New prediction categories in CASP15

Andriy Kryshtafovych[1], Maciej Antczak[2], Marta Szachniuk[2], Rachael C. Kretsch[3], Ramya Rangan[3], Phillip Pham[4], Rhiju Das[4], Xavier Robin[5], Gabriel Studer[5], Janani Durairaj[5], Jerome Eberhardt[5], Aaron Sweeney[6], Maya Topf[6], Torsten Schwede[5], Krzysztof Fidelis[1], and John Moult[7]

[1]University of California Davis Genome Center
[2]Politechnika Poznanska Instytut Informatyki
[3]Stanford University School of Medicine
[4]Stanford University Department of Biochemistry
[5]Universitat Basel Department Biozentrum
[6]Leibniz-Institut fur Virologie
[7]Institute for Bioscience and Biotechnology Research

May 10, 2023

Abstract

Prediction categories in the Critical Assessment of Structure Prediction (CASP) experiments change with the need to address specific problems in structure modeling. In CASP15, four new prediction categories were introduced: RNA structure, ligand-protein complexes, accuracy of oligomeric structures and their interfaces, and ensembles of alternative conformations. This paper lists technical specifications for these categories and describes their integration in the CASP data management system.

New prediction categories in CASP15

Andriy Kryshtafovych [1*], Maciej Antczak[2,3], Marta Szachniuk [2,3],

Rachael C. Kretsch [4], Ramya Rangan[4], Phillip Pham [5], Rhiju Das[5,6],

Xavier Robin [7,8], Gabriel Studer[7,8], Janani Durairaj [7,8], Jerome Eberhardt [7,8],

Aaron Sweeney [9], Maya Topf [9,10],

Torsten Schwede [7,8], Krzysztof Fidelis[1] and John Moult [11]

1 - Genome Center, University of California, Davis, CA 95616, USA

2 - Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland

3 - Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

4 - Biophysics Program, Stanford University School of Medicine, CA USA

5 - Biochemistry Department, Stanford University School of Medicine, CA USA

6 - Howard Hughes Medical Institute, Stanford University, CA USA

7 - Biozentrum, University of Basel, Basel 4056, Switzerland,

8 - SIB Swiss Institute of Bioinformatics, Basel 4056, Switzerland

9 - Centre for Structural Systems Biology (CSSB), Leibniz-Institut für Virologie (LIV), Hamburg, Germany

10 - Universitätsklinikum Hamburg Eppendorf (UKE), Hamburg, Germany

11 - Institute for Bioscience and Biotechnology Research, Department of Cell Biology and Molecular genetics, University of Maryland, 9600 Gudelsky Drive, Rockville, MD 20850, USA

Corresponding author: Andriy Kryshtafovych (*akryshtafovych@ucdavis.edu*)

Abstract

Prediction categories in the Critical Assessment of Structure Prediction (CASP) experiments change with the need to address specific problems in structure modeling. In CASP15, four new prediction categories were introduced: RNA structure, ligand-protein complexes, accuracy of oligomeric structures and their interfaces, and ensembles of alternative conformations. This paper lists technical specifications for these categories and describes their integration in the CASP data management system.

Acknowledgements

1 | INTRODUCTION

CASP is a well-established experiment for tracking progress in computational methods for calculating protein structure from amino acid sequence and related challenges[1-4]. While the core of CASP has remained unchanged since its inception, the scope changes frequently, with some categories becoming obsolete while new ones are introduced. Following the major advances in single protein structure modeling seen at CASP14, three categories were discontinued for round 15: contact prediction, refinement, and third-party model accuracy estimation for monomeric targets. At the same time, discussions among the organizers and advisory board members as well as a survey of the modeling community established interest in the new categories of RNA structure, protein-ligand complexes, accuracy of oligomeric structures and their interfaces, and ensembles of alternative conformations. CASP launched pilot experiments in all these areas, exploring, where possible, collaborations with groups involved in the existing or recently discontinued related experiments. In CASP15, the new categories have yielded interesting results, and their assessment is discussed in dedicated papers in this journal special issue[5-8]. The main goal of this paper is to address technical issues of introducing the new categories in CASP and serve as a point of reference for format specifications and details on how the experiment was conducted.

2 | METHODS

Introducing a new category into CASP requires planning data workflows, designing formats and technical parameters for new types of models, and incorporating those into the existing CASP infrastructure. Subsections 2.1-2.4 below describe the implementation details for four new CASP15 categories.

