Received: 27 May 2019 | Revised: 26 September 2019 | Accepted: 27 September 2019

DOI: 10.1002/prot.25838

PROTEINS WILEY

RESEARCH ARTICLE

Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment

Marc F. Lensink[1] | Guillaume Brysbaert[1] | Nurul Nadzirin[2] | Sameer Velankar[2] | Raphaël A. G. Chaleil[3] | Tereza Gerguri[3] | Paul A. Bates[3] | Elodie Laine[4] | Alessandra Carbone[4,5] | Sergei Grudinin[6] | Ren Kong[7] | Ran-Ran Liu[7] | Xi-Ming Xu[7] | Hang Shi[7] | Shan Chang[7] | Miriam Eisenstein[8] | Agnieszka Karczynska[9] | Cezary Czaplewski[9] | Emilia Lubecka[10] | Agnieszka Lipska[9] | Paweł Krupa[11] | Magdalena Mozolewska[12] | Łukasz Golon[9] | Sergey Samsonov[9] | Adam Liwo[9,13] | Silvia Crivelli[14] | Guillaume Pagès[6] | Mikhail Karasikov[15] | Maria Kadukova[6,16] | Yumeng Yan[17] | Sheng-You Huang[17] | Mireia Rosell[18,19] | Luis A. Rodríguez-Lumbreras[18,19] | Miguel Romero-Durana[18] | Lucía Díaz-Bueno[18] | Juan Fernandez-Recio[18,19,20] | Charles Christoffer[21] | Genki Terashi[22] | Woong-Hee Shin[22] | Tunde Aderinwale[21] | Sai Raghavendra Maddhuri Venkata Subraman[21] | Daisuke Kihara[21] | Dima Kozakov[23] | Sandor Vajda[24,25] | Kathryn Porter[24] | Dzmitry Padhorny[23] | Israel Desta[24] | Dmitri Beglov[24] | Mikhail Ignatov[23] | Sergey Kotelnikov[23,16] | Iain H. Moal[2] | David W. Ritchie[26] | Isaure Chauvot de Beauchêne[26] | Bernard Maigret[26] | Marie-Dominique Devignes[26] | Maria E. Ruiz Echartea[26] | Didier Barradas-Bautista[27] | Zhen Cao[27] | Luigi Cavallo[27] | Romina Oliva[28] | Yue Cao[29] | Yang Shen[29] | Minkyung Baek[30] | Taeyong Park[30] | Hyeonuk Woo[30] | Chaok Seok[30] | Merav Braitbard[31] | Lirane Bitton[32] | Dina Scheidman-Duhovny[31,32] | Justas Dapkūnas[33] | Kliment Olechnovič[33] | Česlovas Venclovas[33] | Petras J. Kundrotas[34] | Saveliy Belkin[34] | Devlina Chakravarty[34] | Varsha D. Badal[34] | Ilya A. Vakser[34] | Thom Vreven[35] | Sweta Vangaveti[35] | Tyler Borrman[35] | Zhiping Weng[35] | Johnathan D. Guest[36,37] | Ragul Gowthaman[36,37] | Brian G. Pierce[36,37] | Xianjin Xu[38] | Rui Duan[38] | Liming Qiu[38] | Jie Hou[39] | Benjamin Ryan Merideth[38,40] | Zhiwei Ma[38,41] | Jianlin Cheng[39,40] | Xiaoqin Zou[38,40,41,42] | Panagiotis I. Koukos[43] | Jorge Roel-Touris[43] | Francesco Ambrosetti[43] | Cunliang Geng[43] |

Correction added on 11 Nov 2019, after first online publication: ORCID ID added for Mireia Rosell. Author name "Panos I. Koukos" has been updated to "Panagiotis I. Koukos."

Jörg Schaarschmidt[43] | Mikael E. Trellet[43] | Adrien S. J. Melquiond[43] | Li Xue[43] | Brian Jiménez-García[43] | Charlotte W. van Noort[43] | Rodrigo V. Honorato[43] | Alexandre M. J. J. Bonvin[43] | Shoshana J. Wodak[44]

[1]University of Lille, CNRS UMR8576 UGSF, Unité de Glycobiologie Structurale et Fonctionnelle, Lille, France

[2]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK

[3]Biomolecular Modelling Laboratory, The Francis Crick Institute, London, UK

[4]CNRS, IBPS, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), Sorbonne Université, Paris, France

[5]Institut Universitaire de France (IUF), Paris, France

[6]Université Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, Grenoble, France

[7]Institute of Bioinformatics and Medical Engineering, School of Electrical and Information Engineering, Jiangsu University of Technology, Changzhou, China

[8]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

[9]Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland

[10]Institute of Informatics, Faculty of Mathematics, Physics, and Informatics, University of Gdańsk, Gdańsk, Poland

[11]Polish Academy of Sciences, Institute of Physics, Warsaw, Poland

[12]Polish Academy of Sciences, Institute of Computer Science, Warsaw, Poland

[13]School of Computational Sciences, Korea Institute for Advanced Study, Seoul, South Korea

[14]Department of Computer Science, UC Davis, Davis, California

[15]Department of Computer Science, ETH, Zurich, Switzerland

[16]Moscow Institute of Physics and Technology, Dolgoprudniy, Russia

[17]School of Physics, Huazhong University of Science and Technology, Wuhan, Hubei, China

[18]Barcelona Supercomputing Center (BSC), Barcelona, Spain

[19]Instituto de Ciencias de la Vid y del Vino (ICVV-CSIC), Logroño, Spain

[20]Instituto de Biología Molecular de Barcelona (IBMB-CSIC), Barcelona, Spain

[21]Department of Computer Science, Purdue University, West Lafayette, Indiana

[22]Department of Biological Sciences, Purdue University, West Lafayette, Indiana

[23]Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York

[24]Department of Biomedical Engineering, Boston University, Boston, Massachusetts

[25]Department of Chemistry, Boston University, Boston, Massachusetts

[26]University of Lorraine, CNRS, Inria, LORIA, Nancy, France

[27]Physical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

[28]Department of Sciences and Technologies, University of Naples "Parthenope", Napoli, Italy

[29]Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas

[30]Department of Chemistry, Seoul National University, Seoul, Republic of Korea

[31]Department of Biological Chemistry, Institute of Live Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

[32]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel

[33]Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, Lithuania

[34]Computational Biology Program and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas

[35]Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts

[36]University of Maryland Institute for Bioscience and Biotechnology Research, Rockville, Maryland

[37]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, Maryland

[38]Dalton Cardiovascular Research Center, University of Missouri, Columbia, Missouri

[39]Department of Computer Science, University of Missouri, Columbia, Missouri

[40]Informatics Institute, University of Missouri, Columbia, Missouri

[41]Department of Physics and Astronomy, University of Missouri, Columbia, Missouri

[42]Department of Biochemistry, University of Missouri, Columbia, Missouri

[43]Computational Structural Biology Group, Department of Chemistry, Faculty of Science, Utrecht University, Utrecht, The Netherlands

[44]VIB Structural Biology Research Center, VUB, Brussels, Belgium

Correspondence
Marc F. Lensink, University of Lille, CNRS UMR8576 UGSF, Unité de Glycobiologie Structurale et Fonctionnelle, F-59000 Lille, France.
Email: marc.lensink@univ-lille.fr

Shoshana J. Wodak, VIB Structural Biology Research Center, VUB, Pleinlaan 2, B-1050 Brussels, Belgium.
Email: shoshana.wodak@vib-vub.be

Funding information
Agence Nationale de la Recherche, Grant/Award Number: ANR-15-CE11-0029-03; Cancer Research UK, Grant/Award Number: FC001003; H2020 European Institute of Innovation and Technology, Grant/Award Numbers: 675728, 777536, 823830; Lietuvos Mokslo Taryba, Grant/Award Number: S-MIP-17-60; Medical Research Council, Grant/Award Number: FC001003; National Institutes of Health, Grant/Award Numbers: R01GM074255, R01GM123055, R35GM124952; National Natural Science Foundation of China, Grant/Award Number: 31670724; National Research Foundation of Korea, Grant/Award Number: 2016M3C4A7952630; National Science Foundation, Grant/Award Number: DBI1565107; Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: 718.015.001; Spanish "Programma Estatal I+D+i", Grant/Award Number: BIO2016-79930-R; University Parthenope, Grant/Award Number: Finanziamento per il Sostegno alla Ricerca Individ; Wellcome Trust, Grant/Award Number: FC001003

Abstract

We present 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 heterocomplexes. 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. A similar performance "gap" was displayed by scorer groups, highlighting 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. Our analysis also indicates that residues in binding interfaces were less well predicted in this set of targets than in previous Rounds, providing useful insights for directions of future improvements.

KEYWORDS
blind prediction, CAPRI, CASP, docking, oligomeric state, protein assemblies, protein complexes, protein-protein interaction, template-based modeling

1 | INTRODUCTION

Protein-protein interactions and multiprotein complexes, which often include other macromolecular components such as DNA or RNA, play crucial roles in many cellular processes. Their disruption or deregulation often leads to disease.[1,2] Charting these interactions and elucidating the principles that governs them remains an important frontier that molecular biology and medicine strive to conquer.

Data on the three-dimensional structures of protein complexes determined by experimental methods and deposited in the PDB (Protein Data Bank)[3] have taught us much of what we currently know about these complexes.[4-7] So far however, detailed structural information has been available for only a small fraction of protein complexes, and more particularly multiprotein complexes, that are active in the cell and can be detected by modern proteomics and other methods. But this is changing rapidly thanks to recent spectacular advances in single molecule cryo-EM techniques, specifically geared at determining the structure of large macromolecular assemblies at atomic resolution.[8,9]

In the meantime, structural biology is continuing to successfully characterize the structural and folding landscape of individual proteins, many of which form the building blocks of larger assemblies. This increasingly rich structural repertoire in conjunction with the recent explosion of the number of available protein sequences and progress in computational methods are making it possible to model the 3D structure of individual proteins with accrued accuracy from sequence information alone. Most commonly, this is done using as templates the structures of related proteins deposited in the PDB.[10-12]

Owing to the recent explosion of the number of available protein sequences, and progress in computational methods for exploiting them to predict residue-residue contacts,[13-15] the ability to model protein structures in absence of available templates has also made significant strides forward. Furthermore, information on structure and sequence features of proteins (see for examples[16-19]) is being exploited much more efficiently thanks to new developments in Artificial Intelligence Deep Learning techniques,[20,21] enabling the prediction of the 3D structure of proteins from sequence information alone, as most recently demonstrated in the CASP13 ab-initio structure prediction challenge.[22]

Protein structures from this increasingly rich repertoire, determined either experimentally or computationally, may be used as

templates or scaffolds for designing artificial proteins with many useful medical applications.[23-25] Designing large artificial multiprotein assemblies has also been attempted, but remains considerably more challenging.[26] Modeling natural higher order protein assemblies is likewise difficult and involves sophisticated hybrid modeling techniques,[27,28] which integrate sequence and structural information on individual proteins with various other types of data.

Computational approaches play a very important role in the efforts to populate the uncharted landscape of protein assemblies. Of particular relevance here are methods for modeling the 3D structures of protein assemblies starting from the known structures of the individual components, the so-called "docking" algorithms, and the associated energetic criteria for singling out stable binding modes.[29-31] CAPRI (Critical Assessment of PRedicted Interactions) (http://pdbe.org/capri/; http://www.capri-docking.org/) is a community-wide initiative inspired by CASP (Critical Assessment of protein Structure Prediction). CAPRI was established in 2001 with the mission of offering computational biologists the opportunity to test their algorithms in blind predictions of experimentally determined 3D structures of protein complexes, the "targets," provided to CAPRI prior to publication. CASP has been very instrumental in stimulating the field of protein structure prediction. CAPRI has played a similar role in advancing the field of modeling protein assemblies. Initially focusing on testing procedures for predicting protein-protein complexes, CAPRI now also deals with protein-peptide, protein-nucleic acids, and protein-oligosaccharide complexes. In addition, CAPRI has conducted challenges geared at evaluating computational methods for estimating binding affinity of protein-protein complexes[32-34] and predicting the positions of water molecules at the interfaces of protein complexes.[35]

The task of modeling the atomic structure of protein complexes has likewise evolved. It was initially limited to the classical docking procedures. These procedures sample and score putative binding poses of two or more proteins starting from the known unbound structures of the individual components of a complex (see Reference [29]). In recent years however, thanks to the growing ease with which structural templates can be found in the PDB, docking calculations can take as input homology-built models of the components, with an increasing degree of success. It is furthermore not uncommon to find templates for the entire protein assembly. Such cases occur most often for assemblies of identical subunits (homodimers, or higher order homo-oligomers), because closely related proteins tend to adopt the same assembly mode (oligomeric state).[36,37] In such instances, classical docking calculations may no longer be required because the protein assembly can be modeled directly from the template, a task also called "template-based docking."[10,38,39]

In a significant number of cases however, the modeling task remains challenging because the template structure may differ significantly from the structure of the protein to be modeled, or adequate templates cannot be identified. Overcoming these important road blocks calls for a much closer integration of methods for predicting the 3D structure of individual protein subunits and those for modeling protein assemblies, and developing means for improving the accuracy of the resulting structures. An important step in this direction has been to establish closer ties between the CASP and CAPRI communities by running joint CASP-CAPRI assembly prediction experiments. Two such experiments were conducted in the summers of 2014 and 2016, respectively, with results presented at the CASP11 and CASP12 meetings in Cancún, Mexico, and Gaeta, Italy, and published in two special issues of Proteins.[40-42]

Here we present the results of the CASP13-CAPRI challenge, the third joint assembly prediction experiment with CASP, representing Round 46 of CAPRI. This prediction Round was held in the summer of 2018 as part of the CASP13 prediction season. Round 46 also included scoring experiments, where participants are invited to identify the correct models from an ensemble of anonymized predicted complexes generated during the docking experiment.[43,44]

CAPRI Round 46 comprised a total of 20 targets including 14 homo-oligomers, and 6 heterocomplexes, for which predicted models were assessed. These represented about half the total number of targets (42) offered for the assembly prediction challenge to CASP predictors. The targets of Round 46 were selected by the CAPRI management, as representing tractable modeling problems for the CAPRI community. The selection criteria were less strict than in previous joint CASP-CAPRI experiments. A target was considered a tractable modeling problem even when it was a dimer, or higher order assembly, for which only distantly related templates could be identified, for at least a portion of the components of the target complex, using available tools such as HHpred.[45,46] But targets where even such templates could not be identified were considered as particularly difficult ab-initio fold prediction problems, since both the 3D structures of the subunits and their association modes need to be predicted simultaneously. Such problems are very challenging even for CASP groups expert in ab-initio fold prediction, but remain intractable for CAPRI groups where this expertise is mostly lacking. As in previous Rounds, such targets where therefore not offered to CAPRI groups in this Round.

