CAPRI-Q: The CAPRI resource for assessment of modeling protein interactions

Keeley W. Collins,[1] Matthew M. Copeland,[1] Guillaume Brysbaert,[3] Shoshana J. Wodak,[4] Alexandre M.J.J. Bonvin,[5] Petras J. Kundrotas,[1]* Ilya A. Vakser,[1,2]* and Marc F. Lensink[3]*

[1] Computational Biology Program, and [2]Department of Molecular Biology, The University of Kansas, Lawrence, KS 66045, USA

[3]Univ. Lille, CNRS, UMR 8576 - UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, F-59000 Lille, France

[4]VIB-VUB Center for Structural Biology, Brussels, Belgium

[5]Bijvoet Centre for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, The Netherlands

*Corresponding authors:

Marc Lensink, University of Lille, marc.lensink@univ-lille.fr.

Ilya Vakser, University of Kansas, vakser@ku.edu,

Petras Kundrotas, University of Kansas, pkundro@ku.edu,

Keywords: protein docking, CAPRI, quality assessment, blind prediction, protein modeling

ABSTRACT

Protein interactions are essential for cellular processes. In recent years there has been significant progress in computational prediction of 3D structures of individual protein chains, with the best-performing algorithms reaching sub-Angstrom accuracy. These techniques are now finding their way into the prediction of protein interactions, adding to the existing modeling approaches. The community-wide Critical Assessment of Predicted Interactions (CAPRI) has been a catalyst for the development of procedures for the structural modeling of protein assemblies by organizing blind prediction experiments. The predicted structures are assessed against unpublished experimentally determined structures using a set of metrics with proven robustness that have been established in the CAPRI community. In addition, several advanced benchmarking databases provide targets against which users can test docking and assembly modeling software. These include the Protein-Protein Docking Benchmark, the CAPRI Scoreset, and the DOCKGROUND database, all developed by members of the CAPRI community. Here we present CAPRI-Q, a stand-alone model quality assessment tool, which can be freely downloaded or used via a publicly available web server. This tool applies the CAPRI metrics to assess the quality of query structures against given target structures, along with other popular quality metrics such as DockQ, TM-score and *l*-DDT, and classifies the models according to the CAPRI model quality criteria. The tool can handle a variety of protein complex types including those involving peptides, nucleic acids, and oligosaccharides. The source code is freely available from https://gitlab.in2p3.fr/cmsb-public/CAPRI-Q and its web interface through the DOCKGROUND resource at https://dockground.compbio.ku.edu/assessment/.

INTRODUCTION

Protein interactions play a central role in most biological functions. Structural characterization of protein interactions is important for fundamental biology, as well as for prevention and treatment of diseases. Experimental techniques for protein structure determination have provided a wealth of data over the past half century, as evidenced by the rapidly growing Protein Data Bank (PDB)[1]. However, these are often time consuming and encounter difficulties on certain types of proteins [2, 3]. Experimentally determined protein structures are available for only a fraction of known proteins [4]. Due to those limitations, computational approaches are essential and complementary to the experimental methods for structural studies of proteins and their complexes. Methods for predicting the structure of protein complexes from structures or sequences of the individual components (protein docking or modeling of protein assemblies) have improved over many years through integration of sophisticated energy functions, better sampling of the energy landscapes, and incorporation of existing knowledge on protein structures and interactions, including deep learning [5, 6]. Many protein docking and assembly modeling tools are publicly available, including a number of fully automatic servers. The number of these tools is growing as a result of the continuous methodology developments by the structure prediction community.

Adequate assessment of the quality of docking/modeling tools is a complicated task. Over the past decade there has been a sharp increase in the number of docking tools, with > 60 available for commercial and academic use [7]. This trend has accelerated with the introduction of powerful deep-learning approaches [8, 9]. Benchmarking new tools is often accompanied by comparison to a limited number of alternative tools, tested

on a relatively small number of structures, typically randomly selected and lacking structural diversity [10]. Furthermore, since many docking programs are regularly updated, studies comparing different docking techniques quickly become obsolete, creating demand for new evaluations [11]. The authors of new docking tools also tend to provide their own evaluations, often achieving better results than the average user, due to a deeper understanding of their own algorithm and of the associated tunable parameters [10, 12]. They may also opt to focus on quality assessment approaches that are favorable to their tool. These reasons point to the need for a standard publicly available resource for the assessment of assembly modeling, which would offer objective assessment criteria agreed upon by the community.

The CAPRI experiment has been serving as the community-wide standard for docking and scoring evaluation since 2001. CAPRI satisfies the needs of the computational biology community by testing the performance of protein modeling procedures in blind prediction of three-dimensional structures of unpublished protein complexes. Prediction Rounds are launched on a rolling basis pending availability of suitable experimental protein complex structures. Participants register to predict the target structures of a Round. Each submitted model is assessed and ranked by the CAPRI assessors [9, 13, 14]. The results of each Round are published on the CAPRI web site as soon as the assessment is completed. However, several Rounds are typically grouped together into assessment periods, published in a CAPRI-dedicated journal issue. These journal issues describe current capabilities of the methods and define the state of the art in protein assembly modeling.

Here, to facilitate the development and comparison of protein docking/modeling techniques, we present *CAPRI-Q*, a tool implementing the prediction quality assessment used in the CAPRI experiment. *CAPRI-Q* is publicly available for download, as well as accessible through a web server. We show how the tool can be used to evaluate assembly models generated by different structure prediction servers including AlphaFold-multimer [15]. The tool performs automatic sequence alignment, and we show that the quality assessment of the model structures is not significantly affected by missing residues and mutations in the protein sequence, evidencing robustness of the model classification.

METHODS

Overview of the program

CAPRI-Q compares multiple models of a predicted protein-protein complex against a reference structure provided by the user (usually an experimentally determined structure of the complex). In addition to protein-protein complexes, CAPRI-Q handles complexes including peptides, nucleic acids, and oligosaccharides. Comparisons are performed on pairs of components, focusing on the binding interfaces between the interacting entities (two protein components or a protein component and another binding partner). For larger assemblies, distinct component pairs must be defined and evaluated independently, following the procedure laid out in Refs [6, 16].

The input files containing the coordinates of respectively, the models for the predicted structure and the reference structure (in PDB format), are filtered to remove hydrogen atoms, residues missing backbone atoms, and nonstandard residues.

Equivalent larger and smaller components in the reference and predicted structures are designated as “receptor” and “ligand,” respectively. Each protein chain of the models of the predicted complex is matched to the chains of the reference structure using the EMBOSS Needleman-Wunsch sequence alignment [17]. The residues contributing to the binding interface are defined, enabling the subsequent calculations of a set of parameters (metrics) evaluating the quality of the predicted models.