2.1 | RNA structure prediction. Prediction of RNA structure from nucleic sequence is a challenging task as RNA molecules, like proteins, can fold into a wide variety of 3D shapes. Several research groups have been actively working in this area, and in 2010 Eric Westhof pioneered a CASP-like RNA-Puzzles

challenge[9] to track the state of the art in the RNA structure prediction and provide a forum for discussing methodological advances. Over the course of 12 years (2010-2021) there were 22 evaluated RNA-Puzzles challenges, which attracted the attention of around 10 returning participants[10]. In 2022, on the initiative of Rhiju Das, Eric Westhof and CASP organizers, the RNA-Puzzles joined forces with CASP, and RNA structure prediction became a prediction category in CASP15. This helped expand the target and predictors base of the RNA-modeling experiment (12 targets, 25 research groups in CASP15), stimulate development of new RNA prediction methods through the exchange of ideas and techniques with the protein prediction community, where deep learning methods recently made a significant impact on the modeling accuracy[11,12], increase visibility of the field, and use CASP's standardized platform for managing predictions and evaluating and comparing different prediction methods.

To incorporate RNA prediction into CASP, we adhered as closely as possible to the requirements and recommendations of the RNA-Puzzles experiments [9].

2.1.1 | RNA prediction format(*https://predictioncenter.org/casp15/index.cgi?page=format#TS*). Similarly to protein structure prediction, a CASP RNA submission file starts with the CASP header including format specification code, target identifier, author identifier, and description of methods used for modeling. The file can include up to five RNA 3D models, each encompassed by the MODEL/END keywords. Models are formatted according to the established standards of the RNA-Puzzles community[9]:

- 3D coordinates are provided for the complete list of atoms for all nucleotides from the target FASTA file.
- only natural nucleotides (A, C, G, U) are allowed.
- if present, modified monomeric units are transformed into unmodified ones by discarding atypical atoms.
- only atoms from the following sets – (C2, C4, C6, C8, N1, N2, N3, N4, N6, N7, N9, O2, O4, O6) for nucleobases, and (C1', C2', C3', C4', C5', O2', O3', O4', O5', OP1, OP2, P) for sugar-phosphate backbone are allowed.

In case of protein-RNA complexes, protein chains are designated with letters (A, B, C, . . .) and RNA chains with numbers (0, 1, 2, . . .).

An example of RNA prediction is provided in Example 3 on the CASP15 format page*https://predictioncenter.org/casp15/index.cgi?page=format*.

2.1.2 | Preparation of targets and model templates. The CASP organizers prepare a FASTA file with the sequence of targeted RNA. The file begins with a header containing target ID (e.g., >R1117) and chain IDs (i.e., numbers from 0 to 9) of all strands in the target structure. The body of the file includes nucleic acid sequence(s). In addition, the organizers generate a 3D structure template using the RNA-Puzzles formatting tool[13]. The template is a PDB file containing all the required ATOM records with zeroed coordinate values. The information on targets is communicated to participating groups via the CASP web portal (e.g.,*https://predictioncenter.org/casp15/target.cgi?id=30&view=all*).

Prior to submission, predictors can verify compatibility of their models with the provided templates by running the RNA-Puzzles tool that checks the number and ordering of residues and atoms in the submission[13]. If a prediction file does not comply with the requirements, error messages are reported to a log file. Non-compliant files can be reformatted with the rna_pdb_toolsx.py script available from the rna-tools toolbox[13,14].

2.1.3 | Setting the acceptance system. At the target release time, each target is assigned a prediction time window, which is typically 3 days for servers and 3 weeks for expert groups. RNA structure models are accepted within the specified prediction window via email or dedicated CASP prediction submission webform. The CASP submission system automatically checks submissions for compliance with the deadlines and format requirements and provides feedback to predictors. The prediction format is checked with the same tools used to generate model templates (section 2.1.2). If a prediction is rejected, an error message is sent to the submitter, and they have until the target deadline to fix the reported issue(s) and resubmit.

3

Accepted predictions are stored in the CASP system and eventually evaluated after the target structure becomes available.

The same submission rules apply to other prediction categories discussed further in this paper.