Combining the still distinct methods and expertise of both communities into an integrated modeling approach to the problem of protein assembly prediction has been an important goal of the CASP-CAPRI collaboration. In this third joint prediction Round, important steps in this direction included relaxing the criteria for selecting CAPRI targets, and the fact that subunit models made available by CASP servers were increasingly used as input for docking calculations by CAPRI groups, as will be further discussed.

A summary of the results of this CASP13-CAPRI assembly prediction challenge was presented at the CASP13 meeting held in Cancún, Mexico in December 2018. Here we present the complete results of this challenge, which also include those of the predicted protein-protein interfaces,[41,47] for example, the amino acids residues that are part of the recognition surfaces of the target proteins.

A separate evaluation of the CASP13 assembly prediction performance, reported at the CASP13 meeting and in this Special issue,[22] was carried out by a team of independent assessors in collaboration with the CASP team.

2 | THE TARGETS

The 20 targets of the CASP13-CAPRI assembly prediction experiment, which is henceforth denoted as Round 46, are listed in Table 1. The targets are designated by their CAPRI target ID followed by their corresponding CASP target ID, prefixed by "T" for homo-oligomers, and by "H" for hetero-oligomers.

As in previous CASP-CAPRI challenges the majority of the targets (14) were homo-oligomers. The remaining six targets were hetero-complexes. These targets were proteins from different organisms with the size of individual subunits spanning a very wide range (70-1589 residues). They were offered to this challenge by individual structural biology laboratories. All the targets were high-resolution X-ray structures, with three exceptions: targets **T144/T0984** and **T147/T0995**, and the 18-mer hetero complex (**T159/H1021**), determined by cryo-EM. Most of the targets had annotated biological function and the majority had an author-assigned oligomeric state of the protein, although in a few cases these assignments may have been tentative. In several cases analysis of the target crystal contacts and the prediction results, with further support from computational procedures such as PISA[48] and EPPIC,[49] suggested alternative oligomeric states to those assigned by the authors, as will be described below.

We classified the 20 targets of Round 46 into two categories: easy targets and difficult (to model) targets. The nine easy targets (Table 1) included eight homo oligomers (five homodimers, one homotrimer, one homotetramer, and one homo-octamer), for which good structural templates were available either for the full assembly, or for the main interfaces (of the higher-order homomers).

These homo-oligomers comprised enzymes, transporters and channels from bacteria, bacteria phages, plant, and human. Their sizes ranged from 79 residues for the putative membrane transporter magnetosome dimer from *C. desulfamplus* (**T153/T1006**), to

TABLE 1 CASP13-CAPRI assembly targets, divided into "Easy" and "Difficult" targets, depending on template availability

Easy Target	Targets ID	Stoich.	#Int.	Area (Å^2)	#Res.	PDB	Description
T139	T0961	A4	2	2530 / 670	505	N/A	Acyl-CoA dehydrogenase from *Bdellovibrio bacteriovorus*
T140	T0973	A2	1	3610	146	N/A	Bacteriophage ESE058 coat protein
T143	T0983	A2	1	920	245	N/A	Cals10 protein
T144	T0984	A2	1	4385	752	6NQ1	Two-pore calcium channel protein; EM
T147	T0995	A2/A4/A8	3	1980-520	330	N/A	Cyanide dihydratase (*B. pumilus*); EM
T152	T1003	A2	1	4645	474	6HRH	ALAS2, 5′-Aminolevulinate synthase 2
T153	T1006	A2	1	590	79	6QEK	Putative membrane transporter (*C. desulfamplus*)
T158	T1020	A3	1	1130	577	N/A	SLAC1 protein
T142	H0974	A1B1	1	670	70/80	N/A	Repressor-antirepressor complex (lysogeny switch)
Difficult Target	**Targets ID**	**Stoich.**	**#Int.**	**Area (Å^2)**	**#Res.**	**PDB**	**Description**
T137	T0965	A2	2	1270-1050	326	6D2V	NADP-dependent reductase
T138	T0966	A2	2	1730-900	494	5W6L	RasRap1 site-specific endopeptidase
T141	T0976	A2	1	2700	252	2MXV	Rhodanese-like family protein, bacteria
T148	T0997	A2	1	1060	228	N/A	LD-transpeptidase
T149	T0999	A2	5	1710-400	1589	N/A	Pentafunctional AROM polypeptide: five main enzymes of the shikimate pathway
T150	T0999						Idem; with SAXS data
T151	T0999						Idem; with crosslinking data
T154	T1009	A2	1	2370	718	6DRU	Alpha-xylosidase
T146	H0993	A2B2	3	1910-630	275/112	N/A	Lipid-transport, bacterial outer membrane
T155	H1015	A1B1	1	1220	89/129	N/A	CDI_213 protein, bacteria
T156	H1017	A1B1	1	1025	111/129	N/A	201_INDD4 protein, *E. coli*
T157	H1019	A1B1	1	820	58/88	N/A	CDI207t protein, *E. coli*
T159	H1021	A6B6C6	7	1615-560	148/351/295	N/A	18-mer heterocomplex; EM

Note: The columns present respectively the CAPRI and CASP target ID, stoichiometry of the assembly, the number of interfaces, the surface area range (largest to smallest) of the interfaces, the number of residues per monomer, the PDB-RCSB code (if available), and a textual description of the target. For target structures not yet deposited in the PDB (N/A in column 7) structural details could not be revealed here.

752 residues for the TTPC2 calcium channel dimer from human (**T144/T0984**). The *B. pumilus* cyanide dehydratase (**T147/T0995**) was potentially a higher order assembly, adopting a helical assembly of dimeric repeats, featuring up to three interfaces. Also classified as an easy target (although more challenging than the remaining eight targets in this category), was the putative *Lactococcus phage* repressor-antirepressor hetero complex (**T142/H0974**); templates for this target were not available for the complex as a whole.

The more difficult targets, 11 in all (Table 1), included 6 homodimers and 5 heterocomplexes, all were from bacteria or fungi. They were classified as "medium difficulty" targets in CASP, where the main focus is prediction of the 3D structure of the protein, but were deemed difficult to model in CAPRI, where the focus is to correctly model the binding interfaces between proteins. The difficult homodimers were rather large proteins, with sizes ranging from 326 residues for the NADP Dependent Oxidoreductase TerB (**T137/T0965**) to 1589 residues for the pentafunctional AROM polypeptide (**T149/T0999**), a protein comprising four structural domains. In addition, mostly distantly related templates were available for the individual subunits. To facilitate the task the pentafunctional AROM polypeptide was also offered as a data driven modeling challenge, guided by SAXS (small angle X-ray scattering) data (**T150/S0999**) and XLMS (cross linking mass spectrometry) data (**T151/X0999**). The difficult heterocomplexes comprised three heterodimers, one heterotetramer and one 18-mer hetero complex, a cryo-EM structure, comprising three different subunits. These hetero complexes were composed of smaller subunits (88-295 residues) than their difficult dimer counterparts.

3 | OVERVIEW OF THE PREDICTION EXPERIMENT

As in the previous CASP-CAPRI challenges and in standard CAPRI Rounds, predictor groups were provided with the amino-acid sequence or sequences of the target proteins. Predictors were also provided with information on the biologically relevant oligomeric states of the proteins, provided by the authors for most targets, and occasionally some additional relevant details about the protein.

Predictors generally start by querying public resources[46,50,51] or their own, for structures of protein homologs that can be used as templates for modeling the structure of the target protein. Modeling is greatly facilitated when templates for the full assembly can be identified (which is more commonly the case for homodimers or homo-oligomers). For such targets (which are most often homodimers), the modeling problem does not involve docking calculations to sample different association modes between the subunits. Instead, it reduces to the simpler homology-based modeling problem whereby the target complex as a whole is modeled on the basis of the known complex in the templates. But the difficulty increases significantly when templates can be found only for individual subunits of the complex and even more so when such templates correspond to proteins distantly related to those of the target. Prediction of targets in this category first requires building models of the individual subunits based on the available templates. These models are then used as input for docking calculations in order to identify the most likely association mode between the subunits. Previous CAPRI evaluations clearly showed that the prediction performance for such targets critically depends on the accuracy of the built subunit models and tends to decrease drastically when the available templates are more distantly related to the components of the target complex.[40,41] To help tackle these more difficult cases, 3D models of the target proteins (mainly those of individual subunits) predicted by participating CASP servers were made available to all predictor groups (of both CASP and CAPRI), 1 week into the prediction round, and a good number of CAPRI groups used them (see Supplementary Methods).

Lastly, it is important to note that predicting the structure of higher order assemblies using as input homology modeled structures (or even unbound versions) of the individual subunits is particularly challenging, as was highlighted in previous evaluations.[40,41,52] Many docking algorithm are built to deal with higher order assemblies adopting simple dihedral or cyclic symmetries. Some methods impose the required symmetry constraints from the onset, thereby reducing the rigid-body search space.[53-55] Several docking servers, such as SymmDock,[56] HEX[55] and CLUSPRO[57] offer them as well. When modeling higher order assemblies, a common approach is to proceed in a hierarchical fashion: predicting individual binary associations first, and applying the symmetry constraints to select an optimal combination of interfaces (eg, a pair of interfaces in the case of D2 symmetry) in a defined order.[58] Often however, even small inaccuracies in the predicted binary interfaces tend to propagate, making it difficult to build a correct model for the full assembly.[40,41]

Following a recent practice in CAPRI, Round 46 predictors were invited to submit 100 models for each target, to be used for the scoring challenge (see below). It was stipulated however, that only the five top ranking models should be evaluated, in compliance with CASP regulations. To enable comparisons with the performance in previous CAPRI rounds, prediction results based on the 10 top ranking models, or on the single top ranking models, are also reported.

With the exception of target T137/T096, prediction experiments were followed by the CAPRI scoring experiment. After the predictor submission deadline, all the submitted models (100 per participating group) were shuffled and made available to all the groups participating in the scoring experiment. The "scorer" groups were in turn invited to evaluate the ensemble of uploaded models using the scoring function of their choice, and to submit their own five top-ranking ones. Scorer results based on the top 10, and top 1 ranking models are also reported. For the three target versions of the multidomain AROM polypeptide (**T149**, **T150**, **T151/T0999**), all the models submitted by predictor groups were combined and a single scoring experiment was carried out on the combined set. Typical timelines for the prediction and scoring experiments were 3 weeks and 5 days, respectively.

The number of CAPRI groups submitting predictions and the number of models assessed for each target are listed in the Supplementary Material (Table S1). For Round 46 targets, 27 CAPRI groups submitted on average ~2000 models per target of which 43 075 were assessed

here. On average 17 scorer groups submitted a total of ~170 models per target, of which a total of 3270 models were assessed.

4 | ASSESSMENT CRITERIA AND PROCEDURES

To enable ready comparison with the results obtained in previous CAPRI Rounds, including the two previous CASP-CAPRI experiments,[40,41] models where evaluated using the standard CAPRI assessment protocol. This protocol was complemented with the DockQ score,[59] a continuous quality metric that integrates the main quality measures of the standard CAPRI protocol, as detailed below.

Additionally, we assessed the quality of the predicted protein-protein interfaces in the submitted models, for example, the extent to which residues from each of the contacting subunits that line the binding interface are correctly identified. This is a distinct problem from that of accurately predicting the detailed atomic structure of the binding interface and of the protein complex (or assembly) as a whole. It requires identifying only the residues from each subunit that form the interface[47] and was therefore assessed separately.

4.1 | The CAPRI assessment and ranking protocols

The predicted homo- and hetero-complexes were assessed by the CAPRI assessment team, using the standard CAPRI assessment protocol detailed previously.[43,44] This protocol uses three main parameters, *L_rms*, *i_rms* and *f(nat)*, to measure the quality of a predicted model. *f (nat)* represents the fraction of native contacts in the target that is recalled in the model. Atomic contacts below 3 Å are considered clashes and predictions with too many clashes are disqualified (for the definition of native contacts, and the threshold for clashes see reference 43 and Supplementary Material). *L_rms* represents the backbone rmsd (root means square deviation) over the common set of ligand residues after the receptor proteins have been superimposed, and *i_rms* represents the backbone rmsd calculated over the common set of interface residues after the structural superposition of these residues. An interface residue is defined as such, when any of its atoms (hydrogen atoms excluded) are located within 10 Å of any of the atoms of the binding partner. Based on the values of these three parameters models are ranked into four categories: high quality, medium quality, acceptable quality, and incorrect, as previously described.[40]

For targets representing higher order oligomers that feature more than one distinct interface, as well as for some dimer targets with seemingly ambiguous biological unit assignment, all distinct interfaces formed with neighboring subunits in the crystal were examined. Submitted models were then evaluated by comparing each pair of interacting subunits in the model to each of the relevant pairs of interacting subunits in the target, as described previously.[40] The quality score for the assembly as a whole, or for targets where more than one interfaces was assessed, was taken as the score of the best-predicted individual interface for the assembly. This is a much more lenient criterion than used in previous CASP-CAPRI challenges, where the score for the entire assembly was taken as the score of the worst predicted interface. Schemes of intermediate leniency, representing linear combinations of weighted scores for individual interfaces of the assembly were also examined. But such schemes need to be adapted to adequately balance the scores for low quality predictions for several interfaces vs high quality predictions of, say, only one interface. They must also deal with cases where alternative oligomeric state assignments are considered. More work is therefore needed, and approval by the CAPRI community must be obtained before these schemes can be used to rank the prediction performance.

