Modeling protein interactions and complexes in CAPRI

7th CAPRI evaluation meeting April 3-5 EMBL-EBI, Hinxton UK

Shoshana J. Wodak¹, Sameer Velankar², Michael J. E. Sternberg³

¹VIB-VUB Centre for Structural Biology, Pleinlaan 2, 1050, Brussels Belgium

²European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI),

Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, United Kingdom

³Department of Life Sciences, Imperial College London,

London, SW7 2AZ, UK

The crucial role in all cellular processes played by protein complexes, and more generally by protein interactions with other macromolecules underlie fundamental biological functions, and their deregulation or disruption causes disease. Characterizing protein interactions and understanding their underlying principles, is therefore a major goal of present-day molecular biology and medicine.

A substantial part of our knowledge about protein interactions and protein complexes has been derived from data on the three-dimensional structures of these complexes determined by experimental techniques such as X-ray crystallography. So far however, detailed structural information has been available for only a small fraction of the protein assemblies in the cell that can be detected by proteomics and other methods. On the other hand, structural biology has been very effective in mapping out the structural repertoire of individual proteins, many of which represent building blocks of larger complexes. Computational methods for predicting the structure of protein complexes, which exploit the rapidly increasing body of data on sequences and three-dimensional structures of individual proteins and their complexes stored in public

databases, therefore play an increasingly important role in efforts to populate the still uncharted landscape of protein assemblies.

CAPRI (The Critical Assessment of Predicted Interactions) (<http://www.ebi.ac.uk/msd-srv/capri/;capri-docking.org>), has been a major factor fueling progress in this endeavor. CAPRI is a community-wide initiative established in 2001, which offers computational biologists the possibility to test their modeling algorithms in blind prediction Rounds on experimentally-determined structures of protein complexes, the 'targets', provided to CAPRI prior to publication. These predictions Rounds are initiated on a rolling basis, as soon as one or more targets become available, and completed 3-6 weeks later. Registered participants are invited to predict the 3D structure of the target protein assembly using as input sequence information alone, or structures of the unbound components when those are available. CAPRI Rounds also include a scoring challenge in which the correct assembly mode(s) must be identified out of an ensemble of decoys (incorrect models). Groups can participate in either or both challenges (as predictors and scorers). Detailed information on how CAPRI Rounds are conducted, how predicted complexes are evaluated against the target structure, and how the performance of participating groups is ranked can be found at www.capri-docking.org and information on all past CAPRI prediction Rounds can be found at <https://www.ebi.ac.uk/msd-srv/capri/>.

To date, 47 CAPRI prediction Rounds were completed with a total of 160 targets. Results for Rounds 1-36 were presented at 6 Assessment Meetings, held between 2002 and 2016 in Europe and Israel. Initially, taking on only targets representing homo- and hetero- protein-protein complexes, CAPRI has significantly broadened its horizon to include complexes of proteins with other large molecules, such peptides, nucleic acids and sugars. Several CAPRI prediction Rounds also involved the prediction of protein binding affinities and modeling the positions of water molecules in protein-protein interfaces. To foster closer integration between different modeling techniques and between the distinct scientific communities that develop them, CAPRI also conducted joint assembly prediction experiments with CASP (Critical Assessment of Structure Prediction), whose traditional focus has been on the prediction of the 3D structure of individual proteins (<http://predictioncenter.org/>). These experiments were run in the summers of 2014, 2016 and 2018, with results published in three Special Issues of Proteins dedicated to CASP.

This CAPRI Special Issue describes the prediction results for the 16 targets of CAPRI Rounds 38-45 that took place during the years 2016-2019. These results were presented at the 7th CAPRI Evaluation meeting held in April 2018 at the European Bioinformatics Institute in Hinxton, UK (<https://www.ebi.ac.uk/pdbe/about/events/7th-capri-evaluation-meeting>; <http://www.capri-docking.org/events/>) organized by Dr. Michael Sternberg (Imperial College London), Dr. Sameer Velankar (EMBL-EBI), and Dr. Paul Bates (The Francis Crick Institute, London, UK).