2.1.4 | RNA evaluation measures. Predictions in the RNA category are assessed by checking their geometric plausibility and comparing them with target structures. When alternative target structures were available, given the early stage of RNA modeling, we reported the best score per model. Evaluation measures include Clashscore [15], Root Mean Square Deviation (RMSD)[16], Local Distance Difference Test (lDDT)[17], Template Modeling score (TM-score) [18], and Global Distance Test - Total Score (GDT-TS)[19]. These are commonly used measures in protein-CASP evaluation that are also adopted here for RNA evaluation. However, none of these measures are suitable for assessing RNA-specific components, like canonical (G-C, A-U, G-U), non-canonical, and stacking interactions between the nucleobases that contribute to RNA folding and stabilization. Proper prediction of only canonical interactions is usually insufficient to obtain a good model of an RNA molecule (example in Figure 1), while prediction of non-canonical interactions is very valuable but hard to achieve due to high computational demands. We additionally consider an RNA-specific measure, Interaction Network Fidelity (INF)[13,20], which evaluates different types of RNA interactions in models. Calculation of these measures requires prior determination of RNA interactions from the atomic coordinates. This is done using 2D structure annotators such as RNAView[21], MC-Annotate[22], ClaRNA[23] or FR3D[24], which provide base pairs and their classification[25]. Given two sets of interactions, one for the model and another for the target, we identify true positives (correctly predicted base pairs), false positives (unpredicted base pairs), and false negatives (incorrectly predicted base pairs), and then calculate the INF score as the Matthews correlation coefficient[26]. The score ranges from [0.0, 1.0], with higher scores indicating better prediction of base-base interactions. The INF score is determined for all interactions (INF_all), and separately for canonical (Watson-Crick, INF_WC), non-canonical (non-Watson-Crick, INF_nWC), and stacking (INF_stacking) interactions.
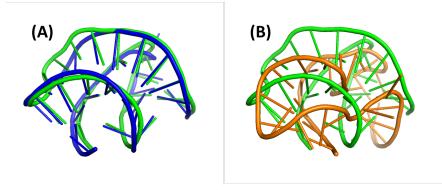

Figure 1. Two models of PreQ1 class I type III riboswitch superimposed with the target structure (R1117, green). Model R1117TS119_3 (orange, panel B) correctly reproduces all canonical base pairs at the same time not reproducing any non-canonical interactions and having the RMSD of 9.90 . Model R1117TS287_2 (blue, panel A) misses 7% of canonical base pairs but is a much better model with the RMSD of 2.25 and non-canonical interactions reproduction rate of 39%. RMSD scores were computed on the 2-30 fragment.

Many RNA targets were derived from low-to-medium resolution (4-7) cryo-EM maps. To remove bias and modeling uncertainties potentially present in the reference structures, we additionally evaluated predictions by directly comparing them to maps. After docking models into maps, predictions were scored with five different measures: CC_{mask} (cross-correlation masked by the area around the model), CC_{peaks} (cross-correlation masked by the model and the highest density peaks in the map), MI (mutual information), SMOC (segment-based Mander's overlap coefficient), and AI (atomic inclusion)[27,28]. It is worth noting that model-independent measures are known to overinflate scores for compressed models with unrealistic geome-

tries. Thus, when using these scores, it is important to first filter out models with poor topology scores or, alternatively, combine these scores with geometric scores, such as Clashscore.

2.2 | Protein-ligand complexes.

The ligand topic is not new to CASP: in CASPs 6 through 10 predicting ligand binding sites was a sub-challenge in the function prediction category [29-32]. Given the recent advances in the accuracy of protein modeling methods[11,12], CASP organizers decided to include prediction of protein- and RNA- small molecule ligand complexes into the scope of CASP15 experiment hoping to boost development of methods in this area. Participants are provided with the sequence and stoichiometry of protein (or RNA) receptors, Simplified Molecular Input Line Entry System (SMILES) codes of bound ligands, and are asked to predict structures of protein- (RNA-) ligand complexes.

2.2.1 | Macromolecule-ligand complex prediction format (https://predictioncenter.org/casp15/ index.cgi?page=format#LG). One important requirement for the ligand prediction format was the need to encode atom connectivity in a robust and reliable manner, as the correct atom connectivity is required for symmetry correction, a necessary step in accurate ligand assessment. Unfortunately, the PDB format, which is commonly used in CASP, is not able to reliably encode connectivity for arbitrary ligands. The MDL molfile format [33] is a common format for ligands which was used in earlier ligand docking challenges such as D3R[34-37]. This is a text-based, fixed column format that besides atom coordinates also encodes the bonds. Unlike the PDB format, atoms are not named and only identified by their element and connectivity. The format allows reporting additional properties such as charge, valence, or isotope, but those were not required nor used here. Bonds between atoms are encoded explicitly, one by line, together with the bond type (single, double, triple, or aromatic) and stereochemistry. The format also includes header lines, a COUNTS line, which can help check the integrity of the file, and an M END line which indicates the end of the ligand data.

