Title

Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking

Permalink

https://escholarship.org/uc/item/69w569vh

Journal

Structure, 27(8)

ISSN

0969-2126

Authors

Wagner, Jeffrey R

Churas, Christopher P

Liu, Shuai

et al.

Publication Date

2019-08-01

DOI

10.1016/j.str.2019.05.012

Peer reviewed

Published in final edited form as:
Structure. 2019 August 06; 27(8): 1326–1335.e4. doi:10.1016/j.str.2019.05.012.

Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking

Jeffrey R. Wagner[1], **Christopher P. Churas**[1], **Shuai Liu**[1], **Robert V. Swift**[1], **Michael Chiu**[1], **Chenghua Shao**[2], **Victoria A. Feher**[1], **Stephen K. Burley**[2,3], **Michael K. Gilson**[*,1,4], **Rommie E. Amaro**[*,1,5]

[1]Drug Design Data Resource; University of California San Diego, La Jolla, CA 92093, USA

[2]RCSB Protein Data Bank, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

[3]Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA

[4]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA

[5]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, USA

[*]Corresponding Authors: MKG mgilson@ucsd.edu, REA ramaro@ucsd.edu.
Lead Contact: Rommie E Amaro, ramaro@ucsd.edu

[7]Author Contributions

Conceptualization and Methodology, J.R.W., S.L., R.V.S., V.A.F., S.K.B, M.K.G, and R.E.A.; Software and Validation J.R.W., C.P.C., S.L., R.V.S, M.C, and C.S.; Investigation J.R.W., S.L.; Resources C.P.C., M.C., C.S., S.K.B.; Writing – Original Draft, J.R.W.; Writing – Review & Editing, J.R.W. S.K.B., M.K.G., R.E.A.; Supervision and Funding Acquisition, V.A.F., S.K.B., M.K.G., R.E.A.

[6]Financial Disclosures

MKG has an equity interest in, and is a cofounder and scientific advisor of, VeraChem LLC. REA has equity interest in, and is a cofounder and scientific advisor of Actavalon, Inc. VAF has equity interest in Actavalon, Inc.

Data and Software Availability

All structures, RMSDs, and workflow settings used in this analysis are available at https://doi.org/10.6075/J0610XPS. Note that:
Predicted structures and calculated RMSD values are available from 2017 week 5 to 2018 week 28, however several weeks in this interval were excluded from the analysis, as less than three predictions were submitted for any target.
2017 week 11 is excluded from the analysis as it had ~300 nearly-identical complexes released.
The files contained in this repository are grouped by submitter, but have had comment fields scrubbed of information to preserve the anonymity of external participants.
The settings and software environment of the GLIDE participant were changed during the "live" data collection period, so all GLIDE predictions were regenerated for this paper using Schrodinger suite version 2015–3 and the code at https://github.com/drugdata/d3r_in_house_glide master branch as of commit fb38a3a.
Sporadic automation and infrastructure issues prevented predictions from being generated or evaluation from being run for some targets. Notably, GLIDE predictions in 2018 between weeks 1–9 had a configuration error and did not generate RMSDs.
External participants enrolled in CELPP at different times, were not required to submit predictions each week, and were not required to use the same method throughout the data collection period.

Data S1: **Complete Table of Targets and Candidates** (Related to STAR Methods: Data Availability). A CSV table containing information about all targets and candidates analyzed in this paper. This table contains a subset of the information that CELPP sends to each participant when a target is selected for a challenge week. Columns in this table are: Target complex PDB ID, Target complex crystallization pH, Target Ligand ID, Target ligand InChI, Target ligand heavy atom count, Target Ligand's number of rotatable bonds, LMCSS candidate PDB ID,SMCSS candidate PDB ID, hiResApo candidate PDB ID, hiResHolo candidate PDB ID, hiTanimoto candidate PDB ID.