The main output file is a simple tab-delimited ASCII text file listing the ID of each model of the predicted complex and all the parameters for that model. Additional output files provide more detailed information, for instance on the sequence alignments of the predicted and reference complexes, and the native residue-residue contacts at the binding interface.

Superposition-dependent CAPRI metrics

For each model, CAPRI-Q evaluates the standard CAPRI quality metrics illustrated in Figure 1. A common metric is the root-mean-square deviation (RMSD) between the equivalent atoms in the optimally aligned model and the reference (native) structure. CAPRI-Q outputs several RMSD-based metrics, computed over specific subsets of atoms (Figure 1B): the *L-RMSD* metric, defined as the RMSD of the ligand atoms after optimally aligning the receptor moiety, and the *i-RMSD* parameter representing the RMSD computed on subsets of the native interface atoms (using a 10 Å interatomic distance cutoff), of *both* receptor and ligand (Figure 1B). Two distinct *i-RMSD* metrics are computed, one considering only the backbone atoms of the interface backbone atoms (*i-*

RMSDbb) and the other considering only the side-chain atoms of the interface residues (*i-RMSDsc*).

The *L-RMSD* may be misleading for large ligands which may align well at the native interface (and thus have small *i-RMSD* values) but display significant structural differences away from the interface. In CAPRI, the definition of the overall quality of the model accounts for this by relying on either *L-RMSD* or *i-RMSD*, depending on which score is better [13] as will be presented below.

To help rationalize the match between the predicted and reference structures, two additional parameters are computed, describing the overall shift of the ligand moiety relative to the receptor in models versus the reference structure. These are the misorientation angle θ_L of the ligand moieties and the residual displacement d_L of the ligand geometric center, computed after the receptors in the model and reference structures have been optimally superimposed [18] (Figure 1B), but these parameters are not considered in defining model quality.

In addition to the RMSD-based metrics described above, CAPRI-Q computes the TM-score [19], representing a superposition-dependent metric normalized by the protein length. The TM-score is calculated using the distances d_i between equivalent residue pairs in the two optimally aligned structures [19]. This score is computed using MM-align [20], specifically designed for the alignment of multi-chain structures (hereafter referred to as MM-score). The MM-score provides the simultaneous evaluation of the quality of the tertiary structure of the protein moieties and their quaternary arrangement. However, the absolute value of the MM-score should be interpreted considering the ratio between the numbers of residues in the two interacting proteins. For example, the MM-score

between two complexes featuring closely similar tertiary structures but different quaternary arrangement can be artificially inflated (higher values indicating a better match) if the receptor is significantly larger than the ligand (even in the ideal case in which all residues of the reference structure are present in the model structure and vice versa).

Superposition-independent CAPRI metrics

Metrics requiring structural superposition of two proteins are highly sensitive to how the superposition or structural alignment is performed, and to local structural distortions of the model or reference structure [21]. It is therefore useful to also consider quality metrics not based on structural superposition. CAPRI-Q computes three such metrics used by CAPRI to define model quality [9, 22]. These are the fraction of correctly predicted 'native' ligand-receptor contacts f_{nat} (with 'native' referring to the reference structure), and the fraction of incorrectly predicted ligand-receptor contacts (contacts not found in the reference structure) $f_{non\text{-}nat}$ (Figure 2). These contacts are computed for interface residues, defined as those having a heavy atom within 5 Å away from any heavy atom of the residues on the opposing side of the interface.

It should be noted that f_{nat} can be deceptively large for a model featuring many steric clashes between the receptor and the ligand due do a significant interpenetration of the two interacting components. Such obviously incorrect models are identified by computing the number of clashes between the ligand and the receptor. When this number exceeds a given limit, the clash threshold, the model is flagged by adding the word "clashes" to the model's CAPRI classification. By default, the clash threshold is set to the number of interface residues.

In addition to the interface residues contact fractions, CAPRI-Q also computes the local distance difference test, *l*-DDT [23], which evaluates the conservation of the local distances in the model compared to the reference structure. To calculate it, all the distances between heavy atoms in the reference structure within a sphere of 15 Å (excluding distances within the same residue) are computed (Figure S1). Then, the distances between all the matching atom pairs in the model structure are determined irrespective of their spatial separation. Finally, the corresponding distances in the reference and the model structures are compared and the fraction of conserved distances is calculated. A distance is considered conserved if it changes less than a given threshold. Following the original *l*-DDT paper [23], we use four different thresholds, 4 Å, 2 Å, 1 Å, and 0.5 Å. The reported *l*-DDT score is the average of the scores with these thresholds.

CAPRI overall model quality

In order to define the quality of the model overall, CAPRI classifies predicted structures onto four quality categories, based on the values of f_{nat}, *L-RMSD*, and *i-RMSDbb* [24]:

- ***(i)*** high quality, ($f_{nat} \geq 0.5$) and either (L-RMSD ≤ 1.0 Å) or (*i*-RMSDbb ≤ 1.0 Å);
- ***(ii)*** medium quality, ($f_{nat} \geq 0.3$) and either (L-RMSD ≤ 5 Å) or (*i*-RMSDbb ≤ 2.0 Å);
- ***(iii)*** acceptable quality, ($f_{nat} \geq 0.1$) and either (L-RMSD ≤ 10 Å) or (*i*-RMSDbb ≤ 4 Å); and
- ***(iv)*** incorrect, ($f_{nat} < 0.1$) or (L-RMSD > 10.0 Å) and (*i*-RMSDbb > 4.0 Å).

These criteria are applied in reverse order.

Model quality based on the DockQ score

CAPRI-Q also computes the DockQ score [25], a continuous quality score that integrates the main quality metrics of the standard CAPRI evaluation protocol [24]. DockQ is defined as the average of the f_{nat}, and scaled L-RMSD and *i*-RMSDbb metrics:

$$\mathrm{DockQ} = \frac{1}{3}\left(f_{nat} + \frac{1}{1+\left(\frac{\text{L-RMSD}}{d_1}\right)^2} + \frac{1}{1+\left(\frac{i\text{-RMSDbb}}{d_2}\right)^2} \right),$$

where d_1 and d_2 are scaling parameters which determine how fast large *L-RMSD* and *i-RMSDbb* values, respectively, can be scaled to zero [25]. The values of these parameters (d_1 = 8.5 Å and d_2 = 1.5 Å) were obtained [25] by maximizing classification performance on the MOAL-set [26]. The range of the DockQ score is 0 to 1. Generally, models with DockQ < 0.23 are considered incorrect. High, medium and acceptable accuracy models tend to have DockQ ≥ 0.80, 0.49 ≤ DockQ < 0.80, and 0.23 ≤ DockQ < 0.49, respectively [25]. It is important to keep in mind that DockQ considers both *L-RMSD* and *i-RMSDbb* simultaneously, whereas CAPRI classification assignments are based on either L-RMSD or i-RMSDbb. Therefore, the model quality categories defined using DockQ may not exactly correspond to those obtained by the CAPRI protocol.