For CASP15, we devised a hybrid submission format where the receptor's model (protein or RNA) and ligand model are submitted as separate files in the same spatial frame of reference. The receptor is submitted in the PDB format, while the ligand in MDL (see below for details). Similarly to the regular protein structure submission, a CASP ligand submission (LG format) starts with a CASP header including format specification code, target identifier, author identifier, and description of the modeling method. Two new keywords are introduced: the LIGAND keyword, which defines ligand name and the beginning of the ligand data, and the POSE keyword, which defines the pose number for the selected ligand. Participants are allowed to submit up to 5 poses of a given ligand for a selected receptor model.

An example of LG prediction is provided in Example 6 on the CASP15 format page.

2.2.2 | Preparation of targets and model templates.

A FASTA file of the receptor sequence is prepared by the CASP organizers. For known small molecules, SMILES are retrieved from the PDB component dictionary. In the case of novel small molecules (not present in the PDB component dictionary), SMILES are provided by the experimentalists. In both cases, SMILES are compared and modified based on those derived from the PDB coordinates. If necessary, stereochemistry is assigned using the *AssignStereochemistryFrom3D* function from RDKit, and the protonation state adjusted by manually editing the SMILES based on the visual inspection of protein-ligand interactions.

The relevance of each small molecule is decided case by case for each target. Only biologically relevant small molecules are retained. Common crystallographic reagents and ions are ignored if not interacting with the small molecules or part of a structural motif (e.g., zinc binding motif).

A script to prepare prediction templates (MDL files) is provided by the CASP organizers. It is implemented in python 3 and RDKit python bindings (*http://www.rdkit.org/*) . The script initially converts the input SMILES strings to RDKit Mol objects using the *rdkit.Chem.MolFromSmiles* method. At this stage, the Mol objects contain only the information related to small molecule properties, like atom types and bond formation. Coordinate section is added to the Mol objects using the RDKit's ETKDG method[38]. Subsequently, the Mol objects are written to the MDL-formatted file[33], which can be used as a ligand submission template.

2.2.3 | Setting up the acceptance system. Validation of ligand predictions is performed with scripts written in python 2.7 and RDKit. Initial checks verify the CASP header section (availability and correctness of PFRMAT, TARGET, AUTHOR, and MODEL/END records). Once submissions have passed this phase, ligand models are converted to RDKit Mol objects and compared with the template for downstream evaluation. Each molecule in the submitted file is validated by comparison with a reference Mol object generated from the corresponding SMILES string as described above. To validate the submissions, comparisons of the following parameters are undertaken:

- number of atoms and their types,
- number of bonds,
- bond types and atom types in bond pairs (e.g., C-C Single, or C=O Double).

Additionally, to account for atom connectivity and chirality in submitted models, the maximum common substructures between the submitted and reference ligands are calculated using the *FindMCS* function in RDKit. To pass the validation, a molecule must have a maximum common substructure equal to the number of atoms in the reference model.

Finally, a validation report is created showing the results of the validation process to aid in troubleshooting invalid submissions.

2.2.4 | Macromolecule-ligand complex evaluation measures. Previous ligand docking challenges like Teach Discover Treat (TDT) [39], Continuous Evaluation of Ligand Prediction Performance (CELPP)[40] and Drug Discovery Data Resource (D3R)[34-37] have used two main types of metrics to assess how well participants can model receptor-ligand complexes. These evaluated how close a predicted ligand is to the target within the binding site in absolute terms with the RMSD metric, and how well the native receptor-ligand interactions are reproduced. CASP experiment brings additional assessment challenges: (1) because the receptor structure is not given but rather modeled, ligands in the model and reference complexes can be bound to different configurations of binding sites, and thus calculation of any superposition-based scores requires preliminary alignment of binding pockets with ligands in two complexes, which is not a trivial task; (2) chain mapping needs to be established; (3) incomplete ligands in some targets require partial graph matching for the symmetry correction; and (4) multiple copies of ligands in the targets and models have to be mapped (assigned) uniquely, in order to avoid scoring target or predicted ligands multiple times.