Contact for Reagent and Resource Sharing

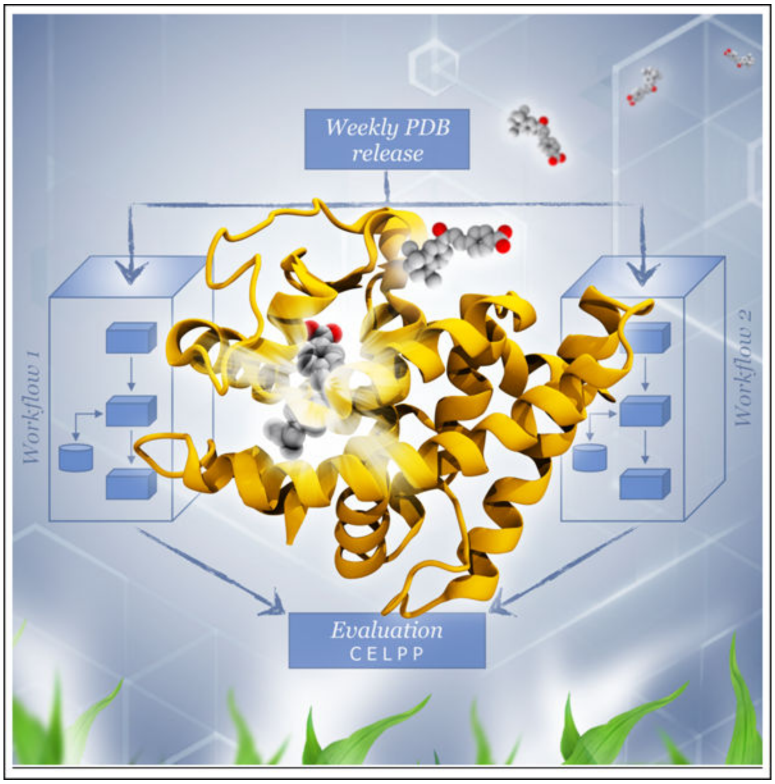
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rommie E Amaro (ramaro@ucsd.edu).

Summary

1

Docking calculations can accelerate drug discovery by predicting the bound poses of ligands for a targeted protein. However, it is not clear which docking methods work best. Furthermore, predicting poses requires steps outside the docking algorithm itself, such as preparation of the protein and ligand, and it is not known which components are most in need of improvement. The Continuous Evaluation of Ligand Protein Predictions (CELPP) is a blinded prediction challenge designed to address these issues. Participants create a workflow to predict protein-ligand binding poses, which is then tasked with predicting 10–100 new protein-ligand crystal structures each week. CELPP evaluates the accuracy of each workflow's predictions and posts the scores online. The results can be used to identify the strengths and weaknesses of current approaches, help map docking problems to the algorithms most likely to overcome them and illuminate areas of unmet need in structure-guided drug design.

Graphical Abstract



2 Introduction

The discovery of a small molecule that binds a disease-related protein with high affinity is a key step in many drug discovery projects. The process is perhaps most efficient when a high-resolution structure of the targeted protein is available, such as from X-ray crystallography, because then structure-based computational methods may be used to accelerate the discovery of high affinity ligands (Amaro et al., 2018; Anna et al., 2017; Bartuzi et al., 2017;

Ganesan and Barakat, 2017; Kuhn et al., 2016; Leelananda and Lindert, 2016; Lybrand, 1995; Raghavendra et al., 2018; Rosenfeld et al., 1995; Santiago et al., 2017; Sledz and Caflisch, 2017; Sliwoski et al., 2014). The computational challenge of structure-based ligand design comprises two main components. One is prediction of the bound conformation, or pose, of a candidate ligand, typically by fast, ligand-protein docking algorithms (Amaro et al., 2008; Chen et al., 2003; Cheng et al., 2012; Doman et al., 2002; Ewing and Kuntz, 1997; Goodsell et al., 1996; Goodsell and Olson, 1990; Guedes et al., 2014; Knegtel et al., 1997; Kuntz, 1992; McGann et al., 2002; Sinko et al., 2013; Sousa et al., 2006; Trott and Olson, 2009; Yuriev et al., 2015). The second involves using the predicted pose to assess the candidate ligand's binding affinity for the targeted protein. Both components have been the subject of intensive research and development in both academic and commercial settings (Beck et al., 2017; Ciancetta and Jacobson, 2017; Hodos et al., 2016; Kim et al., 2017; Li et al., 2018; Mafud et al., 2016; Martinez-Mayorga et al., 2015; Medina-Franco et al., 2015; Morgnanesi et al., 2015; Ogungbe and Setzer, 2016; Rosano et al., 2016; Shunmugam et al., 2017; Singh and Ecker, 2018; Tan et al., 2016; Topiol and Sabio, 2015; Zhu et al., 2015). Nonetheless, computational methods for pose prediction and affinity ranking have yet to fulfill their perceived promise, as neither is yet fully reliable (Gaieb et al., 2017; Gaieb et al., 2019; Gathiaka et al., 2016; Muddana et al., 2012; Muddana et al., 2014; Warren et al., 2006; Yin et al., 2017). It is difficult even to compare the reliability of various methods in a consistent manner, so it is correspondingly difficult to make and verify technical progress.