Additional information and metrics

For the convenience of users, CAPRI-Q generates a “Summary” file containing supplementary information used in calculations of the above metrics. For the receptor and the ligand of the models and the reference structure, the Summary lists residues contributing to the binding interface of the reference structure, mapping of the chains

between the model and the reference structures, solvent accessible surface area (SASA), and sequence identities (percentage of identical residues), seqID. The seqID is also used to flag models generated with sequences significantly deviating (in length or composition) from those in the reference structures (the default threshold used requires 70% sequence overlap with the target). These models marked as having “low_id”, and models flagged with “clashes”, are evaluated here, but not considered in the standard CAPRI evaluation.

Additional metrics reported in the Summary are the interface area (*IA*), fractions of ‘native’ (correctly predicted) (F_{IR}) and ‘non-native’ (incorrectly predicted, or overpredicted) (F_{OP}) interface residues, the RMSD of individual components (monomers) (m-*RMSD*), and the DockQ score.

IA is calculated as

$$IA = \mathrm{SASA}(R) + \mathrm{SASA}(L) - \mathrm{SASA}(RL),$$

where $\mathrm{SASA}(R)$, $\mathrm{SASA}(L)$ and $\mathrm{SASA}(RL)$ are solvent accessible surface areas of the receptor-only, ligand-only and the receptor-ligand complex respectively. SASA is calculated using the C library from the FreeSASA software [27]. The interface residues used to compute F_{IR} and F_{OP} are defined as those losing SASA upon binding. F_{IR} and F_{OP} are computed separately for the receptor and the ligand of the model structure. *m-RMSD* is likewise evaluated separately for the ligand and receptor, based on C^{α} RMSD between the model and the reference.

CASE STUDIES

Comparing different protein docking methodologies

CAPRI-Q functionality is illustrated on models produced by docking calculations performed for three dimers from the DOCKGROUND Docking Benchmark 1 (1B6C, 1GPW, 1K93) [28] and the trimeric CAPRI target T50 (CAPRI round 24 from February-March 2011, PDB code 3R2X) [29]. The PDB codes refer to the experimentally determined structures of the complexes (*i.e.*, bound structures), whereas in docking we used the experimentally determined structures of the unbound proteins: 1BKF (0.6 Å) and 1VJY (1.5 Å) for 1B6C; 1thf (3.6 Å) and 1K9V (0.7 Å) for 1GPW; 1K8T (10.0 Å) and 1CLL (9.0 Å) for 1K93; and 3GBN (0.6 Å) and 1U84 (0.6 Å) for 3R2X (the numbers in parentheses are the RMSD between the unbound and the bound structures). According to these values, 1B6C and 3R2X are easy (rigid-body) targets while 1GPW and 1K93 are medium difficulty and hard targets, respectively [30].

The calculations were performed using the HADDOCK [31] and HDOCK [32] docking servers and a local install of AlphaFold-Multimer [15] version 2.3 with default parameters and databases. HADDOCK performs information-driven docking, clusters the models and reports the top four models for each cluster. The docking was performed using the "*Easy"* interface option of the server, with the interface residues specified. HDOCK performs hybrid free and template-based docking using a distance-dependent knowledge-based scoring function [33]. We also used its option to specify interface residues. All generated HADDOCK and AlphaFold-Multimer models, and the top 50 HDOCK models were evaluated by CAPRI-Q. The output of CAPRI-Q for 1GPW is summarized in Table 1. The best scoring model for each complex for each of the three modeling servers is shown in Supplementary Figure S2.

Effect of missing residues

The CAPRI-Q model quality classification is robust, even when some residues are missing from the model structure. The effect of missing residues based on medium-quality models generated by AlphaFold-Multimer for the 1B6C target is shown in Supplementary Table S1. For this case study, we removed from the initial model 5, 10, 15, or 30 randomly selected ligand residues from regions outside the interface in the model structure. We also removed from the initial model the same numbers of residues randomly selected from the interface. For the models with non-interface residues removed, none of the metrics other than the sequence identity were significantly affected. For the models with removed interface residues, the fraction of native contacts slightly decreased and the interface RMSDs slightly increased. However, these changes marginally affected the overall model quality classification. For the purpose of ranking models in CAPRI, a sequence alignment step ensures that all RMSD calculations are performed on the same number of residues. However, in our example, each model was evaluated separately and RMSD calculations were therefore not necessarily performed over the same amount of residues. Nonetheless, this did not affect the overall classification of the models.

Effect of mutations on model assessment

For this case study, we produced three types of mutated structures by replacing either 5 or 20 randomly selected ligand residues in the native structure of the 3R2X target to alanine, aspartic acid, or proline. Mutated residues were all picked either outside the interface or within the interface of the ligand. Structural models of the mutated proteins were generated using the NEST program from the JACKAL suite [34]. In this case, NEST mostly performed side chain re-packing, with the backbone almost unaffected (the backbone atoms may only be slightly shifted during structure optimization to remove steric

clashes). This becomes apparent when residues that form helices mutate into helix-breaking prolines, since the helices in the 3R2X ligand remained intact after the residue replacements. Table S2 lists the CAPRI-Q metrics of the original and the mutated structures. The native sequence of the 3R2X PDB structure was also run through NEST, resulting in a slight displacement of the original atoms, leading to poorer values of the metrics for the original structure (Table S2).

For the ALA mutations, increasing the number of mutations outside the interface from 5 to 20 had little impact on the metric values. However, increasing the number of such mutations at the interface increased the value of i-RMSDbb and decreased that of f_{nat}, albeit still by a relatively small amount. Nevertheless, none of these changes were sufficient to alter the classification of the model. For the mutations outside the interface, the type of mutation did not significantly affect any metric. For the interface residues, mutations to ALA had the least impact, whereas mutations to PRO had the most impact. Still, none of the metric values changed enough to affect the model classification. Indeed, as long as the backbone of the model is maintained, mutations do not greatly affect the assessment of the predicted models.