The quality of the modeled 3D structure of individual subunits was also evaluated by computing the "molecular" root mean square deviation, M-rms, of backbone atoms of the model vs the target. It was used mainly to gauge the influence of the quality of subunit models on the predicted structure of the assembly.

The performance of predictor and scorer groups and servers was ranked based on their best-ranking model in the 5-model submission for each target. The final score assigned to a group or a server was expressed as a weighted sum of the individual target performance, expressed in each of the three categories (acceptable, medium, and high) as achieved by that group or server over all targets:

$$\text{Score}_G = \omega_1 N_{\text{ACC}} + \omega_2 N_{\text{MED}} + \omega_3 N_{\text{HIGH}}$$

where N_{ACC}, N_{MED}, and N_{HIGH} are the number of targets of acceptable-, medium-, and high-quality, respectively, and the values of weights "ω" were taken as $\omega_1 = 1$, $\omega_2 = 2$, and $\omega_3 = 3$.

This ranking method represents a significant difference with previous ranking protocols, where priority was given to the number of targets for which medium or high quality models were submitted, and then to the number of targets with acceptable models. In particular, it takes into account acceptable models in instances where a similar number of medium and/or high quality models are submitted by a given group.

4.2 | Additional assessment measures

To enable a higher-level analysis of the performance across targets, we used a continuous quality metric as formulated by the DockQ score, to evaluate each modeled interface[59]:

$$\text{DockQ} = [f[\text{nat}] + \text{rms}_{\text{scaled}}[L\_rms, d_1] + \text{rms}_{\text{scaled}}[i\_rms, d_2]]/3$$

$$\text{With } \text{rms}_{\text{scaled}} = 1/\left[1 + \left(\frac{\text{rms}}{d_i}\right)^2\right].$$

where *f*(nat), *L_rms*, and *i_rms* are as defined above. The $\text{rms}_{\text{scaled}}$ represents the scaled rms deviations corresponding to either *L_rms* or *i_rms*, *s* and d_i is a scaling factor, d_1 for *L_rms* and d_2 for *i_rms*, which was optimized to fit the CAPRI model quality criteria, yielding $d_1 = 8.5$ Å and $d_2 = 1.5$ Å (see Reference [59])

4.3 | Evaluating the prediction of interface residues

Models submitted by CAPRI predictor scorer and server groups were also evaluated for the correspondence between residues in the predicted interfaces and those observed in the corresponding structures of the 22 targets of Round 46. A total of 38 distinct protein-protein interfaces, sometimes representing more than one interface for each interacting component, were evaluated. The number of interfaces evaluated for individual targets in both categories (easy and difficult) are listed in Table 1. Interface residues of the receptor (R) and ligand (L) components in both the target and predicted models were defined as those whose solvent accessible surface area (ASA) is reduced (by any amount) in the complex relative to that in the individual components.[47] As in the official CAPRI assessment, the surface area change was computed from the structures of the individual components in their bound form.

The agreement between the residues in the predicted vs the observed interfaces was evaluated using the two commonly used measures, *Recall* (sensitivity) and *Precision* (positive predictive value). *Recall* is denoted as *f*(*IR*), the fraction of interface residues in the target complex recalled in the model. *Precision* = 1 − *f*(*OP*), where *f*(*OP*), is the is the fraction of overpredicted residues (false positives) in the predicted interface.

5 | RESULTS AND DISCUSSION

This section is divided into five main parts. The first part presents the results of human predictors, servers, and scorer groups for the 20 individual CAPRI Round 46 targets for which the prediction and scoring experiments were conducted. In the second part, we present the rankings of the same groups established on the basis of their performance across all targets, and discuss insights gained from ranking the performance of these groups for the easy and more difficult targets, respectively. In the third part, we report results of the binding interface predictions obtained by the different categories of participants for all targets. The fourth and final part analyzes methods and factors that may have influenced the prediction performance.

5.1 | Predictor server and scorer results for individual targets

Detailed results obtained by all groups for individual targets analyzed in this study can be found in Tables S2 and S3 of the Supplementary Material. Values of all the CAPRI quality assessment measures for individual models submitted by CAPRI participants for the 20 Round 46 targets can be found on the CAPRI website (URL: http://pdbe.org/capri). Additional information on the performance of individual groups can be found in the Supplementary Material (Individual Group Summaries).

5.1.1 | Predictor and server results

Easy targets: T139, T140, T142, T143, T144, T147, T152, T153, T158

The easy targets comprised eight homomers and one heterocomplex. Since many predictors and groups performed well on these easy targets, we present the highlights of their performance in general terms, without naming the best performing groups, which can be found in Table S2.

For all the eight easy homomer targets, templates for the full assembly were available in the PDB. Examples of available templates used by predictors can be found in the Supplementary Material (Individual Group Summaries). For five of the homomers (**T139/T0961**, **T143/T0983**, **T147/T0995**, **T152/T1003**, **T153/T1006**) the template quality was excellent. These templates featured sequence identity levels of 50% or higher and backbone rmsd values significantly below 2.0 Å. Lower quality templates (25-45% sequence identity; backbone rmsd values ~3.5 Å), were available for the remaining three homomer targets (**T140/T0973**, **T144/T0984**, **T158/T1020**). For the ninth target, the hetero complex (T142/H0974), lower quality templates (29.3% sequence identify, 2.8 Å rmsd) were only available for the individual subunits and not for the assembly as a whole.

It was therefore not surprising that the prediction performance for all the homomer targets was very good to excellent. For the five homomer targets with excellent templates, an unusually large proportion of the models submitted by individual predictor groups were of high quality (see Table S2). For example, for **T152/T1003**, the seven best performing groups each submitted between 3 and 5 high quality models, whereas for **T153/T1006**, the number of groups with a similar performance was 4 but still significant. For **T139/T0961** the homotetramer, 7 out of the 10 best performing predictor groups each submitted at least 4 high quality models for both interfaces among their 5 top ranking submissions. Interestingly, for **T147/T0995**, the higher order helical assembly, high quality models were submitted for the smaller interfaces of this assembly (respectively 680 $Å^2$ and 520 $Å^2$), whereas only medium quality models were submitted for the larger interface (1980 $Å^2$). But the number of groups submitting high quality models was smaller (only one group for interface 2, and four groups for interface 3), whereas 8 out of the top 10 groups submitted as many as 5 medium quality models for interface 1.

For the three targets with lower quality templates (**T140/T0973**, **T144/0984**, **T158/T1020)**, a large proportion of the submitted models were of medium quality. All eight top ranking predictor groups submitted five medium quality models for target T140. Six of the top ranking groups each submitted five and four such models for T144 and T158, respectively (Table S3).

Lastly, for the relatively more challenging heterocomplex (**T142/H0974)**, the performance was significantly lower overall. Whereas all the 10 best performing groups submitted between 1 and 4 medium quality models, nearly all of the remaining ~20 predictor and server groups submitted incorrect models for this target. In comparison, only a small fraction of participating groups (between 2 and 10, out of a

total of about 30) submitted incorrect models for the remaining easy targets.

Figure 1 displays the best model (medium quality) submitted for the heterodimer **T142/H0974**, and the best high quality models submitted for two of the easy homodimer targets, **T140/T0973**, and **T152/T1003**, illustrating the level of accuracy achieved by predictors for this category of targets.

It is noteworthy that automatic servers ranked frequently among the 5 or 10 best performing groups for all the easy targets, with servers such as LZerD, SwarmDock, GalaxyPPDock, Haddock, and HDock achieving high performance more consistently.

Difficult targets: T137, T138, T141, T146, T148, T149 (T150, T151), T154, T155, T156, T157, T159

As already mentioned, these difficult targets comprised six homodimers, and five heterocomplexes, with the latter including one heterotetramer and an 18-mer assembly obtained by cryo-EM (Table 1). For all of these targets, including the homodimers, distantly related templates were in general available only for individual subunits.

Not too surprisingly, the predictions performance for these targets was in general disappointing. For four of the homodimer targets, (**T137/T0965**, **T138/T0966**, **T148/T0997**, **T154/T1009**) predictions failed completely, with only incorrect models submitted by predictor and server groups alike (Table S3). One of these targets, the RasRap1 site-specific endopeptidase (**T138/T0966**), was likely a case of an ambiguous biological unit assignment for the experimental complex. The biological unit assignment made available to the assessors and predictors at the time of the experiment had the membrane localization domain of the protein forming the rather large (1730 Å^2) dimer interface (interface 1 for this target). Neither the PISA software,[48] nor any of the predictor groups recognized this to represent a stable interface, and failed to predict it. This prompted the assessors to look for potential alternative dimer interfaces among the crystal contacts. This yielded a weaker interface (900 Å^2) between the larger cytoplasmic domains of the proteins, which altered the relative orientation of the subunits, and positioned the two membrane localization domains further apart from each other but pointing in the same direction and seemingly well oriented to fit into a planar bilayer (Figure 2A,B). This case turned out to illustrate well the challenge of assigning the biologically relevant oligomeric state of an assembly from the crystal structure. Indeed, PISA predicts neither interfaces of T138 as stable, whereas EPPIC classifies both interfaces as stable. Furthermore, the membrane localization domain seen to interact in the crystal structure of T138, is found in a number of other known structures listed as monomeric in the PDB. The latter observation together with the contradictory conclusions of the computational assignments lends support to the biological unit being defined by the weaker interface in T138. This interface was ultimately assigned as the biological unit in the PDB entry for this complex (5W6L).

Interestingly, among all the participating groups, only the group of Huang submitted a single acceptable model, which was for the weaker interface of T138. This model was ranked 10th in their list of models and was therefore not considered in the final group ranking.

For the remaining two difficult homodimers the best performance was obtained for the primary interface of the multidomain homodimers **T149/T0999**, also offered as data assisted targets **T150/S0999 (SAXS)** and **T151/X0999 (XLMS)**. For this target, five interfaces were evaluated independently, but only the main interface was well predicted, as high-quality templates were available only for this interface.

Lastly, only a few acceptable models were submitted, by both predictors and servers, for the rhodanese-like family homodimer (**T141/T0976**). The difficulty with this target resided in the fact that the protein comprises two structurally similar domains, and forms an intertwined homodimer, where domain-domain contacts between subunits

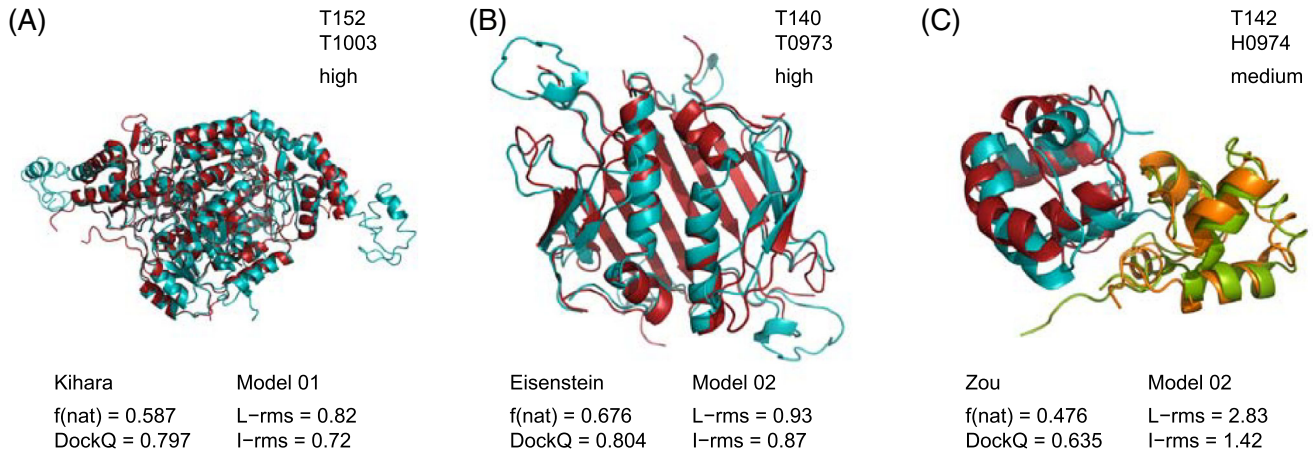

FIGURE 1 Examples of the best quality models obtained for easy targets of Round 46. A, The high quality model by the group of Kihara, obtained for the homodimer **T152/T1003**. B, The high quality model submitted by Eisenstein for the homodimer **T140/T0973**. C, The medium quality model obtained by the group of Zou for the heterodimer **T142/H0974**. The models by Eisenstein and Zou were ranked second among the top five models submitted by these predictors; the Kihara model was their top model. The values of *f(nat)*, *i_rms*, *L_rms*, and the DockQ score for these models are listed