The challenge of rigorously comparing methods derives in part from the difficulty of reproducing the complicated end-to-end computations required for pose and affinity prediction (Jansen et al., 2012). This problem has generated growing interest in automated workflows that clearly memorialize a method (Balasubramanian et al., 2016; Beisken et al., 2013; Dakka et al., 2018; Pronk et al., 2015; Purawat et al., 2017). Also, although many performance comparisons have been published, the results can be difficult to interpret (Abel et al., 2017; Cole et al., 2005; Corbeil et al., 2012; Damm-Ganamet et al., 2013; Grinter and Zou, 2014; Irwin, 2008; Jain, 2008; McGann, 2012; Mobley and Dill, 2009; Neves et al., 2012; Spitzer and Jain, 2012; Warren et al., 2006). Thus, descriptions of new docking algorithms may include comparisons with existing methods, but the comparison is often secondary to the description of the new algorithm and hence not fully developed. Additionally, different methods are typically tested against different sets of protein-ligand complexes, so a consistent set of comparisons may not be available. Finally, studies that benchmark multiple methods against a common dataset, the data often include protein-ligand cocrystal structures that have already been published. Such retrospective studies risk unintentional bias, and the tests may include structures that were used in training the docking algorithms(Weiss et al., 2016).

Several initiatives have addressed these limitations through prospective, or blinded, prediction challenges. In such challenges, researchers evaluate methods against a common set of test cases for which the experimental structures are withheld until after the computational predictions have been made. Prior blinded challenges include the GSK challenge(Warren et al., 2006), CSAR (Carlson, 2016; Carlson et al., 2016; Damm-Ganamet et al., 2013; Smith et al., 2016; Smith et al., 2011), and GPCRDOCK (Kufareva et al., 2014; Kufareva et al., 2011; Michino et al., 2009). Similarly, in recent years, the Drug Design

Resource (D3R) has run blinded prediction challenges called the Grand Challenges (Gaieb et al., 2017; Gathiaka et al., 2016). These efforts have provided insight about best practices and sometimes unexpected results regarding the effectiveness of various technical approaches. However, the challenges to date are not been large enough to afford statistically meaningful distinctions among individual methods or to support an efficient cycle of development and evaluation that can accelerate progress in the field.

Here, we introduce Continuous Evaluation of Ligand Protein Predictions (CELPP), a new blinded prediction challenge which overcomes these limitations. CELPP uses the Protein Data Bank's (PDB) (Berman et al., 2003; Berman et al., 2000; Burley et al., 2018; Rose et al., 2017; Young et al., 2017) weekly publication of a list of structures slated for imminent to drive a weekly, pose-prediction challenge akin to the Continuous Automated Model Evaluation (CAMEO) protein structure prediction challenge (Haas et al., 2017), which served as its inspiration. Here, we detail the challenge, the automation used to enable smooth weekly operations, initial results for a number of docking workflows, and implications and directions for this community science project.

3 Results

3.1 Overview of the CELPP Blinded Cross-Docking Challenge

Each week (Figure 1), in-house CELPP scripts download the list of new PDB entries to be released five days later and identify those with protein-small molecule cocrystal structures suitable for automated docking calculations (https://github.com/drugdata/D3R). At this stage, the only information available about each of these structures, called target complexes, is the identity of the ligand, the amino acid sequence of the protein, and the pH of the crystallographic mother liquor. Additional scripts then search the PDB for crystal structures of each target protein and extract up to five structures appropriate for docking calculations (STAR Methods: Selection of target complexes and receptor structures). These structures are incorporated into the weekly CELPP data package, along with the ligand identities, crystallization pH values, and additional information (STAR Methods: Generation of the challenge data package). CELPP participants download the data package, run their own workflows to predict the ligand binding poses, and submit their predictions to a personal, password-protected web directory before the deadline, which is shortly before release of the new PDB entries containing the actual crystallographic poses. Following the deadline, D3R scripts evaluate the submitted predictions, send the evaluation results to each participant, and add the results to running statistics available online (http://drugdesigndata.org/about/celpp).