To address these challenges, we developed two scores, which are described in more detail in the CASP15 Ligand Assessment paper[7]. The Binding-Site Superposed, Symmetry-Corrected Pose Root Mean Square Deviation (BiSyRMSD) score defines the binding sites and the superpositions to compute RMSDs between target's and model's ligands. The Local Distance Difference Test for Protein-Ligand Interactions (lDDT-PLI) measure assesses how well native contacts between the receptor and the ligand are reproduced in the model with an lDDT-based metric and symmetry correction. When used in combination, these scores give a better account of modeling receptor-ligand complexes.

2.3 | Estimation of model accuracy (EMA) for oligomeric targets.

The EMA category has been an integral part of every CASP experiment starting with CASP7[41-48]. It has attracted the attention of many developers, with over 70 methods tested in the previous CASP experiment[48]. An emphasis on the importance of this category led to very positive developments in protein structure prediction as modelers now routinely integrate quality estimates into their modeling pipelines. In particular, the CASP14-winning AlphaFold2 method offers reliable estimates of global and local accuracy of their models[10,11].

In CASP15, the focus of the EMA category shifted from predicting accuracy of single-sequence proteins to multi-molecular complexes.

2.3.1 | Model accuracy prediction format(*https://predictioncenter.org/casp15/index.cgi ?page=format#QA*)**.** For global (whole model) accuracy prediction (QMODE1), participants are asked to submit a fold similarity score (SCORE, in 0-1 range), which estimates the similarity of model's overall

fold to the target's one, and an interface similarity score (QSCORE, also in 0-1 range), which evaluates reliability of quaternary structure interfaces. Submitting the QSCORE is optional, and predictors can skip it by putting 'X' symbol in the corresponding place of a QA prediction (see the link above). In QMODE2 (local accuracy), in addition to the QMODE1 scores, the predictors are asked to assign confidence scores to the interface residues of the model, indicating their likelihood of being present in the native structure's interface. Interface residues are identified as having contact with at least one residue from a different chain, with a Cβ-Cβ distance not exceeding 8Å (or Cα, if the residue is Glycine).

Examples of EMA predictions in QMODE1 and QMODE2 are provided in Example 5 on the CASP15 format page.

2.3.2 | Submission collecting process. EMA predictions in CASP15 are requested for all (and only) multimeric targets. In contrast with previous CASPs, EMA targets are released after all models (and not only server models) are collected on the corresponding structure prediction target. A tarball with assembly predictions from all CASP groups is created the next day after the TS target closure, and a link to the tarball file is pushed to the EMA servers and posted at the CASP15 website. All EMA groups, regardless of their type (i.e., 'server' or 'human') have 2 days to return accuracy estimates for TS models included in the tarball file. The predictions are checked with the verification scripts, and successful predictions are saved for subsequent evaluation.

2.3.3 | EMA evaluation measures. Global predictions were compared with established evaluation metrics possessing the desired attributes. This is the oligomeric Template Modelling score (TM-score)[49] for overall topology (SCORE) and the contact based QS-score[50] which is interface centric (QSCORE). To ensure a comprehensive evaluation, these metrics were supplemented with additional measures. An oligomeric GDT-like score, referred to as oligo-GDTTS, was employed for overall topology analysis, and a variant of the interface centric DockQ score[51]. Notably, DockQ evaluates pairwise interfaces, necessitating the introduction of a weighted average metric—termed DockQ-wave—to effectively score higher-order complexes. Local predictions were compared against the per-residue lDDT [17] and CAD (AA-variant) [52]scores, which assess the accuracy of relative atom positions in the neighborhood, including neighboring chains. Conceptually the scores are contact-based, but do not penalize for added contacts, which is relevant in case of incorrect interfaces. To address this limitation, two novel local variants of the QS-score and DockQ have been introduced: PatchQS and PatchDockQ. All evaluation metrics are described in detail in the CASP15 EMA Assessment paper[6].

2.4 | Ensembles of alternative conformations.

Following the success of deep-learning methods for single structures, it is increasingly important to assess methods for predicting ensembles of alternative conformations. While deep learning and other methods have the potential to generate ensembles in some circumstances, these abilities have never been rigorously tested. In CASP15, we made a first attempt to include this category. For CASP purposes, we categorize ensembles [8] as: (1) cases where a macromolecule populates multiple conformations under the same environmental and chemical conditions (including intrinsically disordered proteins or parts of proteins; vibrational motion; local alternative conformations; 'ghost' conformations which are present at low level but are dominant in other conditions; and folding intermediates.). (2) Cases where a macromolecule adopts different conformations in response to environment or chemical change (ligand binding; macromolecular complex formation; post-translational modification; mutations; and crystal, pH and other environmental changes). A third category of ensembles we consider is the set of conformations consistent with the experimental data. The latter is an increasingly important category both because of the now common high accuracy of the computed structures and the inclusion of lower resolution data in CASP.