AVAILABILITY OF THE TOOL

Source code

The source code can be freely downloaded at https://gitlab.in2p3.fr/cmsb-public/CAPRI-Q. CAPRI-Q is written in C, except for the included source code for MM-score calculations which is written in C++. The FreeSASA library is included. However, the EMBOSS library needs to be downloaded and installed by the user. It is simple to compile and is called

from within the CAPRI-Q program. Downloading the source code, as opposed to using the web server, enables users to run many models against a reference structure, and gives more control over the scoring parameters, such as the number of acceptable clashes.

Web interface

CAPRI-Q is implemented as part of the DOCKGROUND web server at https://dockground.compbio.ku.edu/assessment/. Users can submit up to ten models per job to evaluate against a reference structure. The submission page is shown in Figure S3A. Users are emailed a link to the results once their job has completed. The example results page is shown in Figure S3B. The results can be downloaded as a text file in a tab-delimited format.

CONCLUDING REMARKS

CAPRI-Q is a stand-alone tool implementing the community-established standard for the quality assessment of protein assembly models. It is an enhanced version of the standard model evaluation procedure in CAPRI, a long-standing community-wide blind prediction experiment, as it also incorporates three complementary metrics: DockQ, a continuous score combining the CAPRI model quality metrics, and the MM-score and *l-DDT*, two commonly used model quality metrics for evaluating predicted protein structures [35]. CAPRI-Q is robust with respect to imperfections of the submitted model structures. The tool can be used either off-line as a command-line standalone program or on-line through the Web interface. The source code is freely available for download at

https://gitlab.in2p3.fr/cmsb-public/CAPRI-Q and the web interface is part of the DOCKGROUND resource at https://dockground.compbio.ku.edu/assessment/.

ACKNOWLEDGEMENTS

This study was supported by NIH grant R01GM074255 and NSF grant DBI 2224122.

REFERENCES

[1] Bittrich S, Bhikadiya C, Bi C, Chao H, Duarte JM, Dutta S, et al. RCSB Protein Data Bank: Efficient Searching and Simultaneous Access to One Million Computed Structure Models Alongside the PDB Structures Enabled by Architectural Advances. J Mol Biol. 2023;435:167994.
[2] Dokholyan NV. Experimentally-driven protein structure modeling. J Proteomics. 2020;220:103777.
[3] Guzenko D, Lafita A, Monastyrskyy B, Kryshtafovych A, Duarte JM. Assessment of protein assembly prediction in CASP13. Proteins. 2019;87:1190-9.
[4] Aloy P, Pichaud M, Russell RB. Protein complexes: structure prediction challenges for the 21st century. Curr Opin Struct Biol. 2005;15:15-22.
[5] Vakser IA. Protein-protein docking: from interaction to interactome. Biophys J. 2014;107:1785-93.
[6] Lensink MF, Brysbaert G, Raouraoua N, Bates PA, Giulini M, Honorato RV, et al. Impact of AlphaFold on structure prediction of protein complexes: The CASP15-CAPRI experiment. Proteins. 2023;91:1658-83.
[7] Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. Biophys Rev. 2017;9:91-102.
[8] Versini R, Sritharan S, Aykac Fas B, Tubiana T, Aimeur SZ, Henri J, et al. A Perspective on the Prospective Use of AI in Protein Structure Prediction. J Chem Inf Model. 2023.
[9] Wodak SJ, Vajda S, Lensink MF, Kozakov D, Bates PA. Critical Assessment of Methods for Predicting the 3D Structure of Proteins and Protein Complexes. Annu Rev Biophys. 2023;52:183-206.
[10] Plewczynski D, Lazniewski M, Augustyniak R, Ginalski K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. J Comput Chem. 2011;32:742-55.
[11] Wang Z, Sun H, Yao X, Li D, Xu L, Li Y, et al. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. Phys Chem Chem Phys. 2016;18:12964-75.
[12] Chen H, Lyne PD, Giordanetto F, Lovell T, Li J. On evaluating molecular-docking methods for pose prediction and enrichment factors. J Chem Inf Model. 2006;46:401-15.
[13] Lensink MF, Nadzirin N, Velankar S, Wodak SJ. Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition. Proteins. 2020;88:916-38.
[14] Janin J, Henrick K, Moult J, Eyck LT, Sternberg MJ, Vajda S, et al. CAPRI: a Critical Assessment of PRedicted Interactions. Proteins. 2003;52:2-9.
[15] Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2022:2021.10.04.463034.
[16] Lensink MF, Brysbaert G, Mauri T, Nadzirin N, Velankar S, Chaleil RAG, et al. Prediction of protein assemblies, the next frontier: The CASP14-CAPRI experiment. Proteins. 2021;89:1800-23.
[17] Madeira F, Pearce M, Tivey ARN, Basutkar P, Lee J, Edbali O, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res. 2022;50:W276-W9.
[18] McLachlan AD. Rapid comparison of protein structures. Acta Cryst. 1982;A38:871-3.
[19] Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins. 2004;57:702-10.
[20] Mukherjee S, Zhang Y. MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. Nucleic Acids Res. 2009;37:e83.
[21] Kufareva I, Abagyan R. Methods of protein structure comparison. Methods Mol Biol. 2012;857:231-57.
[22] Mendez R, Leplae R, Lensink MF, Wodak SJ. Assessment of CAPRI predictions in rounds 3-5 shows progress in docking procedures. Proteins. 2005;60:150-69.

[23] Mariani V, Biasini M, Barbato A, Schwede T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics. 2013;29:2722-8.
[24] Lensink MF, Mendez R, Wodak SJ. Docking and scoring protein complexes: CAPRI 3rd Edition. Proteins. 2007;69:704-18.
[25] Basu S, Wallner B. DockQ: A Quality Measure for Protein-Protein Docking Models. PLoS One. 2016;11:e0161879.
[26] Moal IH, Torchala M, Bates PA, Fernandez-Recio J. The scoring of poses in protein-protein docking: current capabilities and future directions. BMC Bioinformatics. 2013;14:286.
[27] Mitternacht S. FreeSASA: An open source C library for solvent accessible surface area calculations. F1000Res. 2016;5:189.
[28] Gao Y, Douguet D, Tovchigrechko A, Vakser IA. DOCKGROUND system of databases for protein recognition studies: unbound structures for docking. Proteins. 2007;69:845-51.
[29] Lensink MF, Wodak SJ. Docking, scoring, and affinity prediction in CAPRI. Proteins. 2013;81:2082-95.
[30] Mintseris J, Wiehe K, Pierce B, Anderson R, Chen R, Janin J, et al. Protein-Protein Docking Benchmark 2.0: an update. Proteins. 2005;60:214-6.
[31] van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016;428:720-5.
[32] Yan Y, Tao H, He J, Huang SY. The HDOCK server for integrated protein-protein docking. Nat Protoc. 2020;15:1829-52.
[33] Huang SY, Zou X. An iterative knowledge-based scoring function for protein-protein recognition. Proteins. 2008;72:557-79.
[34] Xiang Z, Honig B. Extending the accuracy limits of prediction for side-chain conformations. J Mol Biol. 2001;311:421-30.
[35] Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)-Round XV. Proteins. 2023;91:1539-49.