Thus, CELPP presents what is known as a *cross-docking* challenge, where the protein structure into which the ligand is docked was previously determined either with a different ligand or with no ligand at all (Kumar and Zhang, 2017; Shamsara, 2016; Yuriev et al., 2015). This may be contrasted with the *self-docking* problem, in which the ligand is docked back into the protein structure resolved in complex with the same ligand. Cross-docking is typically more difficult, because the protein structure has not adapted to the ligand being docked and may be in a conformation that the ligand does not fit well. However, cross-docking is a more important problem than self-docking, because it models the real-world applications of docking methods, where the entire purpose of docking is to avoid having to determine cocrystal structures for every ligand of interest in a drug discovery project.

The success of a docking calculation depends not only on the algorithm itself, but also on other methods and parameters in the overall workflow. For example, in cross-docking, one of the most important decisions is what existing structure of the protein to use in the calculation (Gaieb et al., 2017; Gathiaka et al., 2016). Additional issues arise in the preparation of the protein and ligand structures for docking. For the protein, it is often necessary to decide how to resolve ambiguous electron density, whether to remove or retain crystallographic waters, how to account for possible crystallization artifacts resulting from crystal contacts and non-physiologic temperatures, whether to select alternate side-chain conformations, and how to set the protonation states of titratable residues (Alonso et al., 2006; Corbeil et al., 2012; Dolinsky et al., 2007; Forli, 2015; Sastry et al., 2013). For the ligand, issues may include assignment of protonation and tautomer states, and the conformations of flexible rings (Ebejer et al., 2012; Irwin and Shoichet, 2005; Sastry et al., 2013).

3.2 Scale and Character of the CELPP Challenge

During a 66-week period spanning parts of 2017 and 2018, 1,989 targets met the CELPP criteria (STAR Methods: Selection of target complexes and receptor structures) and were submitted to outside participants and our in-house workflows (STAR Methods: Baseline Docking Workflows). To permit meaningful comparisons and reduce the number of test cases with unrecognized problems, such as an artifactual binding site at a crystal-induced interface, the statistic here only includes targets for which at least 3 workflows submitted predictions in the LMCSS category, and at least one LMCSS prediction achieved an RMSD under 8 Å (Table S1).

Future analyses will allow more sophisticated selections and comparisons. Most weeks saw 20–50 targets, and the maximum number of targets per week has remained below 100 (Figure 2, top). The ligands to be docked had an average of 27 heavy atoms and 5 rotatable bonds. The distributions of these descriptors are provided in the bottom panel of Figure 2 and the full list of PDB IDs and ligand SMILES strings is provided in the SI. In some cases, the selection criteria for the candidate categories yielded the same PDB structure in different categories for a target complex. For example, the PDB structure with the largest maximum common substructure (LMCSS) to the target ligand may also be the one with the highest Tanimoto similarity index (hiTanimoto). The frequency of these overlaps is shown in Figure S1. Because apo structures are not always available, there are fewer candidate structures in that category.

3.3 Pose Prediction Performance to Date

The performance records of three anonymous early-adopting external participants (one of them submitting the results from two distinct workflows) and the four in-house workflows over a 66-week period spanning 2017 and 2018 enable informative and illustrative analyses. Previous studies of pose prediction have generally considered a ligand RMSD within 2 Å of the crystal pose to be useful for compound design (Gaieb et al., 2017; Gathiaka et al., 2016;

Trott and Olson, 2009; Warren et al., 2006). In the CELPP dataset, the median prediction RMSD for the best-case prediction categories (LMCSS and hiTanimoto) is around 5 Å (Figure 3, middle). In these categories at best 20% of pose predictions are accurate to within 2 Å RMSD, and only about 40% are accurate to within 4 Å RMSD (Figure S2). These rates are significantly worse than those seen in a prior blinded challenge (Warren et al., 2006), where about 34% of the top scoring poses generated by various docking codes had RMSDs less than 2 Å, when averaged across all protein targets. However, it is important to note that all participants in the prior study were provided with receptor structures hand-picked and prepared by human experts to accommodate the ligands to be docked. In contrast, CELPP receptors are selected automatically and are not prepared by system experts. The CELPP success rates are on par with those in the recent D3R Grand Challenge 3, which yielded a corresponding success rate of 16% (Gaieb et al., 2019). As in CELPP, Grand Challenge participants are not provided with expertly selected and prepared receptor structures. More broadly, it is worth noting that the performance of the fully automated workflows tested here may not be reflective of the best performance the algorithms can provide, particularly when parameters and procedures are tuned to a specific target and/or series of ligands.