Targets for alternative conformers do not require separate prediction formats as they are 3D structures routinely processed in CASP, but they do require a mechanism for submitting multiple models. In CASP15, this need was handled in two different ways. In some cases, different alternative conformations were treated as separate targets. In particular,

- two targets were assigned for modeling an isocyanide hydratase represented by a wild-type structure (target T1110) and its one-point mutant T1109, where amino acid D183 was changed to A183,
- two targets (R1107 and R1108) were assigned for modeling human and chimpanzee CPEB3 ribozymes, which differ by a single nucleic acid A30 (human) - G30 (chimpanzee),
- two pairs of targets (TR1189 and TR1190) were assigned for modeling complexes of the metabolite repressor protein (RsmA) and a non-coding RNA (RsmZ). Both complexes contain one RNA molecule but different number of protein molecules (6 in TR1189 and 4 in TR1190),
- five targets (T1158v0-v4) were assigned for modeling a type IV ABC transporter, where five different conformations have been observed, depending on environmental conditions (ligand binding).

In other cases, participants were encouraged to submit multiple conformers using the standard CASP five models target format. This approach was used for

- three kinases (CASP targets T1195-T1197), each of which has two to three sets of experimental coordinates representing different conformations,
- the Holliday Junction complexes (targets T1170, H1171, H1172), some subunits of which are deformed due to the contact with DNA and other protein molecules in the complex,
- RNA origami target R1138, which was solved in a kinetically trapped young state and the mature state,
- SL5 domain of the RNA betacoronavirus structure BtCoV-HKU (CASP target R1156), where one of the helices accepted multiple relative conformations with respect to the remainder of the structure.

3 | CONCLUSION

This paper summarizes the technical aspects of new prediction categories in CASP15 and can serve as a reference point for future predictors. The systems for handling RNA/ligand/multimer-EMA predictions in CASP15 were implemented shortly before the start of the experiment and, as such, were not extensively tested. Based on the feedback from CASP15 assessors and predictors, we plan to improve the system by further automating the process of verifying RNA and ligand submissions and ensuring better compliance of models with the required format. This investment of time and effort should pay off at the assessment stage with a more seamless evaluation of predictions. Since the PDB format is becoming obsolete, we will motivate predictors and developers of the evaluation measures used in CASP to support the PDBx/mmCIF format in their tools and methods. For the RNA category, we will work to further incorporate and standardize evaluation measures which assess the base-base interaction network. This will require setting community-accepted standards in identification and annotation of RNA interactions[22]. Further, we will work to improve measures comparing predictions to raw data to assess low-resolution targets. For the ligand prediction category, we need to improve the set of targets with more interesting ligand modeling challenges, possibly by tapping into unreleased structures from pharmaceutical companies. We will also improve the acceptance system by validating the identity of submitted ligands in a more robust manner, and we would like to consider introducing metrics for self-assessment of the accuracy of ligand poses within the receptor. For evaluating ensembles of alternative conformations, we will explore acceptance of multi-model entities, with optional probability weights, as submissions.

REFERENCES

1. Moult J, Pedersen JT, Judson R, Fidelis K. A Large-Scale Experiment to Assess Protein-Structure Prediction Methods. Proteins-Structure Function and Genetics 1995;23(3):R2-R4.

2. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)-Round XIII. Proteins 2019;87(12):1011-1020.

3. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)-Round XIV. Proteins-Structure Function and Bioinformatics 2021;89(12):1607-1617.

4. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J. Critical assessment of methods of protein structure prediction (CASP)-Round XV. Proteins 2023(This issue).

5. Das R, Kretsch RC, Simpkin AJ, Mulvaney T, Pham P, Rangan R, Bu F, Keegan RM, Topf M, Rigden DJ, Miao Z, Westhof E. Assessment of three-dimensional RNA structure prediction in CASP15. Proteins 2023(This issue).

6. Studer G, Tauriello G, Schwede T. Assessment of the assessment - All about complexes. Proteins 2023(This issue).

7. Robin X, Walters P, al e. Assessment of ligand prediction in CASP15. Proteins 2023(This issue).

8. Moult J, Kryshtafovych A, al e. Ensembles of alternative confromations in CASP15. Proteins 2023(This issue).

9. Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, Huang LL, Lavender CA, Lisi V, Major F, Mikolajczak K, Patel DJ, Philips A, Puton T, Santalucia J, Sijenyi F, Hermann T, Rother K, Rother M, Serganov A, Skorupski M, Soltysinski T, Sripakdeevong P, Tuszynska I, Weeks KM, Waldsich C, Wildauer M, Leontis NB, Westhof E. RNA-Puzzles: A CASP-like evaluation of RNA three-dimensional structure prediction. Rna 2012;18(4):610-625.