TABLES

Table 1. CAPRI-Q sample output for 1GPW (see text).

Three models with a range of quality classifications from each modeling method are shown in Table 1. The column for sequence identity was removed to reduce the size of the table (all were above 90%). From the Table it can be deduced that the CAPRI classification expectedly correlates with the metrics, as it improves with a higher fnat, and lower L-RMSD or i-RMSDbb scores. Interestingly, the AlphaFold model with the highest fnat value is of medium rather than high quality, due to increased L-RMSD and i- RMSDbb values.

MODELING TOOL	NUMBER OF CLASHES	F_{NAT}	$F_{NON-NAT}$	TM-SCORE	LDDT	LRMSD	IRMSDBB	IRMSDSC	DOCKQ	DISTANCE	THETA	CLASSIFICATION
HADDOCK	8	0.12	0.89	0.72	0.80	16.86	7.06	7.64	0.12	10.76	57.44	incorrect
HADDOCK	21	0.75	0.61	0.90	0.81	5.55	1.93	2.46	0.61	4.30	16.12	medium
HADDOCK	26	0.71	0.60	0.90	0.82	5.35	1.95	2.44	0.60	4.01	16.33	medium
HDOCK	32	0.93	0.32	0.97	0.85	0.88	0.67	1.62	0.92	0.34	0.34	high
HDOCK	29	0.38	0.61	0.86	0.82	6.37	2.67	3.02	0.42	3.39	27.23	acceptable
HDOCK	28	0.07	0.94	0.67	0.80	13.60	7.32	7.90	0.13	2.55	65.35	incorrect
ALPHAFOLD	17	0.90	0.31	0.98	0.89	2.68	1.04	1.66	0.83	2.34	6.23	medium
ALPHAFOLD	18	0.87	0.30	0.98	0.89	2.43	0.99	1.63	0.83	2.04	5.46	high
ALPHAFOLD	15	0.88	0.27	0.97	0.89	1.51	0.78	1.55	0.88	1.25	3.07	high

FIGURES

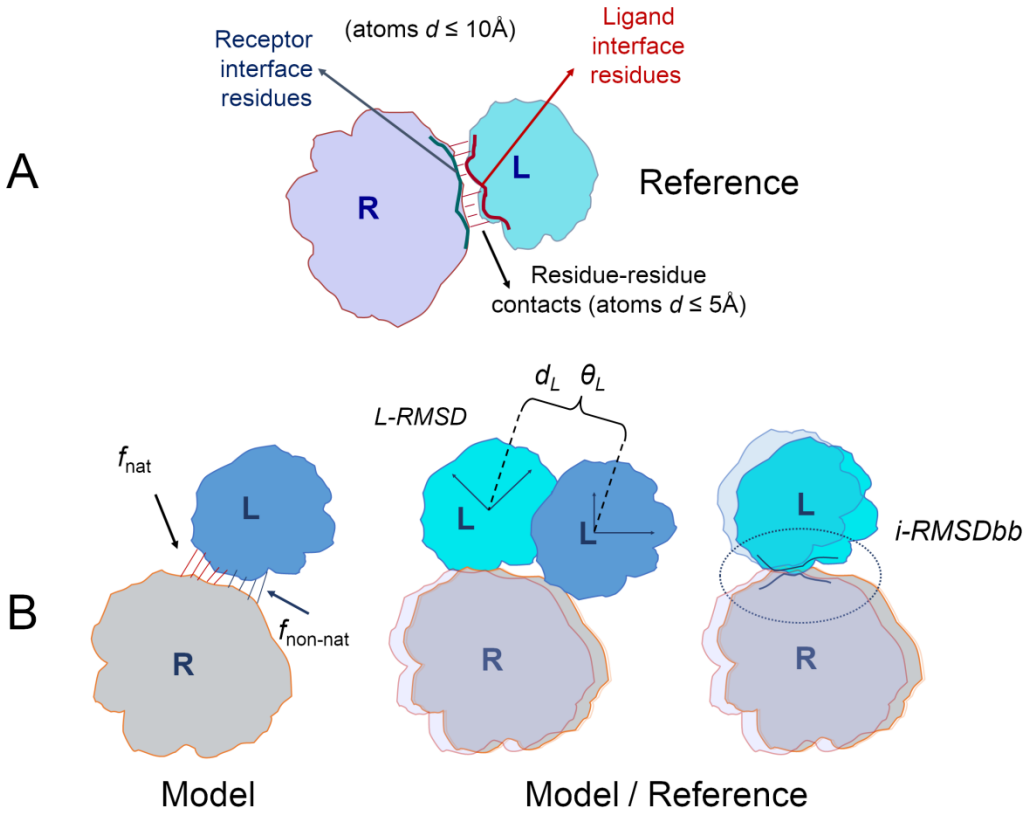


Figure 1. *Schematic illustration of the CAPRI model quality metrics.* (A) Shows the quantities evaluated for the reference structure of a protein complex consisting of two protein components. These are: the residue-residue contacts between the receptor (R) and the ligand (L) (see text for definition), computed using a distance threshold of d ≤ 5 Å between atoms; and the residues of the ligand and receptor, respectively, contributing to the binding interface, using a more lenient threshold of d ≤ 10 Å for the interacting atoms between both subunits. (B) Illustrates the model quality metrics computed for each predicted model of the complex. These are, from left to right, the fractions native (f_{nat}) and non-native ($f_{non\text{-}nat}$) residue-residue contacts in the predicted interface (see text for definition); the root mean square displacement (RMSD) of the backbone atoms of the ligand (*L-RMSD*), the mis-orientation angle θ_L and the residual displacement d_L of the ligand center of mass, after the receptor in the model and experimental structures were optimally superimposed; the RMSD of the backbone atoms of all interface residues (*i-RMSDbb*) after only these residue have been optimally superimposed. Additionally, we also compute the RMSD of interface sidechain atoms (*i-RMSDsc*) (not shown).