As highlighted in the overview of Lensink *et al.* and in the 20 reports of top performing individual groups of this Issue, the 16 targets predictors were challenged with, were particularly diverse and demanding. They included protein-protein hetero-complexes (nearly half of the targets), as well as protein-peptide, and protein-oligosaccharide complexes. Some of the protein hetero-complexes were particularly difficult to model due to the inter domain flexibility

of the component proteins, to the presence of flexible loops contributing to the binding interface, or absence of adequate templates for a protein component. The main difficulty in modeling the protein-peptide complexes, an important category of targets with which members of CAPRI community have achieved significant expertise, resided in the challenge of accurately modeling a relatively small interaction interface, while accounting for the conformational flexibility of the interacting moieties. Handling the high conformational flexibility of the sugar moiety when modeling the protein-oligosaccharide targets was even harder, given the still limited experience of CAPRI groups with this type of targets.

As readers of this issue will learn, the recipe for successful predictions across targets of all categories was the use of increasingly sophisticated ‘integrative’ modeling techniques that deftly leverage information not only on templates for the individual subunits of a complex, but also on templates for the complex as a whole, and on the interfaces of these templates. Information from different templates and interfaces was often also combined, by prioritizing alternative models using guided docking calculations augmented by biochemical information, or computing various model quality scores including approximate binding affinities to evaluate and rank the corresponding solutions.

This integrative strategy is greatly facilitated by the availability of efficient and easy to use tools for various steps of the modeling process. These include a widening palette of docking and scoring procedures, and tools for identifying templates, developed by the CAPRI community, which are made available as software modules or via an increasing number of Web servers whose prediction performance often rivals with that of human predictors (see www.capri-docking.org). As a result, groups can focus on optimizing specific aspects of the modeling procedure and readily evaluate the impact on the prediction exercise as a whole. In parallel, ‘classical’ docking algorithms are shown to gain significantly in speed and sampling capacity, thereby improving the ability to incorporate conformation flexibility and more accurately score and prioritize docking solutions, allowing their authors to rank prominently among the top performing groups.

A lingering bottleneck is the limited ability to model the more consequential conformational changes that proteins often undergo upon association, or those that need to be modeled when only distant templates are available. Overcoming these and other bottlenecks of the modeling process will continue to depend on the availability of challenging test systems. With the spectacular advances in high-resolution single molecule cryo-EM techniques, such systems will most likely include large and dynamic multi-protein assemblies, enough to keep the modeling community busy for years to come. One would also hope that the critical dependence of protein assembly prediction on the availability of templates be gradually eased as *ab-initio* prediction of protein 3D structures becomes accurate enough, thanks notably to Deep Learning techniques.

We wish to dedicate this CAPRI Special Issue of Proteins to Dave Ritchie, who tragically passed away on Sep 15 2019, after a battle with Cancer. This is a huge loss for the CAPRI community and for science in general. The methods and software developed by Dave have been singularly powerful and continue to occupy a unique and important place in the field and beyond. Dave will also be remembered as a most pleasant, modest and helpful colleague.

Acknowledgements

CAPRI is nourished by the enthusiasm of its vibrant community, with the continuous support and guidance of the CAPRI Management Committee (www.capri-Docking.org). INSTRUCT-ERIC is gratefully acknowledged for support to the 2019 CAPRI evaluation meeting.

The entire CAPRI community expresses deep thanks to all the structural biologists who provided the targets of CAPRI Round 38-45 discussed in this issues: Colin Kleanthous (Oxford University, UK), Savvas Savvides (Ghent University, Belgium), Alain Roussel (Université Aix Marseille, France), Ondřej Vaněk (Charles University, Prague, Czech Republic), Han Remaut (Free University of Brussels, and the Flemish Institute for Biotechnology, Brussels Belgium), Sarel Fleishman and Ravit Netzer (Weizmann Institute, Israel), Irina Gutsche (Institut de Biologie Structurale, Grenoble, France), Petri Kursula (University of Bergen Norway), Shifra Lansky and Gil Shoham (Hebrew University of Jerusalem, Israel).