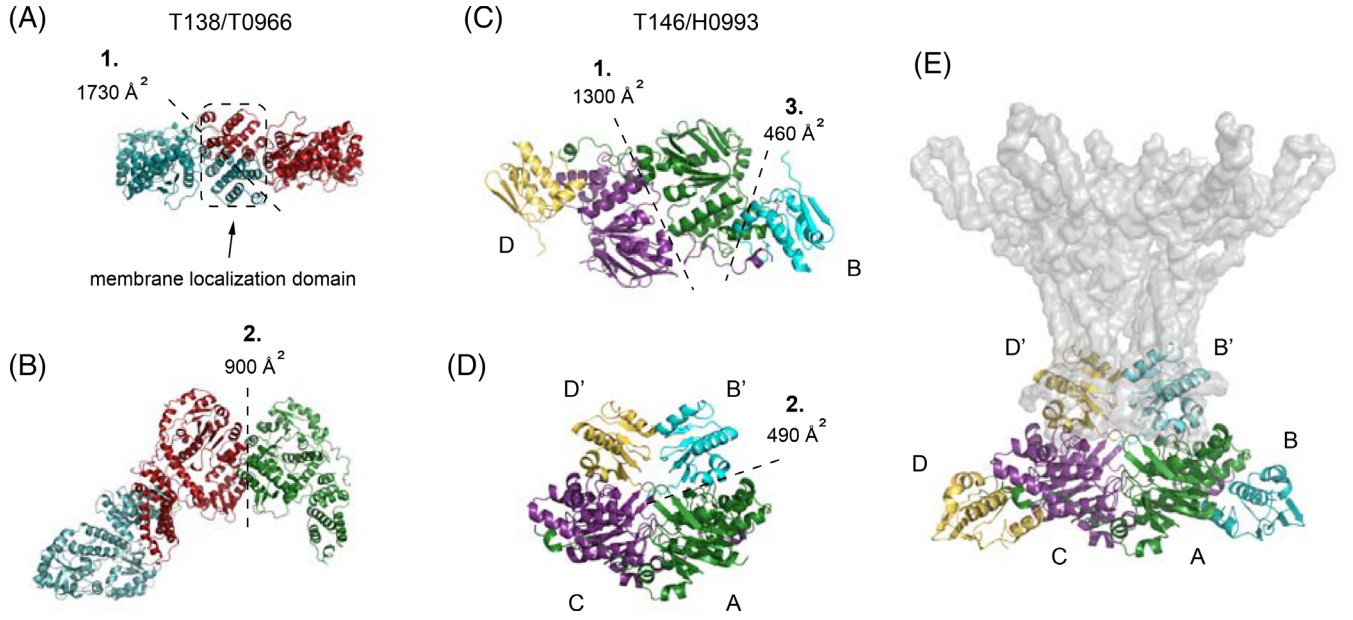

FIGURE 2 Examples of targets with ambiguous biological unit assignments. A, B, Illustrates the case of target **T138/T0966**. A, Displays the dimer association mode communicated by the authors, where the interface (interface 1, 1730 Å^2 buried area) is formed between the two equivalent membrane localization domains. B, Displays the association mode suggested by the assessors after examining crystal contacts, where the dimer interfaces (interface 2, 900 Å^2) is formed between the two cytoplasmic domains. In this new arrangement, the equivalent membrane localization domains are now positioned roughly parallel to one another pointing in the same direction, an arrangement that seems compatible with their insertion into the membrane. C-E, Illustrates the case of the A2/B2 heterotetramer **T146/H0993**. C, Shows the association mode communicated by the authors, where the larger subunits form a homodimer (interface 1, 1300 Å^2) and two copies of the second smaller protein bind at opposite sides of the dimer (interface 3, 460 Å^2), without contacting each other. D, Shows the association mode suggested by the assessors following analysis of the crystal contacts. It forms a more globular complex, featuring the same dimer contact between the large subunits, but involving a different interface between the large and small subunits (interface 2, 490 Å^2) as well as an additional small contact between the two smaller subunits (200 Å^2), thereby reducing the total solvent accessible area upon complex formation. E, Displays both association modes, using the same labels and color-code as in (C,D), with the larger MalF homodimer superimposed onto the cryo-EM structure (PDB 6IC4). The panel illustrates the overlap of the alternative heterointerface of (D) with the interface formed with the other components of the larger complex, lending support to the author assigned assembly of (C). The MalF chains of the EM structure and T146/H0993 align well and for reasons of visibility only those of T146/H0993 are displayed

are more extensive than those within subunits. A hint about how the domains interact in the dimer could be obtained from a number of monomeric templates, featuring related domains that form a roughly similar arrangement to that in the target dimer. Only six groups (four human predictors: Kozakov, Zou, Shen, Eisenstein, and two servers: CLUSPRO, and MDOCKPP) seemed to have successfully exploited this hint and submitted acceptable models among their top five ranking ones (Table S3).

The five heterocomplex targets presented a range of challenges. Availability of poor templates for one or both subunits was a major stumbling block for the prediction of the heterodimer complexes. Only distantly related templates were available for **T155/H1015** (~3 Å rmsd, 34-40% sequence identity), resulting in a single acceptable-quality model among the top 10 submitted by Huang (-Table S3). For **T156/H1017** and **T157/H1019** both complexes of an uncharacterized *E. coli* protein and a partner protein with a putative adhesin/hemagglutinin/hemolysin activity, a relatively good template that revealed some information about the potential interface was available for one of the subunits, but not for the other. Among the top five ranking models, only two acceptable-quality models were submitted for **T156/H1017** (by Venclovas and Zou), whereas four such models were submitted for **T157/H1019** (two by Fernandez-Recio, and one each by the groups of Huang and Chang).

The two higher order heterocomplexes, the heterotetramer **T146/H0993** and the 18-mer complex featuring six copies of three different subunits, **T159/H1021**, posed other major challenges. **T146/H0993**, the complex of the two MlaF proteins involved in lipid transport, was defined as consisting of a homodimer formed by the larger MlaF protein (275 residues), to which two copies of the second smaller protein (112 residues) bind at opposite sides of the dimer without contacting each other. Exploration of the crystal contacts by the CAPRI assessment team suggested an alternative arrangement, which conserved the homodimer, but positioned two different copies of the smaller protein into contact with the dimer, forming a small interface (490 Å^2) with each of the subunits of the dimer, while at the same time contacting each other (200 Å^2 interface), thereby forming a more compact globular complex that buries overall a somewhat larger portion of the solvent accessible surface of the component proteins (Figure 2C, D). The submitted models were therefore assessed against three interfaces, the large homodimer interfaces (~1300 Å^2), and the two

alternative interfaces formed with the smaller protein: the one suggested by the authors (460 Å^2) and the one obtained by the assessors using crystal symmetry operations (490 Å^2).

Not too surprisingly, the larger homodimer interface was predicted with some success, thanks to the availability of homodimer templates. In total 8 medium-quality and 15 acceptable models were submitted by 6 predictor groups and one server (MDockPP), with the groups of Eisenstein and Venclovas submitting 5 and 3 medium-quality models, respectively. Of the two potential interfaces with the smaller protein, the one suggested by the assessors was "detected" only among the top 100 models of one group (that of Chang). On the other hand, two groups, Moal and Kozakov, correctly predicted the interface proposed by the authors among their top five models (Table S3). However, neither of the smaller hetero interfaces were supported by PISA and no templates were available for the A2B2 assembly to help with the assignment. Only after completing the evaluation, was the originally proposed interface for T146 observed in the low-resolution cryo-EM multicomponent *A. baumannii* MLA complex (PDB 6IC4) (deposited December 2018), clearly lending support to the quaternary structure proposed by the authors.

Interestingly however, the alternative interface proposed by the assessors involves the same region of the MalfA dimer that binds the transmembrane component of the larger Cryo-EM complex (Figure 2E). This indicates in turn that the similarly sized alternative interface also conveys biologically relevant information. This case thus illustrates once again the challenge of assigning the biologically relevant association mode from the crystal structure, especially when the latter corresponds to a protein assembly representing a component of a larger complex.

Finally, the most challenging target of Round 46 was indisputably **T159/H1021**, the 18-mer cryo-EM heterocomplex. This complex is composed of three different polypeptides denoted here as A, B, and C comprising 148, 351, and 295 residues, respectively. Each polypeptide is present in six identical copies that form three concentric hexameric rings, stacked on top of each another to form a hat-like structure, with the smallest subunits forming the apical ring (Figure 3). The subunits make extensive contacts within and between rings. These contacts feature an important degree of intertwining, fostered by long extended segments featured primarily by protein C and to a lesser extent by proteins A and B. Whereas good templates were available for protein A, those for proteins B and C were of poorer quality (Figure 3). Protein C, which adopted the least globular fold, had an NMR structure available as template only for its more structured N-terminal domain. The full complex features a total of seven distinct protein-protein interfaces that had to be modeled, of which three were between identical protein subunits. Considering the quality of the templates, the pairwise homo- and hetero-association modes between proteins A and B (interfaces 1, 6, 7 of **T159**) could, in principle, be predicted at some level of accuracy. This was not the case for the remaining four interfaces, involving the least globular protein C.

Prediction results confirmed these expectations (see Table S3). Good prediction performance was obtained for the homomeric interfaces involving two copies of protein A (interface 1 of **T159**), and two copies of protein B (interface 6 of **T159**), respectively. All of the 10 best-performing groups submitted as many as 5 medium-quality models for interface 1, and 5 additional groups each submitted 5 acceptable-quality models for this interface. The 10 best performers counted 8 groups (Weng, Venclovas, Kihara, Shen, Seok, Kozakov,

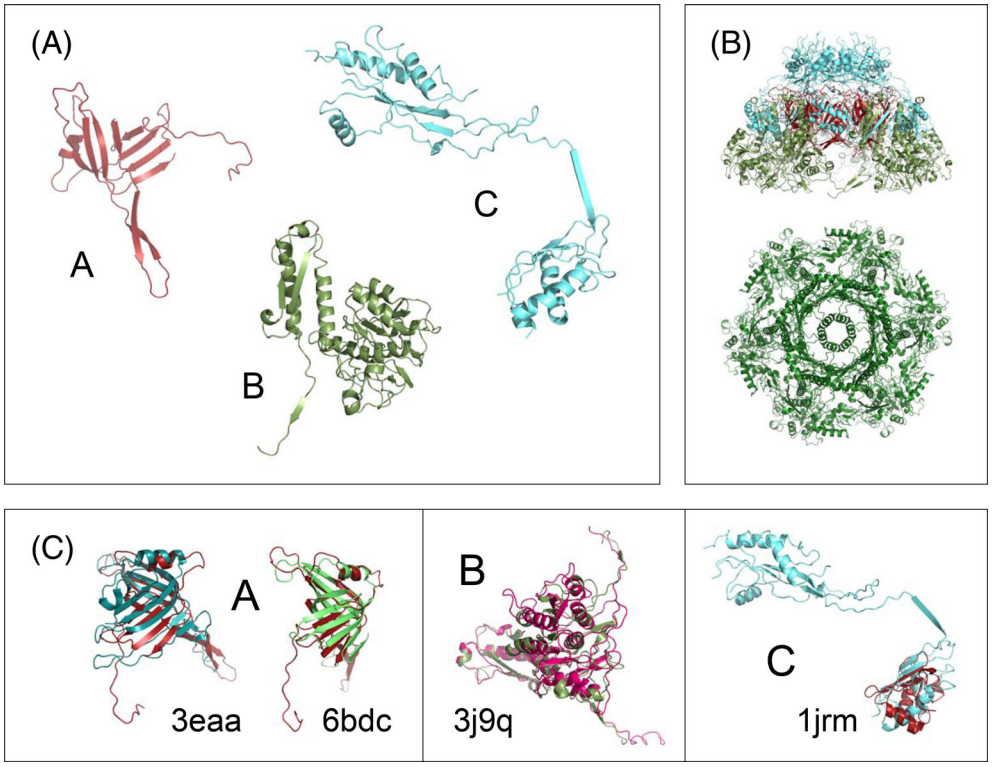

FIGURE 3 Illustration of the modeling challenge for the multiprotein heterocomplex (**T159/H1021**). A, Shows ribbon diagrams of the three different polypeptide chains A, B, and C, forming the complex. Each polypeptide is present in six identical copies that form three concentric hexameric rings, stacked on top of each another to form a hat-like structure, with the smallest subunits forming the apical ring, B. C, Illustrates the quality of the templates available for each of the three subunits (identified by their PDB-RCSB codes), which was particularly poor for subunit C as it only partially covered the structure

Fernandez-Recio and Huang) as well as two servers, LZerD and HDock. The performance for interface 6 was likewise good with 5 out of the 10 top performing groups (including the servers LZerD and HDock), each submitting 5 medium quality models. The performance for the heterodimeric interfaces between protein A and B (interface 7 of **T159**), was significantly lower, with only seven groups (including the CLUSPRO server) submitting at least one acceptable quality model for this interface, and only one group (Seok) submitting a medium quality model.

As expected, only incorrect models were submitted for four association modes involving protein C. Considering the highly intertwined interactions formed by this protein with neighboring subunits in the assembly, these interactions probably form through strong coupling of folding and association. Different modeling techniques, such as those recently developed for modeling interactions with proteins featuring large intrinsically disordered segments,[60] are probably needed to improve the prediction performance of complexes involving nonglobular and more flexible proteins, such as the C subunit of **T159/H1021**.

5.1.2 | Scorer results

Scoring Rounds were held for 19 of the 20 targets or Round 46, with close to 2000 uploaded models offered to scorers per target and the participation of about 17 CAPRI scorer groups (Table S1). Like for the predictor submissions, the performance of scorers was evaluated considering only the top five submitted models. Detailed per target results are provided in Table S2 (easy targets) and Table S3 (difficult targets).

Easy targets: T139, T140, T142, T143, T144, T147, T152, T153, T158

In general, the scorer performance followed the trend of predictor groups for the nine easy targets. But unlike in previous CASP-CAPRI challenges, the performance of scorer groups was more uneven (see Tables S2 and S3 for details). For the five easiest homomer targets, uploaded models contained the largest proportion of high-quality models. But in general scorers identified only a subset of these models. It was also not uncommon to see scorer groups failing to identify some of their own high-quality models submitted as predictors. For example, only three scorer groups produced two or three high-quality models each for target **T152/T1003**, whereas as many as four predictor groups and three docking servers produced between 3 and 5 high-quality models for this target. A similar lower performance was observed for targets **T153/T1006** and **T139/T0961**.

Interestingly, although the scorer performance for **T143/T0983** was overall lower than that of predictors (with only four scorer groups producing high quality models compared with eight predictor groups), the best performing scorer group for this target produced five high-quality models, whereas this number was at most four for the best-performing predictors. For the three somewhat less easy targets, and for the less well-predicted interfaces 1 and 2 of the helical assembly **T147/T0995**, where top-ranking predictor groups produced only medium quality models, the scorer performance was only marginally lower than that of predictor groups.

Several servers were also among the top-performing scorer groups for these easy targets, although not as prominently and consistently as among predictor groups. LZerD, HDock, and MDockPP were the servers that ranked more consistently among the top-performing scorer groups.

Difficult targets: T137, T138, T141, T146, T148, T149 (T150, T151), T154, T155, T156, T157, T159

For the 11 difficult targets, the paucity of models of acceptable quality or higher among the 100 models submitted by predictors was in general the main reason for the inability or difficulty of scorer groups to identify such models in the combined set of uploaded models. Hence, with a few exceptions, scorer results were poorer than those of predictors (Table S3).