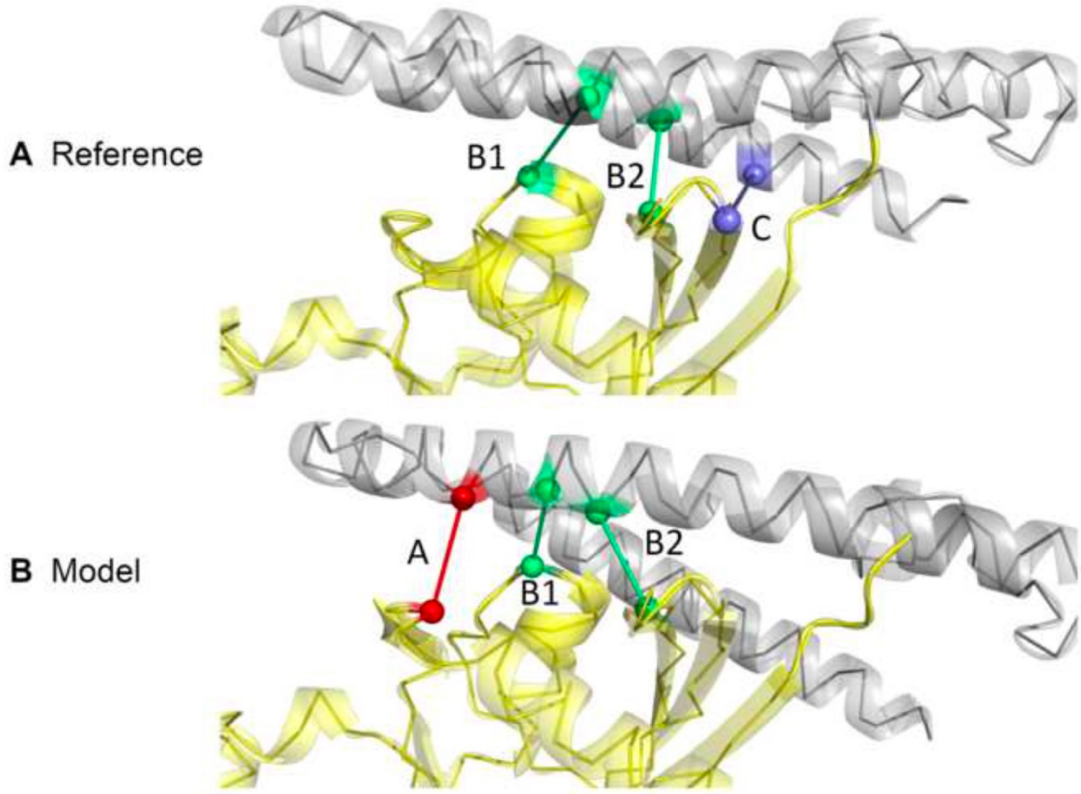


Figure 2. *An illustration of different types of contacts in* f_{nat} *and* $f_{non\text{-}nat}$ parameters. The reference structure is shown on the top and the model is on the bottom. The receptor is in yellow, and the ligand is in gray. A is an incorrect contact in the model that does not exist in the reference structure, which increases the $f_{non\text{-}nat}$ parameter. B1 and B2 are contacts in both the reference and the model structures, which increase f_{nat}. C is a contact in the native structure not present in the model structure, which would decrease f_{nat}.

Supplementary Materials

CAPRI-Q: A CAPRI assessment of docking quality server

Keeley W. Collins,[1] Matthew M. Copeland,[1] Guillaume Brysbaert,[3] Shoshana J. Wodak,[4] Alexandre M.J.J. Bonvin,[5] Petras J. Kundrotas,[1]* Ilya A. Vakser,[1,2]* and Marc F. Lensink[3]*

[1]Center for Computational Biology, and [2]Department of Molecular Biology, The University of Kansas, Lawrence, KS 66045, USA

[3]Univ. Lille, CNRS, UMR 8576 - UGSF - Unité de Glycobiologie Structurale et Fonctionnelle, F-59000 Lille, France

[4]VIB-VUB Center for Structural Biology, Brussels, Belgium

[5]Bijvoet Centre for Biomolecular Research, Faculty of Science - Chemistry, Utrecht University, The Netherlands

*Corresponding authors:

Petras Kundrotas, University of Kansas, email: pkundro@ku.edu,

Ilya Vakser, University of Kansas, e-mail: vakser@ku.edu,

Marc Lensink, University of Lille, e-mail: marc.lensink@univ-lille.fr.

TABLES

Table S1. CAPR-Q scores for one of the medium-accuracy models produced by AlphaFold Multimer for the 1b6c target with certain amounts of ligand residues removed.

NUMBER OF MISSING RESIDUES	NUMBER OF CLASHES	F_{NAT}	$F_{NON\text{-}NAT}$	TM-SCORE	LDDT	LRMSD	IRMSDBB	IRMSDSC	DOCKQ	DISTANCE	THETA	SEQ-ID	CLASSIFICATION
0	54	0.64	0.45	0.89	0.83	4.21	2.67	3.21	0.56	2.04	19.6	100.0	medium
5	54	0.64	0.45	0.89	0.83	4.21	2.67	3.21	0.56	2.04	19.6	90.65	medium
10	54	0.64	0.45	0.89	0.83	4.21	2.67	3.21	0.56	2.04	19.6	85.05	medium
15	54	0.64	0.45	0.89	0.82	4.21	2.67	3.21	0.56	2.04	19.6	85.05	medium
30	54	0.64	0.48	0.87	0.83	3.98	2.66	3.24	0.56	1.65	19.9	64.50	low_id, medium
5*	54	0.64	0.48	0.89	0.83	4.21	2.67	3.21	0.56	2.04	19.6	94.05	medium
10*	28	0.62	0.46	0.89	0.83	4.21	2.67	3.21	0.55	2.04	19.6	89.70	medium
15*	28	0.58	0.46	0.89	0.83	4.24	2.68	3.21	0.55	2.07	19.6	84.10	medium
30*	24	0.59	0.47	0.86	0.83	4.68	3.34	3.24	0.54	1.77	19.2	69.2	medium

* Residues were removed from the interface region.

From the Table it can be deduced that removing ligand residues has only a minimal effect on metric values and does not affect the overall classification of the model.