Further analysis reveals that most methods provide rather similar levels of accuracy, both for the full set of candidate structures (Figure 3, top left) and for the LMCSS set (Figure 3, top right), based on median RMSD, with rDock and one External Participant performing noticeably worse. Focusing on the performance of various methods when provided high-similarity structures (LMCSS and hiTanimoto categories; see next paragraph), we observe that two anonymous public participants performed slightly better than other methods, as measured by the fraction of RMSDs in the 0–2 Å range (Figure S2). The in-house OE Fred and Vina workflows yield rather similar results, with GLIDE variants and rDock trailing slightly by this metric. As noted above, the in-house workflows are not tuned for optimum results and thus may not reflect the best performance these algorithms can provide.

The CELPP data also allows quantification of important trends that have been previously noted (Gaieb et al., 2017; Gathiaka et al., 2016). First, docking into a receptor determined with a chemically similar ligand, as determined by the maximum common substructure (LMCSS) or the fingerprint Tanimoto similarity index (hiTanimoto), more than doubles the success rate (RMSD < 2Å), relative to docking into a structure determined without a bound ligand (hiResApo) (Figure S2). Docking into the highest resolution structure solved with any ligand (hiResHolo) and into the structure with the least similar ligand, based on maximum common substructure (SMCSS), yielded results of intermediate accuracy. Similar results are observed for the individual methods (Figure S2). Second, docking results tend to be less accurate for ligands with more rotatable bonds, but this challenge can be overcome by docking into a protein structure determined with a highly similar ligand (Figure 3, bottom). Best docking results are, therefore, obtained either when the number of rotatable bonds is less than 2 or when the fraction of the heavy atoms in the target ligand that are in its maximal common substructure with the candidate ligand (the MCSS ratio) is above 0.8. The worst results are obtained for target ligands with >10 rotatable bonds and an MCSS ratio lower than about 0.5. In the best-case scenarios, where the candidate structure has 0 or 1 rotatable bonds and an MCSS ratio of at least 0.8, automated docking workflows can achieve a median RMSD of around 3 Å. In more difficult cases, with MCSS ratios between 0.4 and 0.5 and 7 rotatable bonds, the median RMSD rises to about 6 Å.

The "standard" target, 1FCZ, is included in the challenge package each week. As 1FCZ is an existing PDB structure, the LMCSS category poses a self-docking challenge where the correct ligand pose is publicly known (results from 1FCZ are excluded from the general dataset). All workflows regularly achieve RMSD values below 1 Å in the LMCSS category for 1FCZ (Figure S3). Three internal and two external workflows behave deterministically on this target, returning the same poses each week. One internal workflow (rDock) and two external workflows return different poses each week. These inconsistent results indicate a potential source of uncertainty for method comparison. The Glide workflow implemented in CELPP does not produce a prediction for 1FCZ, as the size of the docking region used for all methods in this study is smaller than its recommended value. However, when run with its default size docking box, the GLIDE workflow consistently produces a pose with an RMSD of 0.4 Å.

Directions to download the dataset analyzed in this paper are available in the Supporting Information (Wagner et al., 2019).

4 Discussion

The CELPP challenge is a powerful tool to evaluate and improve protein-ligand pose prediction technologies. Unlike prior blinded prediction challenges in this field, CELPP sets a new challenge each week, each with dozens of new ligand-protein complexes to model and provides rapid and consistent feedback for participants. The >1,900 individual challenge cases set by CELPP in one 66-week period far exceeds the number of cases set by all prior blind pose-prediction challenges, and the CELPP challenge is ongoing. We invite additional researchers to become participants in this new blinded challenge and utilize it to help advance their pose-prediction technologies.