10. Carrascoza F, Antczak M, Miao ZC, Westhof E, Szachniuk M. Evaluation of the stereochemical quality of predicted RNA 3D models in the RNA-Puzzles submissions. Rna 2022;28(2):250-262.

11. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Applying and improving AlphaFold at CASP14. Proteins-Structure Function and Bioinformatics 2021;89(12):1711-1721.

12. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D. Highly accurate protein structure prediction with AlphaFold. Nature 2021;596(7873):583-+.

13. Magnus M, Antczak M, Zok T, Wiedemann J, Lukasiak P, Cao Y, Bujnicki JM, Westhof E, Szachniuk M, Miao ZC. RNA-Puzzles toolkit: a computational resource of RNA 3D structure benchmark datasets, structure manipulation, and evaluation tools. Nucleic Acids Research 2020;48(2):576-588.

14. Magnus M. rna-tools.online: a Swiss army knife for RNA 3D structure modeling workflow. Nucleic Acids Research 2022;50(W1):W657-W662.

15. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35(Web Server issue):W375-383.

16. Kabsch W. A solution for the best rotation to relate two sets of vectors. Acta Crystallogr 1976;A32:922-923.

17. 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(21):2722-2728.

18. Gong S, Zhang CX, Zhang Y. RNA-align: quick and accurate alignment of RNA 3D structures based on size-independent TM-score(RNA). Bioinformatics 2019;35(21):4459-4461.

19. Zemla A. LGA: A method for finding 3D similarities in protein structures. Nucleic Acids Res 2003;31(13):3370-3374.

20. Parisien M, Cruz JA, Westhof E, Major F. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. Rna 2009;15(10):1875-1885.

21. Yang HW, Jossinet F, Leontis N, Chen L, Westbrook J, Berman H, Westhof E. Tools for the automatic identification and classification of RNA base pairs. Nucleic Acids Research 2003;31(13):3450-3460.

22. Gendron P, Lemieux S, Major F. Quantitative analysis of nucleic acid three-dimensional structures. Journal of Molecular Biology 2001;308(5):919-936.

23. Walen T, Chojnowski G, Gierski P, Bujnicki JM. ClaRNA: a classifier of contacts in RNA 3D structures based on a comparative analysis of various classification schemes. Nucleic Acids Research 2014;42(19).

24. Sarver M, Zirbel CL, Stombaugh J, Mokdad A, Leontis NB. FR3D: finding local and composite recurrent structural motifs in RNA 3D structures. J Math Biol 2008;56(1-2):215-252.

25. Leontis NB, Westhof E. Geometric nomenclature and classification of RNA base pairs. Rna 2001;7(4):499-512.

26. Matthews BW. Comparison of Predicted and Observed Secondary Structure of T4 Phage Lysozyme. Biochimica Et Biophysica Acta 1975;405(2):442-451.

27. Liebschner D, Afonine PV, Baker ML, Bunkoczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ, Adams PD. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D 2019;75:861-877.

28. Cragnolini T, Sahota H, Joseph AP, Sweeney A, Malhotra S, Vasishtan D, Topf M. TEMPy2: a Python library with improved 3D electron microscopy density-fitting and validation workflows. Acta Crystallogr D 2021;77:41-47.

29. Soro S, Tramontano A. The prediction of protein function at CASP6. Proteins 2005;61 Suppl 7:201-213.

30. Lopez G, Rojas A, Tress M, Valencia A. Assessment of predictions submitted for the CASP7 function prediction category. Proteins 2007;69 Suppl 8:165-174.

31. Schmidt T, Haas J, Gallo Cassarino T, Schwede T. Assessment of ligand-binding residue predictions in CASP9. Proteins 2011;79 Suppl 10(Suppl 10):126-136.

32. Gallo Cassarino T, Bordoli L, Schwede T. Assessment of ligand binding site predictions in CASP10. Proteins 2014;82 Suppl 2(0 2):154-163.