Table S2. CAPRI-Q scores for the native structure of the 3r2x target with certain amounts of mutated interface and non-interface ligand (chain C in the PDB structure) residues

NUMBER AND TYPE OF MUTATIONS	NUMBER OF CLASHES	F_{NAT}	$F_{NON-NAT}$	TM-SCORE	LDDT	LRMSD	IRMSDBB	IRMSDSC	DOCKQ	DISTANCE	THETA	SEQ-ID	CLASSIFICATION
0	3	1.00	0.02	0.98	0.99	0.22	0.04	0.29	0.99	0.03	0.30	98.90	high
5, ALA	3	1.00	0.02	0.98	0.99	0.05	0.04	0.29	0.99	0.00	0.00	97.50	high
20, ALA	3	1.00	0.02	0.98	0.99	0.12	0.04	0.42	0.99	0.01	0.13	79.00	high
20, ASP	3	1.00	0.02	0.98	0.99	0.10	0.04	0.13	0.99	0.02	0.09	78.00	high
20, PRO	6	1.00	0.04	0.98	0.99	0.11	0.04	0.28	0.99	0.01	0.13	77.80	high
5*, ALA	3	1.00	0.02	0.98	0.99	0.22	0.04	0.29	0.99	0.03	0.30	92.70	high
20*, ALA	1	0.67	0.09	0.98	0.99	0.61	0.41	0.41	0.86	0.09	0.26	75.60	high
20*, ASP	11	0.66	0.17	0.98	0.99	0.64	0.50	0.59	0.85	0.09	0.24	75.60	high
20*, PRO	13	0.75	0.16	0.98	0.99	0.82	0.66	0.96	0.83	0.17	0.76	74.40	high

* Residues were mutated from the interface region.

From the Table it can be deduced that neither an increase in the amount of mutated residues nor the type of mutation affects the overall classification of the model. However, the type of mutation does have an impact albeit small on the metric values. It should be noted that the mutation resulted mostly in side chain repacking, with only minimal change upon backbone conformation.

FIGURES

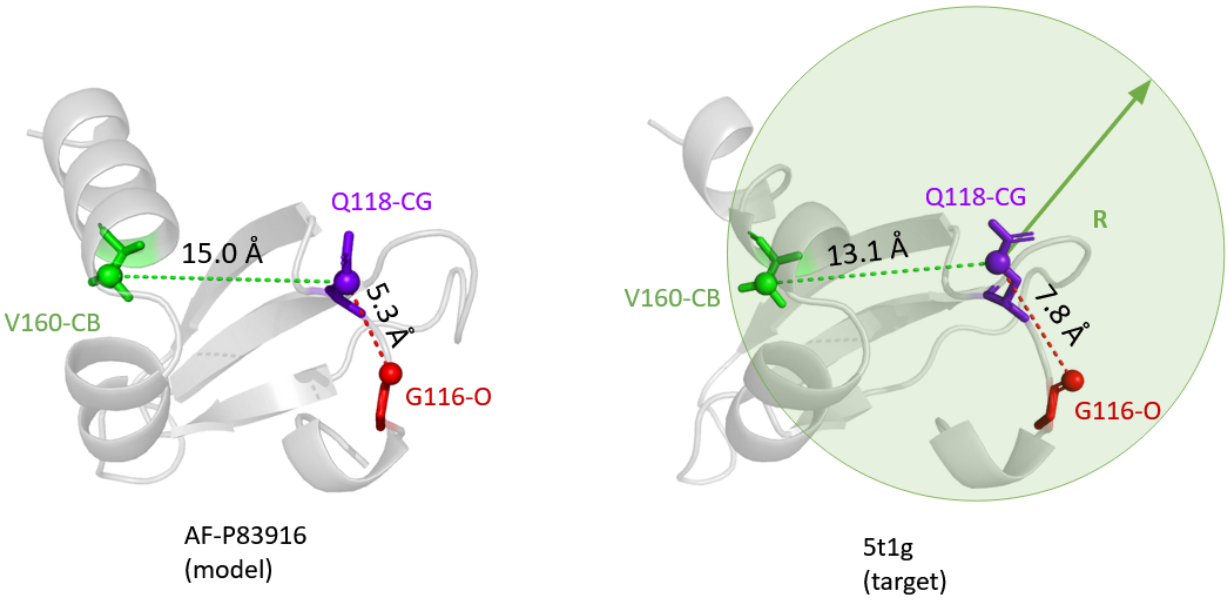


Figure S1. *Examples of conserved and non-conserved distances used in l-DDT calculations.* X-ray structure 5t1g was used as the reference structure (right) and the model AF-P83916 from the AlphaFold database as the model structure (left). Distances are calculated between the CG atom of residue Q118 (purple) in the reference structure and all atoms within R=15 Å sphere (light green circle), which are not within the same residue. These distances are compared to the corresponding distances between equivalent atoms in the model structure. The figure shows a non-conserved distance (at 2 Å threshold) between atoms CG of Q118 and O of G116 residue (red) with 2.5 Å difference and a conserved distance between atoms CG of Q118 and CB of V160 residue (green) with 1.9 Å difference.