The high-throughput nature of CELPP dramatically increases the statistical power of the results and enables sharper distinctions among methods. We anticipate new insights into not only the core docking algorithms, but also procedural details, such as how crystallographic water molecules and protonation states are treated. We also plan to look for characteristics of protein targets and ligands that correlate with the performance of specific methods. For example, some algorithms may do better for hydrophobic sites or for specific protein families, such as serine proteases. Such statistical analyses will be posted on the CELPP website to help practitioners choose methods suited for their specific applications and set meaningful expectations for the quality of predictions on new systems. We will also scan for cases where multiple methods do poorly, checking for situations in which CELPP's automated procedures may generate inappropriate challenges, such as where a cofactor is present in the candidate but not in the target. The volume and tempo of the CELPP challenge also allow its use in iterative optimization of pose prediction methods. Thus, we anticipate that CELPP will help drive the development of increasingly predictive docking workflows.

A possible concern with CELPP is that it may not reveal the accuracy that could be achieved with a given docking method by an expert user devoting time to optimizing the results for a protein target. However, only a fully automated procedure reports on the accuracy of the method itself, as opposed to the expertise of the user, and our aim is to test the method. In addition, one may envision ways of building expertise and background information into an automated method. For example, a docking method could be trained to recognize that a protein target is a kinase and therefore to use kinase-specific settings.

It is also worth noting that, if a participant's pose-prediction method were to change over time, it would become impossible to collect full statistics for a single, defined method. It will therefore be useful to distinguish between those methods which are locked in for a period of time, and for which meaningful statistics therefore can be obtained, from those which are mutable, such as when CELPP is used to guide ongoing improvements in a pose-prediction method.

Ultimately, we anticipate that participants will package their stable methods into shareable workflows, using technologies such as Docker (Merkel, 2014) and Singularity (Kurtzer et al., 2017), which can then be executed automatically on machines hosted by the CELPP project, and also shared with other users. Participants also would benefit by not having to manage the processes or allocate computer time for the calculations. Ideally, such workflows will be structured into modular steps with standardized I/O, enabling the creation and benchmarking of new strategies that recombine steps from various workflows. Such derivatization of workflows could, for example, make it possible to evaluate how various ligand-preparation methods affect the quality of the final pose predictions. This direction promises to build a beneficial culture of generating methods that are rigorously evaluated through ongoing blinded challenges and that are readily shared, so that effective methods can easily be used.

9 STAR Methods

Method Details

Hosting the Challenge

Selection of target complexes and receptor structures: Every Friday, the PDB provides files (http://www.wwpdb.org/files/) listing the new crystal structures that will be released the at the end of the following Tuesday. For each forthcoming PDB entry, this pre-release notification contains the PDB ID, the protein sequence(s), the identities of the ligands, if any, in the form of InChi strings (Heller et al., 2015), and the pH at which the structure was determined. To be designated as a CELPP "target", a structure must include a single druglike ligand (see below). There also must be at least one X-ray crystal structure of the same protein extant in the PDB to serve as a suitable "candidate" for cross-docking, and the target must have only one unique protein sequence, to avoid situations in which the target and candidate ligands are bound to different binding sites (Figure 4). A ligand is considered drug-like if it is not a single metal ion or typical solvent molecule, and if it is not on an exclusion list of common cosolutes and cofactors (e.g., Zn^{++}, ethylene glycol, and NADH; see Scheme S3). We also exclude ligands with more than 100 self-symmetries (i.e., automorphs), because evaluating symmetry-corrected root-mean-square deviations between predicted and crystal poses becomes excessively time-consuming in such cases (OpenEye Scientific Software). Finally, we include one “standard” target, PDB ID 1FCZ, in the challenge package each week to monitor workflow stability.