For the five difficult homodimer targets for which scoring rounds were organized, scorers submitted only incorrect models for targets **T138/T0966**, **T148/T0997**, and **T154/T1009**, since predictors submitted mostly incorrect models for these targets. For **T141/T0976**, scorer groups performed reasonably well, with seven groups identifying at least one of the acceptable models submitted by only five predictor groups. For **T149/T0999**, and the two data-driven variants of this target (**T150/S0999**, **T151/X0999**), scoring results were combined for all three targets. This increased the size of the total set of models offered to scorers, but likely affected only marginally the fraction of correct models included in the set, as most participants either did not use the SAXS or crosslinking data and submitted very similar models, or participated in at most two of the three targets.

Results for the heteromeric targets depended on the interfaces involved. For **T155/H1015**, scorers did not identify the single acceptable model submitted by predictors. One acceptable model was submitted for **T156/H1017** by the Venclovas team (albeit not among their top five models). Likewise, a single scorer team (Bonvin) identified only one of the few acceptable-quality predictor models submitted for **T157/H1019**. The results were not better for the three interfaces of **T146/H0993**. Scorers failed to identify the 23 medium-quality models submitted by predictors (in fact only two predictor groups: Venclovas and Eisenstein) for the dimer interface of this target (interface 1). Instead, only two servers, HDOCK and MDOCKPP, produced a total of three acceptable quality models in their top-5 submission. But for the smaller heteromeric interface of this target (interface 3), scorers were able to pick out the few acceptable models submitted by predictors.

Finally, the performance of scorers was surprisingly good for the three interfaces of **T159/H1021**, for which predictors submitted models of acceptable quality or better (Table S3). Scorers outperformed predictors groups for interface 1 of T159 (between the two copies of protein A), with all 15 participating groups submitting at least 1 medium quality model, and 10 of these groups each submitting 5 medium quality models. Scorers also significantly outperformed predictors for interface 7 of this target (between proteins A and B). Three groups and the LZerD server each produced five correct models of

TABLE 2 Overall CAPRI performance ranking for top-1, top-5, and top-10 submissions

Rank	Predictor group Human	#[a]	Rank	Top-1	Rank	Top-5	Rank	Top-10
1	Venclovas	20	3	9/4***/4**	1	13/6***/6**	1	13/6***/6**
2	Fernandez-Recio	19	1	11/3***/7**	2	12/5***/6**	2	13/5***/6**
3	Seok	20	4	10/2***/6**	3	11/4***/7**	3	11/4***/7**
4	Kihara	20	2	10/3***/6**	4	11/4***/6**	4	11/4***/6**
5	Weng	20	4	9/4***/3**	5	10/5***/4**	4	11/5***/4**
	Kozakov	19	8	10/9**	5	13/11**	6	13/11**
7	Vakser	20	4	11/1***/7**	7	11/2***/8**	6	11/3***/7**
	Huang	19	8	9/3***/4**	7	11/4***/4**	6	12/4***/4**
	Zou	20	8	11/8**	7	13/2***/6**	6	13/3***/5**
	Bates	18	4	9/3***/5**	7	9/5***/4**	10	9/5***/4**
11	Chang	20	12	10/8**	11	10/3***/6**	11	10/3***/6**
12	Eisenstein	12	8	9/3***/4**	12	9/4***/3**	11	10/4***/4**
13	Pierce	20	14	6/2***/3**	13	7/4***/3**	13	8/4***/3**
	Shen	20	12	11/3***/5**	13	11/3***/5**	14	11/3***/5**
15	Elofsson	20	16	6/3**	15	8/2***/3**	15	8/2***/3**
16	Czaplewski	17	17	5/3**	16	7/1***/4**	16	7/2***/3**
	Grudinin	20	18	3/1***/2**	16	7/1***/4**	16	7/1***/5**
18	Moal	17	15	6/1***/2**	18	6/2***/2**	18	6/2***/3**
19	Carbone	20	20	2/1***/1**	19	5/2***/1**	19	5/2***/2**
20	Schneidman	12	18	3/2***	20	3/2***	20	4/2***/1**
21	Hou	11	21	2**	21	2/1***/1**	21	3/2***
22	Ritchie	4	22	1**	22	1**	22	1**
23	Liwo	11	23	0	23	0	23	1
	Crivelli	12	23	0	23	0	23	1
	EMBO 2017 course	1	23	0	23	0	23	0
	Del Carpio	13	23	0	23	0	23	0
Rank	**Server**	**#[a]**	**Rank**	**Top-1**	**Rank**	**Top-5**	**Rank**	**Top-10**
1	HDOCK	20	3	7/4***/3**	1	10/5***/5**	1	12/5***/6**
2	SWARMDOCK	20	1	9/3***/5**	2	9/5***/4**	2	9/5***/4**
3	CLUSPRO	20	3	10/8**	3	12/10**	3	12/10**
4	LZERD	20	2	8/3***/5**	4	9/3***/6**	4	9/3***/6**
5	MDOCKPP	20	5	10/1***/4**	5	11/1***/5**	5	11/2***/4**
	HADDOCK	19	5	8/3***/2**	5	9/3***/3**	6	9/3***/3**
7	GALAXYPPDOCK	17	7	6/1***/4**	7	7/3***/2**	7	8/3***/2**
8	HAWKDOCK	7	8	1**	8	2/1**	8	2/1***
Rank	**Scorers[b]**	**#[a]**	**Rank**	**Top-1**	**Rank**	**Top-5**	**Rank**	**Top-10**
1	Fernandez-Recio	19	10	7/2***/5**	1	12/5***/6**	1	12/5***/6**
2	Oliva	19	1	11/4***/7**	2	12/4***/7**	1	12/5***/6**
3	Zou	19	4	10/8**	3	12/2***/9**	3	13/3***/8**
	MDOCKPP	19	10	8/1***/6**	3	13/2***/8**	3	13/3***/8**
5	Chang	19	2	10/1***/8**	5	11/2***/9**	7	12/3***/8**
	HDOCK	19	4	9**	5	12/1***/10**	7	12/3***/8**
7	Venclovas	19	2	10/1***/8**	7	11/2***/8**	3	13/2***/10**
	Kihara	19	4	9/1***/7**	7	11/3***/6**	3	12/5***/5**

(Continues)

TABLE 2 (Continued)

Rank	Scorers[b]	#[a]	Rank	Top-1	Rank	Top-5	Rank	Top-10
	Huang	19	8	8/2***/5**	7	11/3***/6**	7	12/4***/6**
10	LZERD	19	4	10/8**	10	10/3***/6**	10	11/5***/4**
	Bates	19	12	8/6**	10	11/2***/7**	12	11/2***/8**
12	Bonvin	18	13	8/5**	12	12/1***/7**	11	12/3***/6**
13	Carbone	19	8	8/1***/7**	13	9/3***/5**	12	11/3***/6**
14	Weng	19	14	6/5**	14	9/1***/6**	12	12/1***/9**
15	Seok	17	16	4/1***/3**	15	8/1***/5**	15	10/2***/5**
16	Grudinin	18	15	5**	16	6/1***/5**	16	8/2***/5**
17	HAWKDOCK	13	16	4/1***/3**	17	5/2***/3**	17	6/2***/4**
18	QASDOM	13	18	3**	18	5**	17	7**

Note: Server groups are listed in all-caps. Target performance shows the number of targets for which an acceptable model or better was submitted, followed by the number of these that were of high (***) or medium (**) quality. For any multi-interface target, the best performance over the interfaces was taken; T149, T150, and T151 are grouped together.

[a]Target participation, out of 20 (for predictors) or 19 (for scorers).

[b]Human and server together.

which 1-2 per group were of medium quality, while only two predictors groups featured the same performance level. Interface 6 (between the two copies of protein B) was also well predicted by scorers. Whereas only three scorer groups and the LZerD server produced 3-4 medium quality models, the number of scorer groups submitting acceptable models or better was higher than for predictors (Table S3).

5.1.3 | Performance across CAPRI predictors, servers and scorers

Groups (predictors, servers, and scorers) were ranked according to their prediction performance for the 20 assembly targets of Round 46. In addition, we ranked participants according to their performance for the easy and difficult targets, respectively. The idea of providing separate performance ranking for different target categories, was repeatedly raised in previous CASP-CAPRI challenges and CAPRI rounds, but was not implemented owing to the fact that the number of targets, notable of difficult targets, was too small to enable a useful assessment. With roughly the same number of targets in the two categories (11 difficult vs 9 easy targets) in Round 46, it seemed worthwhile to also evaluate the performance on the basis of target difficulty as this may help better detect strength and potential weaknesses of the modeling methods used.

All the rankings presented here consider, as usual, the best model submitted by each group among the five top ranking models for each target. The group rankings across targets were performed using the revised ranking protocol, which uses a more balanced weighting scheme for models of different accuracy levels, as detailed in the section on the assessment and ranking procedures. The present rankings differ somewhat from those presented for Round 46 at the CASP13 meeting in Cancun, since they include the assessment results of target **T137/T0956**, which were missing from those presented at the meeting. Other small differences with the "Cancun rankings" were introduced by the revised ranking protocol, which corrected consequential inconsistencies in the scorer rankings, without significantly affecting the rankings of the 10-15 best performing predictor and server groups.

5.1.4 | Performance across all targets

The ranking of participating groups (predictors, scorers, and server) based on their performance across all targets is provided in Table 2.

The five top ranking predictor groups in Round 46 are Venclovas, Fernandez-Recio, Seok, Kihara, Weng, Kozakov, with Weng and Kozakov both ranking fifth. Venclovas ranked first, with a total of 13 out of 20 targets for which this group submitted 6 high-quality, 6 medium-quality, and 1 acceptable model, respectively. The runner-up, Fernandez-Recio submitted correct models for a total of 12 targets, of which 5 were of high-quality, 6 of medium quality and 1 acceptable model. The Seok team submitted correct models for 11 targets, all of which all were of medium (7) or high quality (4), whereas the group of Kihara did nearly as well as the Seok team, by submitting the same number of correct models, but one less medium quality model. The fifth rank position of Kozakov and Weng rewards somewhat differently the achievement by the two groups. Like Venclovas, Kozakov submitted correct models for a total of 13 targets, including 11 medium-quality models and 2 acceptable ones, but no high-quality models. Weng, on the other hand submitted correct models for only 10 targets, but of these 10, five are of high quality and four of medium quality. The higher weight assigned to the higher quality models leads to ranking these two groups equally.

The performance of the Vakser, Huang, Zou, Bates and Chang groups, further down the rank of Table 2 should also be noted. For example, Vakser submitted correct predictions for 11 targets, with a total of 10 models of medium quality or higher, and Zou submitted correct models for 13 targets, of which a smaller number of models (8) where of medium or high quality.

It was rather satisfying to see that the group ranking based on the best 10 submitted models (top-10 in Table 2) differs only marginally from that based on the best 5 models (top-5, Table 2), as this suggests that predictors have improved their ability to rank models in comparison to earlier prediction rounds.

A total of eight servers participated in CAPRI Round 46, four of which did not participate in the CASP12-CAPRI challenge (HDock, MDockPP, GalaxyPPDock, and HawkDock). Six of the servers submitted predictions for all 20 targets. Overall, the server performance was lower than that of "human" predictor groups, likely reflecting the lower performance for the 11 difficult targets, as will be discussed below. The four top-ranking servers were HDock (new, this round), SwarmDock, ClusPro, and LZerD, who submitted correct models for 9 to 12 targets, including as many as 9 to 10 medium quality and 3 to 5 high quality models, each.

The performance of the scorer groups was also lower than that of predictors. The total of 18 scorer groups (including servers) participated in the scoring challenge, which was offered for 19 of the 20 targets. The scoring experiment was not run for **T137/T0956**, and a single scoring experiment was run for all the models submitted for **T149/T0999** and its data driven versions **T150/S0999** and **T151/X0999**.

The five top-performing human scorer groups were Fernandez-Recio, Oliva, Zou, and Chang, followed by Venclovas, Kihara, and Huang, with the latter three groups occupying a shared seventh position in the rank. The groups of Fernandez-Recio, Oliva, and Zou submitted correct models of 12 targets; the models of Fernandez-Recio included 6 and 5 medium and high quality models, respectively, whereas those of Oliva and Zou included a somewhat different mix of medium and high quality models each. Chang, Venclovas, Kihara, and Huang each submitted 11 correct models, which included a different proportion of medium and high quality models.

In all, only five servers participated in the scoring experiments, with three of these, MDockPP, HDock, and LZerD performing rather well. MDockPP and HDock performed on par or better than several of the top human performers, with MDockPP producing correct models for no less than 13 targets (more than other scorer groups), including 10 of medium and high quality. HDock scored correct models for 12 targets no less than 11 of these of medium or high quality. LZerD ranked third, with 9 medium or higher quality models, and 1 or better submitted for 10 targets.

5.1.5 | Performance across easy and difficult targets

Dividing the 20 targets of Round 46 into easy and difficult targets was done mainly in order to identify trends in how human predictors and servers deal with different type of assembly prediction problems. For the majority of the easy targets, high-quality templates were available for the assembly as a whole. The prediction exercise was therefore essentially reduced to the optimization of the homology-built model. For the more difficult targets, predictors and servers were faced with the more standard CAPRI challenge, where an adequate template for the subunit(s) (often only distantly related) must be identified, a homology model built, and the association modes predicted using mainly docking calculations. For both the easy and difficult targets, most CAPRI groups relied on third party software for template identification and homology modeling as will be seen in the next section. For scorers the difference between easy and difficult targets was mainly the level of enrichment in acceptable or higher accuracy models in the combined anonymized set of models to be evaluated, as the latter is directly proportional to the probability of identifying correct models by chance.