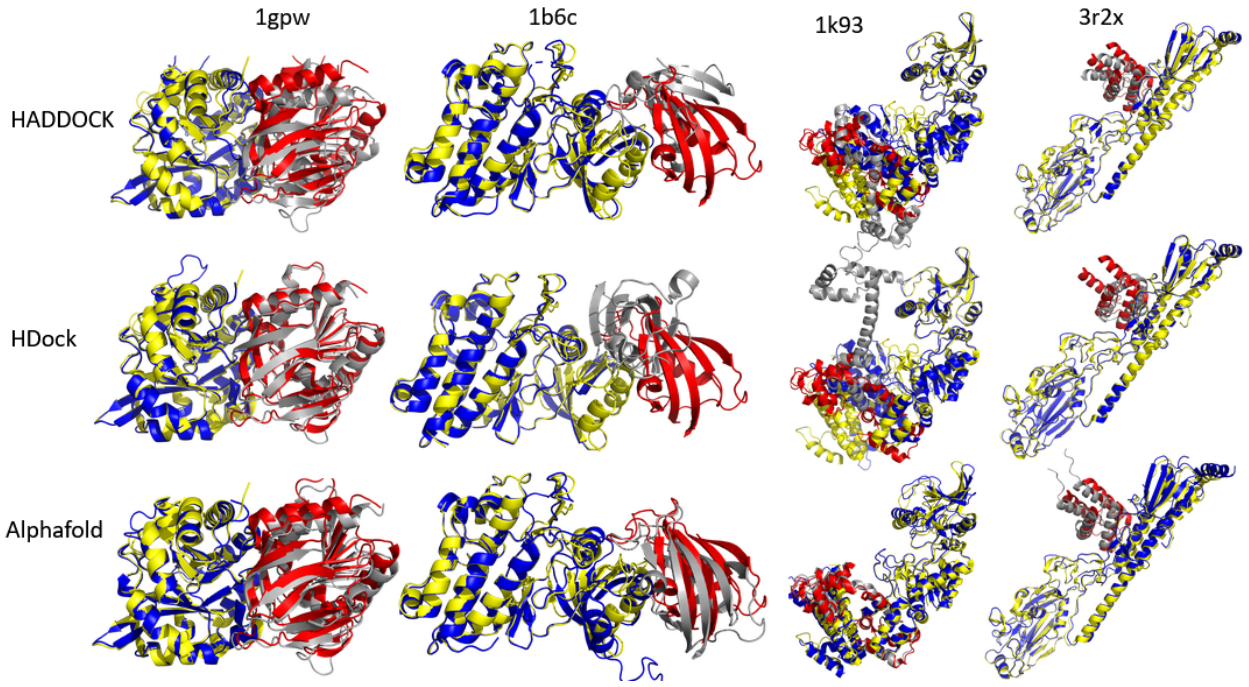
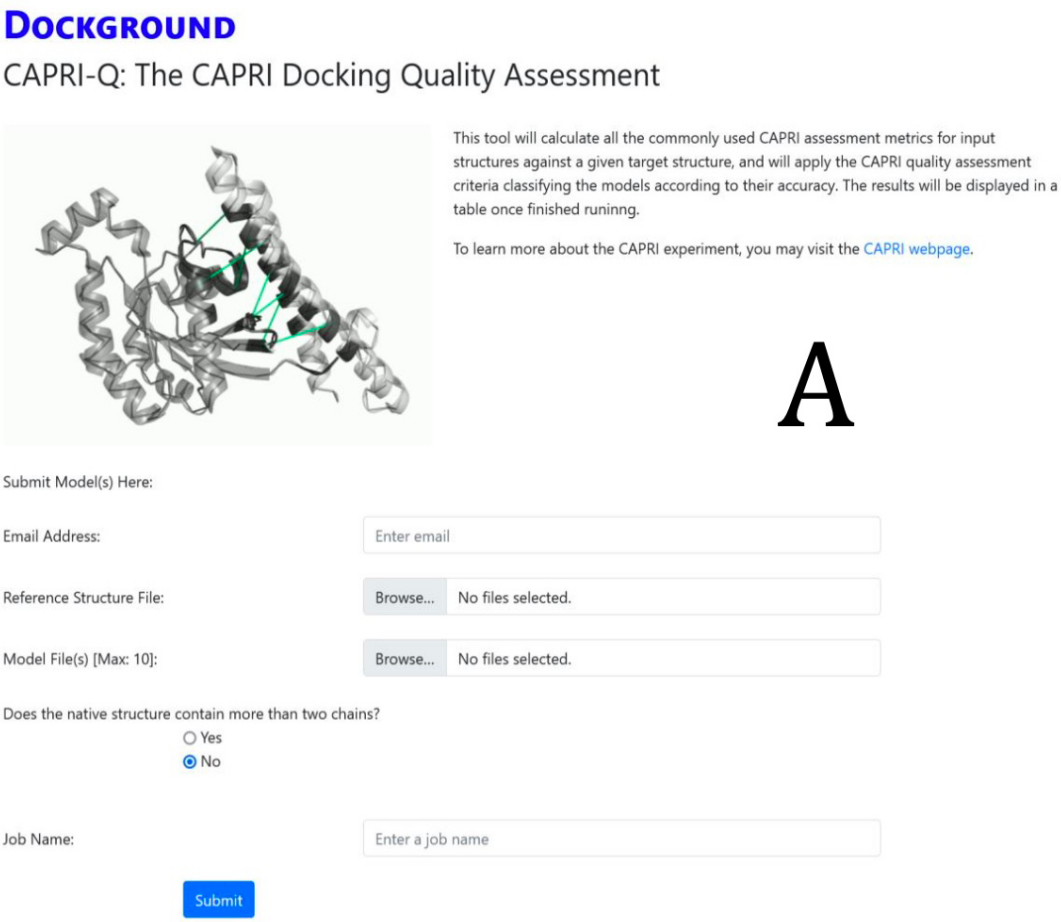


Figure S2. *Best models produced by three methodologies (denoted at the left of the Figure) for four exemplar complexes aligned to corresponding native structures (PDB codes of which are shown on the top of the Figure)*. The receptors (ligands) of the native and model structures are shown as yellow (red) and blue (gray) cartoons, respectively.



model	nclash	fnat	fnonnat	tm-score	lddt	lrmsd	irmsdbb	irmsdsc	distance	theta	seqid	dockq	classification
cluster2_2.pdb	13/64	0.1471	0.8667	0.6977	0.7926	15.5325	7.3758	8.0693	6.1807	64.8583	99.20	0.1391	incorrect
cluster2_4.pdb	8/64	0.1176	0.8919	0.7171	0.7909	16.8556	7.0603	7.6406	10.7667	57.4456	99.20	0.1212	incorrect
cluster3_1.pdb	21/64	0.7500	0.6107	0.9048	0.8103	5.5465	1.9275	2.4569	4.2957	16.1156	99.20	0.6095	medium
cluster3_2.pdb	26/64	0.7059	0.6033	0.9018	0.8184	5.3495	1.9542	2.4410	4.0054	16.3316	99.20	0.5976	medium
cluster3_4.pdb	26/64	0.6912	0.6270	0.9041	0.8184	5.3638	1.9627	2.4541	4.1689	15.7484	99.20	0.5917	medium
cluster4_1.pdb	15/64	0.3235	0.7800	0.8558	0.7988	7.3552	3.1330	3.6799	6.0994	20.4300	99.20	0.3606	acceptable
cluster4_2.pdb	15/64	0.2647	0.8085	0.8211	0.7957	8.6316	3.7753	4.1409	6.9637	25.1936	99.20	0.2978	acceptable
cluster4_3.pdb	12/64	0.2500	0.8247	0.8485	0.8043	7.6148	3.2620	3.8103	6.1087	22.6794	99.20	0.3264	acceptable
cluster4_4.pdb	16/64	0.2941	0.8020	0.8464	0.8037	7.5161	3.2941	3.7847	5.8424	23.3965	99.20	0.3424	acceptable
cluster5_2.pdb	25/64	0.6029	0.5393	0.8663	0.8213	6.0320	2.8377	3.3214	3.2199	21.3318	99.20	0.4955	acceptable

Figure S3. *The submission page and results page for the CAPRI-Q Assessment web interface. Part A shows the submission page and part B shows the results table.*