**. *Methods Mol Biol* 2020, **2112**:131-144.

48. Jimenez-Garcia B, Pons C, Svergun DI, Bernado P, Fernandez-Recio J: **pyDockSAXS: protein-protein complex structure by SAXS and computational docking**. *Nucleic Acids Res* 2015, **43**:W356-361.

49. Karaca E, Bonvin AMJJ: **On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys**. *Acta Crystal D* 2013, **69**:683-694.

50. Schneidman-Duhovny D, Hammel M: **Modeling structure and dynamics of protein complexes with SAXS profiles**. *Methods Mol Biol* 2018, **1764**:449-473.

51. Schneidman-Duhovny D, Hammel M, Sali A: **Macromolecular docking restrained by a small angle X-ray scattering profile**. *J Struct Biol* 2011, **173**:461-471.

52. Schindler Christina EM, de Vries Sjoerd J, Sasse A, Zacharias M: **SAXS data alone can generate high-quality models of protein-protein complexes**. *Structure* 2016, **24**:1387-1397.

53. Ignatov M, Kazennov A, Kozakov D: **ClusPro FMFT-SAXS: ultra-fast filtering using small-angle X-ray scattering data in protein docking**. *J Mol Biol* 2018, **430**:2249-2255.

54. Vreven T, Schweppe DK, Chavez JD, Weisbrod CR, Shibata S, Zheng C, Bruce JE, Weng Z: **Integrating cross-linking experiments with *ab initio* protein-protein docking**. *J Mol Biol* 2018, **430**:1814-1828.

55. Rosell M, Rodríguez-Lumbreras LA, Romero-Durana M, Jiménez-García B, Díaz L, Fernández-Recio J: **Integrative modeling of protein-protein interactions with pyDock for the new docking challenges**. *Proteins* 2019. (in press).