The separate performance ranking of predictors for the easy and difficult targets is listed in Tables S4 and S5. The top 10 performing groups for the easy targets are virtually the same as for all targets (Table 2), with however negligible differences in the exact rank position. The exception are the performances of Weng and Kozakov, who rank fifth when all targets are considered, but ninth on the ranking for the easy targets. Unsurprisingly, this indicates that the performance over all targets is, in general, dominated by the performance for the easy targets. Exceptions such as that of the groups of Weng and Kozakov are quite interesting. Both rank ninth as predictors of easy targets, but move up to second position in the rank for difficult targets, which propels them to the fifth position in the rank for all targets. Such cases suggest that the corresponding groups have better methods for dealing with difficult targets where the performance of docking algorithms is more important, than when mostly template-based modeling needs to be mastered.

It is also noteworthy that a number of CAPRI predictor groups seem to be at relative ease with both types of approaches. The group of Venclovas ranked first for the predictions of both the easy and difficult targets, and thus for all targets combined, with several other groups (Fernandez-Recio, Seok, Kozakov, Kihara, and Zou) likewise performing well in both target categories and therefore also overall.

The analysis of the scorer performance (human and servers) for the two target categories is also informative (Tables S4 and S5). However, since most scorer groups successfully predicted the same limited subset of difficult targets, multiple groups were ranked at the same level for these targets, making it more difficult to identify trends. Nonetheless, we see for example that the three best-performing scorer groups (Fernandez-Recio, Oliva, Zou) in the global ranking, also rank among the best performers for both the easy and difficult targets. A number of other scorer groups performed differently between the two target categories, with some groups, such as Seok, Kihara, Bonvin, and Bates, ranking higher for the difficult targets than for the easy ones. This seems to suggest that their scoring functions are better at singling out correct models (binders) from incorrect alternatives (nonbinders) than discriminating between correct models displaying different accuracy levels (acceptable vs medium or high accuracy).

Lastly, we confirm that the global performance of servers was dominated by the ability to predict the easy targets, as indeed the rankings

of prediction servers for the easy targets (Table S4) and for all the targets (Table 2) were very similar. The server performance for the difficult targets therefore played only a marginal role, but we do note that none of the top three servers in either list (HDock, SwarmDock, Haddock for the easy targets; GalaxyPPDock, LZerD, ClusPro for the difficult targets) occupies a top three position in both lists.

As far as scoring servers are concerned, the best performing servers overall (in order: MDockPP, HDock, LZerD; Table 2) owe their high rank relative to other scorer groups to their good performance for the difficult targets (Table S4).

5.2 | Prediction of binding interfaces

Interface predictions were evaluated for 47 binary association modes in the top 5 scoring models submitted for 22 targets by CAPRI predictors groups (human and servers), as well as for 36 binary association modes in the top 5 models submitted by CAPRI scorer groups (human and server) for 19 targets. The correspondence between the residues defining the interfaces of the individual protein components of each binary association mode in the predicted models and those in the target structure was evaluated using the *Recall* and *Precision* measures (see section 4).

5.2.1 | Global trends

Figure 4 presents scatter plots of the recall and precision values of predicted interfaces for components (receptor and ligand, respectively) of the top 5 models submitted for each of the 47 evaluated association modes by predictor and scorer groups. Individual points represent values averaged separately over interfaces of association modes in each of the four categories (incorrect, acceptable, medium, and high) submitted by a given group for a given target.

Inspection of the scatter plots reveals that predicted interfaces in the models submitted by both predictors (Figure 4A) and scorers (Figure 4B) span a wide range of recall and precision values. Confirming our previous reports[41,47] we observe that a sizable fraction of the points corresponding to interfaces of incorrect models cluster loosely along the diagonal at very low values, whereas the vast majority of acceptable and higher quality models feature interfaces with recall and precision values ≥50% (upper-right quadrant of the scatter plots in Figure 4), which we designate here as correct interface predictions. In addition, a sizable fraction of the points in Figure 4 is spread widely above and below the diagonal. Like in previous analyses, a higher fraction of interfaces in scorer models (all quality levels) tend to have higher recall (55% of the interfaces) than precision values (27% of the interfaces, Figure 4B). On the other hand, interfaces of predictor models show little preference (Figure 4A). Among the latter interfaces about 37% feature higher recall than precision, ~35% feature higher precision than recall, and ~28% have equal recall and precision values.

We likewise confirm that, (a) a fraction of incorrect models feature in fact correctly predicted interfaces and (b) a fraction of correctly predicted interfaces corresponds to incorrect models.[41,47] Intriguingly however, in Round 46, the fraction of correctly predicted interfaces in incorrect models has gone down to

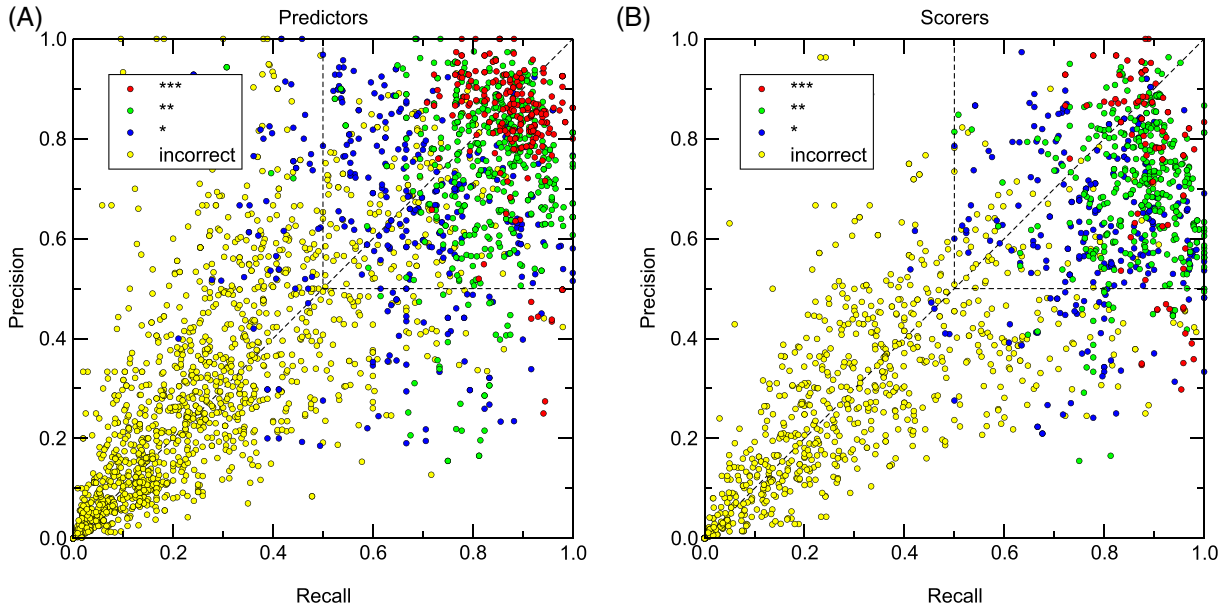

FIGURE 4 Global landscape of the interface prediction performance. Scatter plot showing the average Recall and Precision values (see main text for definition) of the interfaces in models submitted by all Predictors (A) and Scorers (B) for the 22 targets of Round 46. Each point represents average values for the interfaces of individual protein components in models submitted by individual participants for one association mode. Averaging was performed separately over models in the four CAPRI accuracy categories (incorrect, acceptable, medium, and high). Individual points are color-coded by the CAPRI model quality category (as indicated in the legend displayed in the upper left corner of each graph). The upper right hand quadrant of the graph, with Recall and Precision values above 0.5, contains all points corresponding to "correct" interface predictions

~11-12% (11.35 for predictors and 11.9% for scorers) from 16%, in the CASP12-CAPRI challenge[41] and 24% in the initial CAPRI evaluation of 2010.[47] In parallel, the fraction of incorrect assembly models in the submissions with correctly predicted interfaces decreased to 19%, from 27.2% in the CASP12-CAPRI challenge. These trends reflect a more general decline in interface prediction performance. Indeed, the fractions of acceptable and higher quality models featuring correctly predicted interfaces are now 70% and 92%, respectively, down from 87% and 98%,[41] and from 92% and 100%, respectively, in the 2010 evaluation.[47] Thus, acceptable models and, surprisingly, also models of medium quality or better submitted in Round 46 feature a significantly larger fraction of incorrectly predicted interfaces than previously documented.

Insights into the origins of these trends can be obtained from the scatter plots of Figure 4. These plots show indeed that a significant fraction of the correct assembly models correspond to points located above and below the diagonal. Points above the diagonal, which feature higher precision than recall values, correspond to predicted interfaces of smaller size that capture only a fraction of the native interfaces, but little else, and may hence be of predictive value. Interfaces with lower precision than recall values, corresponding to points located below the diagonal, and more particularly the points in the lower left quadrant of the plots in Figure 4 are problematic. Strikingly, a number of these latter points correspond to medium- and high-quality assembly models with interfaces featuring high recall values between 0.6 and 1.0, but nonetheless very low precision (less than 40%). While such predicted interfaces capture rather well the native interface, they also include a large fraction (0.6-0.8) of "false-positives," for example, residues incorrectly predicted to be part of the target interface, which drastically reduces the predictive value of the corresponding assembly models.

Closer examination of some of major outliers in the plots of Figure 4, primarily those corresponding to very low precision and high recall values, revealed that the corresponding assembly models were for targets where more drastic adjustments of the template conformation were required in order to correctly model the assembly. Examples of such cases include the high and medium quality models submitted for **T149/T0999** and **T151/X0999**, both corresponding to the same difficult multidomain homodimer of the pentafunctional AROM polypeptide (Table 1B). Other cases are the medium quality models for the heterodimer **T146/H0993**, another difficult target for which only distant templates were available for individual protein subunits, but also for **T147/T0995** and **T158/T1020**, two targets with higher-order assembly modes, that were classified as easy targets since adequate templates were available (Table 1). Analysis of several of these models indicates that the predicted interfaces tend to include portions of the modeled subunits that did not belong to the native interface, likely due to an effort to maximize the interface size. None of these models exceeded the allowed level of atomic clashes, which is closely monitored in the evaluated models and may be cause for disqualifying the submission.[43,44]

5.2.2 | Performance of predictor server and scorer groups

The ranking of groups based on the interface prediction performance is listed in Table S6. Group performance was ranked based on the fraction of correctly predicted interfaces (interfaces with both recall and precision ≥0.5), in the top five submitted models for each target. Nine human predictor groups (Venclovas, Eisenstein, Seok, Bates, Fernandez-Recio, Chang, Zou, Kozakov and Kihara) and four prediction servers (Haddock, SwarmDock, HDOCK, MDockPP) submitted correct predictions for at least 30% of the interfaces. The best performing predictor groups were Venclovas with correct predictions for 44.6% of the evaluated interfaces and Eisenstein with correct predictions for 43.3% of the predicted interfaces. The best-performing prediction servers Haddock and SwarmDock performed less well, with correct predictions for 35.4% and 33.5% of the interfaces, respectively. The winners of the interfaces prediction challenge were the scorers, both human and servers. Ten human scorer groups submitted correct predictions for at least 30% of the interfaces. Of these, four groups (Oliva, Venclovas, Huang, Zou) achieved correct predictions for at least 40% of the interfaces, with the groups of Oliva and Venclovas topping the rank with 47.3% and 43.5% of the interfaces correctly predicted, respectively. A total of four scoring servers submitted correct predictions for at least 30% of the interfaces, of which MDockPP and LZerD performed best, both with about 40% of correctly predicted interfaces.

The last four columns of Table S6 list the average recall and precision values for interfaces of individual models (top 5) submitted by each group, as well as the corresponding standard deviations. It is noteworthy that the average recall and precision values achieved by the best performing groups or servers rarely exceed 50%, compared to 60% in the CASP12-CAPRI challenge.[41] The standard deviations are also larger, routinely exceeding 30%, compared to previous values of about 25%. These results indicate that models for individual targets (even those by the best performing groups) tend to vary substantially in terms of the interface prediction accuracy, and that the interface prediction accuracy has in general declined, relative to achievements in previous CAPRI Rounds.

Lastly, it should be noted that most published interface prediction methods reach average recall and precision levels of ~50% and ~25%, respectively, when applied to transient complexes (see Reference 61 for review). The best-performing groups of Round 46 achieve similar recall levels but significantly higher precision (45-56%) (Table S6), still suggesting that interface prediction methods which model the association modes with the cognate binding partner retain an advantage over most extant interface prediction methods, which do not use such information.

5.3 | Factors influencing the prediction performance

Round 46 comprised 20 targets and these targets spanned a wide range of modeling difficulties. By the CAPRI management choice, the majority of the targets had some templates available in the PDB. The

majority of the targets were homo-oligomers—mostly homodimers. For a significant fraction of these targets (the easy targets) the assembly prediction task boiled down to template-based modeling of the entire complex and model refinement. The prediction of the more difficult targets required modeling the structures of individual subunits, followed by docking calculations and usually some form of model refinement.

Critical factors influencing the prediction performance were therefore (a) the ability to identify templates whose 3D structure and association modes were close enough to those of the target, to enable building an accurate model of the target assembly, and (b) the extent to which these models were adequately optimized.

The influence that model accuracy of individual subunits had on the assembly prediction performance can be gleaned from Figure 5, which displays the M-rms values (the backbone rms values of the individual subunits of the submitted models vs those of the target). For the majority of the easy targets, these values rarely exceed 2.3-3 Å, whereas the models for the difficult targets feature much higher M-rms values. For the more poorly predicted heterocomplexes **T155**, **T156**, and **T157**, M-rms values for a least one of the subunits displays a significant spread into higher values (10-12 Å), culminating at values as high as 25 Å for the partially unstructured subunit C of T159. High M-rms values (10-15 Å) are also displayed for domain B of the multidomain AROM polypeptide (**T149/T0999**), for which only poor templates were available, although a few predictors nonetheless succeeded in generating acceptable models for the interface involving this domain.