33. Dalby A, Nourse JG, Hounshell WD, Gushurst AKI, Grier DL, Leland BA, Laufer J. Description of Several Chemical-Structure File Formats Used by Computer-Programs Developed at Molecular Design Limited. J Chem Inf Comp Sci 1992;32(3):244-255.

34. Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB, Jr., Carlson HA, Burley SK, Walters WP, Amaro RE, Feher VA, Gilson MK. D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity predictions. Journal of computer-aided molecular design 2016;30(9):651-668.

35. Gaieb Z, Liu S, Gathiaka S, Chiu M, Yang H, Shao C, Feher VA, Walters WP, Kuhn B, Rudolph MG, Burley SK, Gilson MK, Amaro RE. D3R Grand Challenge 2: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. Journal of computer-aided molecular design 2018;32(1):1-20.

36. Gaieb Z, Parks CD, Chiu M, Yang H, Shao C, Walters WP, Lambert MH, Nevins N, Bembenek SD, Ameriks MK, Mirzadegan T, Burley SK, Amaro RE, Gilson MK. D3R Grand Challenge 3: blind prediction of protein-ligand poses and affinity rankings. Journal of computer-aided molecular design 2019;33(1):1-18.

37. Parks CD, Gaieb Z, Chiu M, Yang H, Shao C, Walters WP, Jansen JM, McGaughey G, Lewis RA, Bembenek SD, Ameriks MK, Mirzadegan T, Burley SK, Amaro RE, Gilson MK. D3R grand challenge 4: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. Journal of computer-aided molecular design 2020;34(2):99-119.

38. Riniker S, Landrum GA. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. Journal of Chemical Information and Modeling 2015;55(12):2562-2574.

39. Jansen JM, Cornell W, Tseng YJ, Amaro RE. Teach-Discover-Treat (TDT): collaborative computational drug discovery for neglected diseases. Journal of molecular graphics & modelling 2012;38:360-362.

40. Wagner JR, Churas CP, Liu S, Swift RV, Chiu M, Shao C, Feher VA, Burley SK, Gilson MK, Amaro RE. Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking. Structure 2019;27(8):1326-1335 e1324.

41. Cozzetto D, Kryshtafovych A, Ceriani M, Tramontano A. Assessment of predictions in the model quality assessment category. Proteins-Structure Function and Bioinformatics 2007;69:175-183.

42. Cozzetto D, Kryshtafovych A, Tramontano A. Evaluation of CASP8 model quality predictions. Proteins-Structure Function and Bioinformatics 2009;77:157-166.

43. Kryshtafovych A, Fidelis K, Tramontano A. Evaluation of model quality predictions in CASP9. Proteins-Structure Function and Bioinformatics 2011;79:91-106.

44. Kryshtafovych A, Barbato A, Fidelis K, Monastyrskyy B, Schwede T, Tramontano A. Assessment of the assessment: Evaluation of the model quality estimates in CASP10. Proteins-Structure Function and Bioinformatics 2014;82:112-126.

45. Kryshtafovych A, Barbato A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano A. Methods of model accuracy estimation can help selecting the best models from decoy sets: Assessment of model accuracy estimations in CASP11. Proteins-Structure Function and Bioinformatics 2016;84:349-369.

46. Kryshtafovych A, Monastyrskyy B, Fidelis K, Schwede T, Tramontano A. Assessment of model accuracy estimations in CASP12. Proteins-Structure Function and Bioinformatics 2018;86:345-360.

47. Won J, Baek M, Monastyrskyy B, Kryshtafovych A, Seok C. Assessment of protein model structure accuracy estimation in CASP13: Challenges in the era of deep learning. Proteins-Structure Function and Bioinformatics 2019;87(12):1351-1360.

48. Kwon S, Won J, Kryshtafovych A, Seok C. Assessment of protein model structure accuracy estimation in CASP14: Old and new challenges. Proteins-Structure Function and Bioinformatics 2021;89(12):1940-1948.

49. Zhang C, Shine M, Pyle AM, Zhang Y. US-align: universal structure alignments of proteins, nucleic acids, and macromolecular complexes. Nature methods 2022;19(9):1109-1115.

50. Bertoni M, Kiefer F, Biasini M, Bordoli L, Schwede T. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci Rep-Uk 2017;7(1):10480.

51. Basu S, Wallner B. DockQ: A Quality Measure for Protein-Protein Docking Models. PLoS ONE 2016;11(8):e0161879.

52. Olechnovic K, Kulberkyte E, Venclovas C. CAD-score: a new contact area difference-based function for evaluation of protein structural models. Proteins 2013;81(1):149-162.