Candidate structures for use in the docking challenge are identified by using the sequence comparison program blastp (Altschul et al., 1990) to find PDB entries with >95% sequence identity and >90% sequence coverage of the target sequence (Figure 4). The resulting proteins are then further filtered to a set that were determined by X-ray crystallography (rather than NMR, for example) and which comprise only a single unique protein sequence. Any ligands bound to these candidates have their maximal common substructure with the target ligand calculated using RDKit. For each target, up to five candidate structures meeting these criteria are selected from the PDB for use as receptors in the cross-docking challenge. The five candidates, which are chosen to test the effects of various criteria for selecting the receptor used in cross-docking, are as follows: **(i) Largest Maximum Common Substructure (LMCSS):** This contains the ligand with the largest maximal common substructure to the target ligand. The center of mass of the ligand in this complex is used to define the binding pocket for all five candidates for this target. If two candidate complexes tie for the largest maximal common substructure, the highest-resolution one is used. **(ii) Smallest Maximum Common Substructure (SMCSS):** This contains the ligand with the smallest maximal common substructure to the target ligand. If two candidate complexes tie for the smallest maximal common substructure, the highest-resolution one is used. **(iii) Highest Tanimoto Similarity (hiTanimoto):** This contains the ligand with the highest ligand Tanimoto score, using the RDKit default fingerprint method, to the target ligand. If two candidate complexes tie for the highest mutual Tanimoto score, the highest-resolution one is used. This candidate may be the same as the LMCSS candidate. **(iv) Highest Resolution Holo (hiResHolo):** This has the highest crystallographic resolution limit of any determined with a druglike ligand. **(v) Highest Resolution Apo (hiResApo):** This has the highest crystallographic resolution limit of any for this protein determined with no druglike ligand.

Generation of the challenge data package: The results of the processing described in STAR Methods: Selection of target complexes and receptor structures are incorporated into a data package for use by CELPP participants. This is deposited in a public Box.com folder by about 00:00 Pacific Time every Sunday (Figure 1). For each target, the challenge data package (Figure 4) contains: (i) Structures of the candidate proteins in PDB format, aligned to the LMCSS structure coordinates but otherwise unmodified from the PDB entries from which they were drawn; (ii) PDB structure of the LMCSS ligand, drawn from the LMCSS candidate structure; (iii) The suggested binding pocket center (center of mass of the LMCSS ligand) in .txt format; (iv) SMILES, InChI, and 2D MOL files of the target ligand; (v) A parseable text file containing the PDB ID of the forthcoming entry, the crystallization pH, the HETID of the target ligand, and additional information about the candidate cocrystal structures, such as their PDB IDs and crystallographic resolution limits. See Scheme S1 for sample.

Evaluation of Predictions: After the close of the submission window and release of the experimental cocrystal structure by the PDB (Figure 1), automated scripts evaluate the pose predictions by calculating the symmetry-corrected RMSD of each predicted ligand pose relative to the crystallographic pose, using Schrödinger and OpenEye tools (Scheme S2). When the crystal structure has multiple instances (protein chains) of the target protein, the predicted pose is assigned the lowest RMSD that can be achieved by aligning the predicted protein-ligand complex to each instance of the chain, as detailed in Scheme S2. In the future, we anticipate incorporating additional evaluation metrics, such as the fraction of native ligand-protein contacts made by a predicted pose. Such additions will be facilitated by the modular architecture of the CELPP software.

Participating in CELPP

Enrollment and Information: In order to obtain upload/download credentials for CELPP data, one must register as a CELPP participant. Information for how to participate in CELPP is available at the D3R website (https://drugdesigndata.org/about/celpp), including links to a CELPP Developers User Group on Google Groups and to the CELPP GitHub Wiki.

Developing a Prediction Workflow: Based on typical throughput (Results: Scale and character of the CELPP challenge), CELPP participants should construct a pose-prediction workflow that can process up to 100 targets in the 63-hour submission window (Figure 2, top). This requires executing up to 100 ligand preparation tasks, 400–500 protein preparation tasks, and 400–500 docking tasks. Participants are encouraged to submit predictions for all targets, so that methods can be compared on an equal footing. To help participants construct their workflows, D3R provides a workflow template, called CELPPade (Figure 5, https://github.com/drugdata/cookiecutter-pycustomdock). CELPPade handles the download, unpackaging, repackaging and upload of the CELPP challenge data, letting the participant focus on implementing their docking solution by writing a few specific methods in Python.

The five user-written Python functions in CELPPade are located in three files (Figure 5, protein_prep.py, ligand_prep.py, and dock.py). These five functions mirror the stages in a general pose prediction workflow: protein and ligand structure preparation, protein and ligand file format conversion, and docking. Each function receives input filenames as arguments, and participants are responsible for populating the function bodies with commands to run the respective stage of their workflow. Participants can implement their docking workflow in these functions using Python commands or using Python's interfaces to the command line to execute shell commands. Once implemented, CELPPade is able to run these functions in sequence on each prediction target in the current CELPP challenge week. Combined with the download and upload scripts bundled in CELPPade, implementing these five functions results in a functioning CELPP workflow.