Clearly, identifying the most adequate template is often not an easy task, as multiple templates are often available either for the full complex or for the independent subunits, requiring adequate strategies for exploiting these data. As can be seen from the summaries by the individual CAPRI groups co-authoring this paper (see Supplementary Material), a variety of approaches were used to tackle this important step. A number or groups successfully exploited homology models generated by the best performing CASP13 servers and made available during the prediction Rounds, or used third party tools such as Modeller.[62] Successful approaches involved searching a database of known structures, clustered on the basis of sequence and structure similarity, and relying on various scoring schemes to select the most suitable templates, or a reduced set of templates, for further refinement. Querying the PPI3D web server[63] for suitable subsets of templates by the group of Venclovas, or running HHblits[45] against a sequence profile database of known structures clustered at 70% sequence identity, as done by the Bates group, are good examples of such approaches.

Further filtering and refining models built from identified templates is likewise important, and here too, different approaches were rather successful (see supplementary section on Individual Group Summaries). For example, the group of Bates used fragments from different templates coupled with optimization techniques employing biophysical force fields and information on residue contacts, whereas fragment-guided molecular dynamics was used by Venclovas. For some targets, close integration of classical template-based modeling with docking calculation, as done by the group of Fernandez-Recio, was likewise quite effective.

Several of the best-performing CAPRI groups also highlighted the importance of specialized, often custom-developed, functions for scoring and ranking protein-protein interfaces for the entire modeled assembly. But the type of functions differed substantially between participants. Examples are the VoroMQA score developed by the Venclovas group,[64] the combined use of three scoring functions, GOAP,[65] Dfire,[66] and ITScore[67] by the Kihara group, or the multiterm scoring function of the

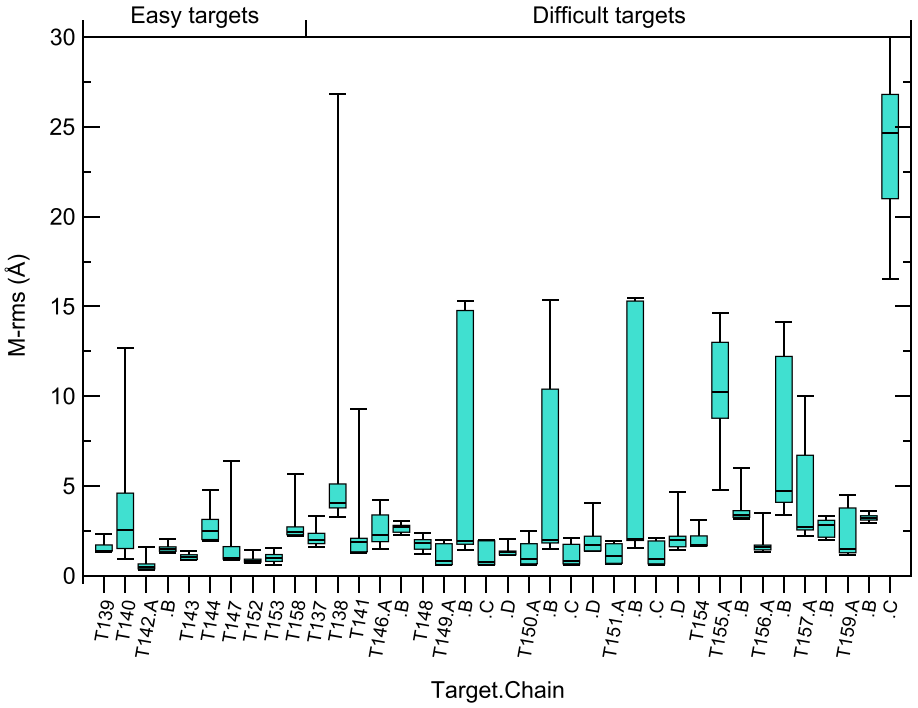

FIGURE 5 Model quality of individual protein subunits in assembly models of Round 46. Shown are whisker plots (displaying the median, 1st and 3rd quartile, and 9th and 91st percentile) representing the distributions of M-rms values of individual protein subunits in models submitted for each of the targets of Round 46. Targets are labeled by their CAPRI target number; distributions for the easy targets are shown on the left side of the graph, and those for the difficult targets are shown on the right side. For all homomeric targets only one subunit was analyzed, except for the multidomain homodimer targets T149/T150/T151, where individual structural domains (A-D) were considered. For the heterodimer targets (T142, T146, T155, T156, T157) two subunits (A, B) were analyzed and for hetero-18-mer (T159) 3 subunits (A, B, C) were evaluated

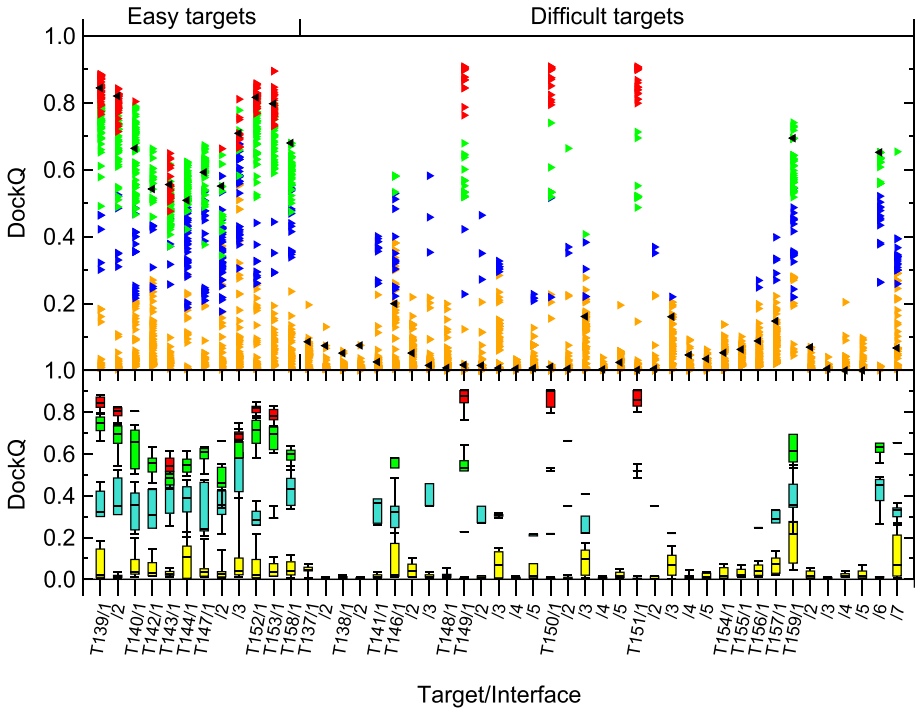

FIGURE 6 Global overview of the prediction performance for targets of Round 46. Shown are the distributions of the DockQ values computed for the top-five models submitted by all predictor groups for individual targets of Round 46. The targets are labeled by their CAPRI target number and interface rank. Distributions for the easy targets are shown on the left side of the graph, and those for the difficult targets are shown on the right side. Individual points are color-coded according to the CAPRI model quality category; yellow: incorrect; blue: acceptable; green: medium; red: high. For each target, a baseline-level prediction, represented by the best model of the top-performing automatic server (HDock, see Table 2), is represented by black triangles. The boxplot distributions (whiskers at 9th and 91st percentiles) of each target and prediction category are shown on the lower panel; color-coding is as for the upper panel, but with a lighter shade of blue for better visibility

Vakser group, additionally complemented with sequence-based measures for individual subunits[68] and with functional annotations. The quite successful scoring performance of the Oliva group relied on their CONSRANK-based methods to score and rank multiple models, based on the most frequent interface residue contacts observed in these models.[69]

For the more difficult targets (Table 1), the full assembly was predicted using models of the individual subunit, built on the basis of more distantly related templates and performing "pure" (ab-initio) docking calculations. Interestingly, a number of groups relied on reputable CAPRI docking servers such as CLUSPRO[57] and/or algorithms such ZDOCK,[35] or HEX,[70] developed by other groups, to generate their docking poses. Some teams, like that of Grudinin/Laine/Carbone, exploited the fast sampling speed of the HEX and SAM[71] docking programs to perform cross-docking calculations, whereby sets of models are docked to one another, yielding a large set of assembly models that are then scored and optimized. Increasing use was also made of docking algorithms that incorporate symmetry operations (eg, HSYMDOCK-lite[72]), or of algorithms that handle multiple chains (eg, Multi-LZerD[73]) or better account for conformational flexibility. But ultimately the performance crucially depended on how similar the homology-built independent subunits were to those of the target.

For the difficult homodimer targets, failures were mainly attributable to the availability of very poor and often incomplete templates. A combination of factors contributed to the poor prediction performance for **T159/H1020**, the large 18-mer heterocomplex (Figure 3): the partial template available for the nonglobular subunit (C), the intertwined association modes formed by this subunit with its neighbors in the complex, and the large number of interfaces that all needed to be accurately modeled. The latter problem also hampered the accurate modeling of the multidomain homodimer of **T149/T0999**, despite the availability of good quality templates for three of the four independent structural domains of the protein. These results indicate yet again that modeling large-order protein assemblies in absence of adequate templates for the full assembly remains a major challenge, especially when symmetry operations cannot be applied to all the components, as for the intersubunit multidomain association of **T149/T0999**.

6 | CONCLUDING REMARKS

This report presented an assessment of the assembly prediction results for CAPRI Round 46, the CASP13-CAPRI challenges held during the summer of 2018. The 20 targets of Round 46 included 6 heterocomplexes, a larger number than previously, in addition to 14 homo-oligomers, still representing the majority.

The CAPRI management selected these targets as those with structural templates in the PDB, which therefore represented tractable modeling problems for the CAPRI community. But the selection criteria were somewhat relaxed this time, allowing the inclusion of a

significant number of more challenging targets than in previous joint experiments. These comprised some large complex assemblies and those with significantly poorer templates. Nevertheless, the larger total number of targets, of which a significant fraction was more difficult to model, allowed us to evaluate not only the prediction performance across all targets, as done previously, but also to measure how groups performed on the roughly similar number of easy (9/20) and more difficult (11/20) targets, respectively.

A global overview of the quality of models submitted by predictor groups for the two targets categories is presented in Figure 6. The top panel of Figure 6 displays the DockQ scores, color-coded by the CAPRI model quality categories for all the interfaces in individual models submitted for the 22 targets of Round 46 (including the 2 data-assisted versions of **T149/T0999**). These scores are contrasted with those obtained for the best of the five models submitted by HDock, the top performing automatic server in this evaluation, which we use as the baseline performance, analogous to that produced by the "naïve" predictions considered previously.[41] The lower panel of Figure 6 represents the same data using box plots, illustrating the score distributions per model quality and target interface. Not too surprisingly, the prediction performance, as measured by the fraction of models of acceptable quality or higher submitted across the ~30 human predictor and server groups, was very good to excellent for the nine easy targets, comprising mostly homomers for which templates were available for the entire assembly For this category of targets the baseline predictions produced by the HDock automatic server were in general on par with the best performing manual predictors. On the other hand, a much lower performance was achieved for the 11 difficult targets. For example, whereas top predictor groups submitted quite accurate models (medium and high quality) for all of the 9 easy targets, only 3 of the 11 difficult targets were predicted at a similar level of accuracy by top performers and only for one of their interfaces (see also Table S3). For four of the 11 difficult targets, only incorrect models were submitted for either interface. The automatic server produced incorrect models for 10 out of the 11 difficult targets, including the primary interface of T149/T0999, for which high quality models were produced by the top manual predictors (Figure 6). It successfully predicted interfaces 1 and 6 of T159, two of the easier interfaces of this target, submitting medium quality models of similar quality to that obtained by the manual predictions, while failing to predict the third "easy" interface of T159.

This prediction "gap" for easy vs difficult targets was also apparent in the performance of scorers, the ~17 groups participating in the scoring experiments. Scorers performed very well and on par with predictor groups for the easy targets. But their performance was weak for the difficult targets, likely due to the much lower fraction of correct models in the uploaded set.

Thus, the performance of predictors and scorer groups on the set of easy targets weighed heavily on their ranking for the full set of targets in Round 46. But ranking separately the performance of predictor, server and scorer groups on the easy and difficult targets (Tables S4 and S5), respectively, led to interesting observations. Although the lists of top 5 to 10 performing groups for the two target categories overlapped significantly, several groups such as those of Shen, Weng, Kozakov, or Huang, performed better than their colleagues on the difficult targets, but ranked lower on the easy ones. Since most of the difficult targets involved ab-initio docking of homology built models, the expertise in ab-initio docking and scoring of these groups was probably a determining factor. A number of scorer groups also performed differently between the two target categories, providing useful insights into the strength and weaknesses of their scoring functions. For more detailed information on factors potentially influencing the performance of individual groups see Supplementary Material (Individual Group Summaries).

Analyzing how well predictor and scorer groups were able to identify the residues on each of the interacting subunits that contribute to the recognition interfaces also led to useful observations. Overall the average interface prediction performance achieved in Round 46 was significantly lower than previously (eg, in the CASP12-CAPRI challenge). This might be due to the larger number of poor models submitted for the difficult targets. However, a significant number of submitted medium and high quality models had poorly predicted interfaces nonetheless. In particular, some of these interfaces were extensively "overpredicted" and included a large fraction of "false positives"; residues not belonging to the target interface. Although this surprisingly high degree of interface "over prediction" occurred most frequently for models of difficult targets, it indicates that the criteria used by many predictors to score and rank models remain suboptimal. It likewise suggests that the CAPRI evaluation criteria should routinely incorporate $f_{non\text{-}nat}$, the fraction of non-native contacts in the predicted interface, in addition to the f_{nat}, the fraction of native contacts. This option is currently under discussion with the CAPRI community.

Finally, the following main general conclusions can be drawn from the present evaluation. Modeling of homo-oligomers, especially homodimers, when templates for the full assembly are available, is a problem that can be tackled by many groups, but highly accurate models are an exception rather than the rule, indicating that further efforts should be devoted to better model refinement. The prediction of targets for which good templates for individual subunits are available is increasingly successful, thanks to more efficient docking algorithms and better exploitation of template data, although, here too, model refinement remains suboptimal.

On the other hand, generating accurate 3D structures of assemblies for which only distantly available templates are available, remains out of reach for modeling tools such as those currently available to the CAPRI community. To tackle the very challenging problem of predicting protein assemblies from sequence information and limited prior information on the structures of the individual subunits, novel tools are needed. These tools must closely integrate sequence information with 3D as well as quaternary structure prediction, a very valid justification to continue bringing the CASP and CAPRI communities together in the future. Likewise, the protocol for scoring and ranking models of higher order assemblies, which currently takes into account only the best-predicted interface of the assembly, is clearly suboptimal as it does not reflect the quality of the full predicted complex. A possible approach might be to combine the scores for individual interfaces with those that measure the relative displacements of the interacting subunits.

ACKNOWLEDGEMENTS

We thank the CASP Management and in particular Andriy Kryshtafovych, for valuable help and support in running the assembly prediction challenge. We also express gratitude to the structural biologists who provided the targets for this challenge and to the CAPRI management team and predictor groups for stimulating discussion, valuable input, and cooperation.

[Correction added on 11 Nov 2019, after first online publication: Preceding ORCID ID added.]

ORCID

Marc F. Lensink https://orcid.org/0000-0003-3957-9470
Raphaël A. G. Chaleil https://orcid.org/0000-0003-2759-2088
Paul A. Bates https://orcid.org/0000-0003-0621-0925
Alessandra Carbone https://orcid.org/0000-0003-2098-5743
Shan Chang https://orcid.org/0000-0001-7169-9398
Cezary Czaplewski https://orcid.org/0000-0002-0294-3403
Sheng-You Huang https://orcid.org/0000-0002-4209-4565
Mireia Rosell https://orcid.org/0000-0002-7787-8896
Juan Fernandez-Recio https://orcid.org/0000-0002-3986-7686
Charles Christoffer https://orcid.org/0000-0002-6163-6323
Daisuke Kihara https://orcid.org/0000-0003-4091-6614
Dima Kozakov https://orcid.org/0000-0003-0464-4500
Iain H. Moal https://orcid.org/0000-0002-4960-5487
Isaure Chauvot de Beauchêne https://orcid.org/0000-0002-7035-3042
Yang Shen https://orcid.org/0000-0002-1703-7796
Chaok Seok https://orcid.org/0000-0002-1419-9888
Justas Dapkūnas https://orcid.org/0000-0002-0496-6107
Česlovas Venclovas https://orcid.org/0000-0002-4215-0213
Ilya A. Vakser https://orcid.org/0000-0002-5743-2934
Thom Vreven https://orcid.org/0000-0003-3413-806X
Zhiping Weng https://orcid.org/0000-0002-3032-7966
Cunliang Geng https://orcid.org/0000-0002-1409-8358
Alexandre M. J. J. Bonvin https://orcid.org/0000-0001-7369-1322
Shoshana J. Wodak https://orcid.org/0000-0002-0701-6545

REFERENCES

1. Ideker T, Sharan R. Protein networks in disease. *Genome Res*. 2008;18(4):644-652.
2. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet*. 2011;12(1):56-68.
3. Berman HM, Battistuz T, Bhat TN, et al. The protein data bank. *Acta Crystallogr Sect D: Biol Crystallogr*. 2002;58(1):899-907.
4. Lo Conte L, Chothia C, Janin J. The atomic structure of protein-protein recognition sites. *J Mol Biol*. 1999;285(5):2177-2198.
5. Dey S, Pal A, Chakrabarti P, Janin J. The subunit interfaces of weakly associated homodimeric proteins. *J Mol Biol*. 2010;398(1):146-160.
6. Ponstingl H, Kabir T, Gorse D, Thornton JM. Morphological aspects of oligomeric protein structures. *Prog Biophys Mol Biol*. 2005;89(1):9-35.
7. Nooren IM, Thornton JM. Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol*. 2003;325(5):991-1018.
8. Bai XC, McMullan G, Scheres SH. How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci*. 2015;40(1):49-57.
9. Cheng Y, Glaeser RM, Nogales E. How Cryo-EM became so hot. *Cell*. 2017;171(6):1229-1231.
10. 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(24):9438-9441.
11. Berman HM, Coimbatore Narayanan B, Di Costanzo L, et al. Trendspotting in the protein data Bank. *FEBS Lett*. 2013;587(8):1036-1045.
12. Ovchinnikov S, Park H, Varghese N, et al. Protein structure determination using metagenome sequence data. *Science*. 2017;355(6322):294-298.
13. Marks DS, Hopf TA, Sander C. Protein structure prediction from sequence variation. *Nat Biotechnol*. 2012;30(11):1072-1080.
14. Ovchinnikov S, Kim DE, Wang RY, Liu Y, DiMaio F, Baker D. Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins*. 2016;84(suppl 1):67-75.
15. Jones DT, Singh T, Kosciolek T, Tetchner S. MetaPSICOV: combining coevolution methods for accurate prediction of contacts and long range hydrogen bonding in proteins. *Bioinformatics*. 2015;31(7):999-1006.
16. Kocher JP, Rooman MJ, Wodak SJ. Factors influencing the ability of knowledge-based potentials to identify native sequence-structure matches. *J Mol Biol*. 1994;235(5):1598-1613.
17. Rooman MJ, Kocher JP, Wodak SJ. Prediction of protein backbone conformation based on seven structure assignments. Influence of local interactions. *J Mol Biol*. 1991;221(3):961-979.
18. Rooman MJ, Kocher JP, Wodak SJ. Extracting information on folding from the amino acid sequence: accurate predictions for protein regions with preferred conformation in the absence of tertiary interactions. *Biochemistry*. 1992;31(42):10226-10238.
19. Rooman M, Wodak SJ. Generating and testing protein folds. *Curr Opin Struct Biol*. 1993;3(2):247-259.
20. Min S, Lee B, Yoon S. Deep learning in bioinformatics. *Brief Bioinform*. 2017;18(5):851-869.
21. Wang S, Sun S, Li Z, Zhang R, Xu J. Accurate de novo prediction of protein contact map by ultra-deep learning model. *PLoS Comput Biol*. 2017;13(1):e1005324.
22. Guzenko, D, Lafita, A, Monastyrskyy, B, Kryshtafovych, A, Duarte, JM. Assessment of protein assembly prediction in CASP13. *Proteins*. 2019;87(12):1190-1199. https://doi.org/10.1002/prot.25795
23. Whitehead TA, Baker D, Fleishman SJ. Computational design of novel protein binders and experimental affinity maturation. *Methods Enzymol*. 2013;523:1-19.
24. Whitehead TA, Chevalier A, Song Y, et al. Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat Biotechnol*. 2012;30(6):543-548.
25. Silva DA, Yu S, Ulge UY, et al. De novo design of potent and selective mimics of IL-2 and IL-15. *Nature*. 2019;565(7738):186-191.
26. Yeates TO. Geometric principles for designing highly symmetric self-assembling protein nanomaterials. *Annu Rev Biophys*. 2017;46:23-42.
27. Ward AB, Sali A, Wilson IA. Biochemistry. Integrative structural biology. *Science*. 2013;339(6122):913-915.
28. Webb B, Viswanath S, Bonomi M, et al. Integrative structure modeling with the integrative modeling platform. *Protein Sci*. 2018;27(1):245-258.
29. Wodak SJ, Janin J. Structural basis of macromolecular recognition. *Adv Protein Chem*. 2002;61:9-73.
30. Ritchie DW. Recent progress and future directions in protein-protein docking. *Curr Protein Pept Sci*. 2008;9(1):1-15.
31. Vajda S, Kozakov D. Convergence and combination of methods in protein-protein docking. *Curr Opin Struct Biol*. 2009;19(2):164-170.

32. Fleishman SJ, Whitehead TA, Strauch EM, et al. Community-wide assessment of protein-interface modeling suggests improvements to design methodology. *J Mol Biol*. 2011;414(2):289-302.
33. Moretti R, Fleishman SJ, Agius R, et al. Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions. *Proteins*. 2013;81(11):1980-1987.
34. Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. *Proteins*. 2013;81(12):2082-2095.
35. Lensink MF, Moal IH, Bates PA, et al. Blind prediction of interfacial water positions in CAPRI. *Proteins*. 2014;82(4):620-632.
36. Xu Q, Canutescu AA, Wang G, Shapovalov M, Obradovic Z, Dunbrack RL Jr. Statistical analysis of interface similarity in crystals of homologous proteins. *J Mol Biol*. 2008;381(2):487-507.
37. Levy ED, Teichmann S. Structural, evolutionary, and assembly principles of protein oligomerization. *Prog Mol Biol Transl Sci*. 2013; 117:25-51.
38. Negroni J, Mosca R, Aloy P. Assessing the applicability of template-based protein docking in the twilight zone. *Structure*. 2014;22(9): 1356-1362.
39. Szilagyi A, Zhang Y. Template-based structure modeling of protein-protein interactions. *Curr Opin Struct Biol*. 2014;24:10-23.
40. Lensink MF, Velankar S, Kryshtafovych A, et al. Prediction of homo-protein and heteroprotein complexes by protein docking and template-based modeling: a CASP-CAPRI experiment. *Proteins*. 2016; 84(suppl 1):323-348.
41. Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ. The challenge of modeling protein assemblies: the CASP12-CAPRI experiment. *Proteins*. 2018;86(suppl 1):257-273.
42. Lafita A, Bliven S, Kryshtafovych A, et al. Assessment of protein assembly prediction in CASP12. *Proteins*. 2018;86(suppl 1):247-256.
43. Lensink MF, Mendez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd edition. *Proteins*. 2007;69(4):704-718.
44. Lensink MF, Wodak SJ. Docking and scoring protein interactions: CAPRI 2009. *Proteins*. 2010;78(15):3073-3084.
45. Remmert M, Biegert A, Hauser A, Soding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. 2012;9(2):173-175.
46. Zimmermann L, Stephens A, Nam SZ, et al. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J Mol Biol*. 2018;430(15):2237-2243.
47. Lensink MF, Wodak SJ. Blind predictions of protein interfaces by docking calculations in CAPRI. *Proteins*. 2010;78(15):3085-3095.
48. Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol*. 2007;372(3):774-797.
49. Bliven S, Lafita A, Parker A, Capitani G, Duarte JM. Automated evaluation of quaternary structures from protein crystals. *PLoS Comput Biol*. 2018;14(4):e1006104.
50. Soding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics*. 2005;21(7):951-960.
51. Hildebrand A, Remmert M, Biegert A, Soding J. Fast and accurate automatic structure prediction with HHpred. *Proteins*. 2009;77(suppl 9):128-132.
52. Lensink MF, Velankar S, Wodak SJ. Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition. *Proteins*. 2017;85(3):359-377.
53. Berchanski A, Eisenstein M. Construction of molecular assemblies via docking: modeling of tetramers with D2 symmetry. *Proteins*. 2003;53 (4):817-829.
54. Pierce B, Tong W, Weng Z. M-ZDOCK: a grid-based approach for Cn symmetric multimer docking. *Bioinformatics*. 2005;21(8):1472-1478.
55. Ritchie DW, Grudinin S. Spherical polar Fourier assembly of protein complexes with arbitrary point group symmetry. *J Appl Cryst*. 2016; 49:158-167.
56. Mashiach E, Schneidman-Duhovny D, Peri A, Shavit Y, Nussinov R, Wolfson HJ. An integrated suite of fast docking algorithms. *Proteins*. 2010;78(15):3197-3204.
57. Kozakov D, Hall DR, Xia B, et al. The ClusPro web server for protein-protein docking. *Nat Protoc*. 2017;12(2):255-278.
58. Comeau SR, Camacho CJ. Predicting oligomeric assemblies: N-mers a primer. *J Struct Biol*. 2005;150(3):233-244.
59. Basu S, Wallner B. DockQ: a quality measure for protein-protein docking models. *PLoS One*. 2016;11(8):e0161879.
60. Peterson LX, Roy A, Christoffer C, Terashi G, Kihara D. Modeling disordered protein interactions from biophysical principles. *PLoS Comput Biol*. 2017;13(4):e1005485.
61. de Vries SJ, Bonvin AM. How proteins get in touch: interface prediction in the study of biomolecular complexes. *Curr Protein Pept Sci*. 2008;9(4):394-406.
62. Webb B, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol*. 2017;1654:39-54.
63. Dapkunas J, Timinskas A, Olechnovic K, Margelevicius M, Diciunas R, Venclovas C. The PPI3D web server for searching, analyzing and modeling protein-protein interactions in the context of 3D structures. *Bioinformatics*. 2017;33(6):935-937.
64. Olechnovic K, Venclovas C. VoroMQA: assessment of protein structure quality using interatomic contact areas. *Proteins*. 2017;85(6):1131-1145.
65. Zhou H, Skolnick J. GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. *Biophys J*. 2011;101(8):2043-2052.
66. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci*. 2002;11(11):2714-2726.
67. Huang SY, Zou X. Statistical mechanics-based method to extract atomic distance-dependent potentials from protein structures. *Proteins*. 2011;79(9):2648-2661.
68. Kundrotas PJ, Anishchenko I, Badal VD, Das M, Dauzhenka T, Vakser IA. Modeling CAPRI targets 110-120 by template-based and free docking using contact potential and combined scoring function. *Proteins*. 2018;86(suppl 1):302-310.
69. Chermak E, Petta A, Serra L, et al. CONSRANK: a server for the analysis, comparison and ranking of docking models based on inter-residue contacts. *Bioinformatics*. 2015;31(9):1481-1483.
70. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. *Proteins*. 2000;39(2):178-194.
71. El Houasli M, Maigret B, Devignes MD, Ghoorah AW, Grudinin S, Ritchie DW. Modeling and minimizing CAPRI round 30 symmetrical protein complexes from CASP-11 structural models. *Proteins*. 2017; 85(3):463-469.
72. Yan Y, Tao H, Huang SY. HSYMDOCK: a docking web server for predicting the structure of protein homo-oligomers with Cn or Dn symmetry. *Nucleic Acids Res*. 2018;46(W1):W423-W431.
73. Esquivel-Rodriguez J, Yang YD, Kihara D. Multi-LZerD: multiple protein docking for asymmetric complexes. *Proteins*. 2012;80(7):1818-1833.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Lensink MF, Brysbaert G, Nadzirin N, et al. Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins*. 2019;87: 1200–1221. https://doi.org/10.1002/prot.25838