We also provide a tutorial for creation of a model workflow based on CELPPade. This uses Chimera DockPrep (Pettersen Eric et al., 2004) to prepare both the protein and ligand, and AutoDock for pose prediction (Lang et al., 2009; Moustakas et al., 2006; Trott and Olson, 2009). The model workflow provides examples of running shell commands from within Python, uses software that is free for use by academic labs, and can run on most 64-bit

Linux systems with Python, Chimera, RDKit, OpenBabel, and Autodock Vina installed (Lang et al., 2009; Moustakas et al., 2006; O'Boyle et al., 2011; Pettersen Eric et al., 2004). Note that participants are not required to use the CELPPade template; it is provided only as a convenience.

Even if a participant does not use the full CELPPade package, two helper scripts it contains may be of interest. The first, **getchallengedata.py**, helps participants access the correct challenge data package each week. It reads the Box.com login credentials of participants from the user's customized file ftp_config and uses these to access the online folder. It then reads the file latest.txt file in the Box.com folder to determine the name of the most recent challenge package, downloads the package, and unzips it in the user's local folder. The last script, **packdockingresults.py**, takes as input a formatted directory of docking results generated by the participant, compresses the directory into a tar file, and uploads the tar file to the participant's private submission folder. These scripts facilitate automation by providing straightforward upload/download functionality for CELPP data, independent of the specific details of the prediction workflow.

Definition and Submission of Predictions: For each pose prediction, for a given candidate structure, a valid submission comprises the receptor structure in PDB format and the ligand structure in MOL format (Dalby et al., 1992; McNaught and Wilkinson, 1997), with ligand coordinates in the receptor frame of reference. Participants need not use the candidate receptor structures provided by D3R for docking but must inform us if they do not. Deviation from these file formats may result in improper evaluation or disqualification of a prediction. Using the CELPPade workflow template ensures that the docking results directory is appropriately formatted for upload. Pose predictions are uploaded to an online, password-protected Box.com folder provided by D3R. This upload must be completed before 15:00 U.S. Pacific time on Tuesday to be considered valid for scoring.

Evaluation Reporting: Detailed evaluation results are emailed directly to participants and will soon be publicly accessible at the CELPP website. Participants may choose to remain anonymous, in which case their methods and results will be posted without identifying information.

Baseline docking workflows—To illustrate CELPP, test our procedures, and provide a baseline of performance, we used the CELPPade template (STAR Methods: Developing a prediction workflow) to create four in-house CELPP workflows, based on the Autodock Vina (Trott and Olson, 2009), FRED (McGann, 2011, 2012), Glide (Friesner et al., 2004; Halgren et al., 2004), and rDock (Morley et al., 2004; Ruiz-Carmona et al., 2014) docking suites. These use the protein and ligand preparation tools that accompany the respective docking codes where possible, and open-source tools otherwise. These workflows represent default implementations of their respective methods. To standardize the testing, all workflows are set to consider the same docking region. Algorithms that require a docking box are set to use a 15×15×15 Å region, and software that requires a sphere is set to use a 10 Å radius region. As we have made no effort to optimize the workflows, their performance may not be reflective of the best performance the algorithms can provide, particularly when tuned for a specific protein target and/or ligand series. All the in-house prediction workflows are available on GitHub. In addition, the AutoDock Vina workflow has been documented in a tutorial on CELPP workflow development. This can be accessed on the D3R GitHub page, as noted in STAR Methods: Developing a prediction workflow.

Quantification and Statistical Analysis—Prediction RMSDs were evaluated using the Schrodinger and OpenEye toolkits, first by aligning the target and predicted complexes, and then by calculating the symmetry-corrected ligand heavy atom RMSD. More information on the evaluation process is available in STAR Methods: Evaluation of Predictions, Scheme S2, and the evaluation codebase is available at the D3R GitHub repository.

Python's numpy and matplotlib packages were used for numerical analysis and plotting. Means and medians in Figure 3 were calculated using matplotlib for the violin plots, and numpy for the heatmap.

For each ligand-protein pose prediction, we compute the symmetry-corrected RMSD of the ligand's predicted coordinates versus its crystallographic coordinates. This is a widely used and understood metric in the field. We use the median across the full set of predictions to represent the overall performance in a manner that is insensitive to variations in individual RMSD values for large outliers. For example, we regard two predictions with RMSD of 9 Å and 7 Å as having equally failed to find the correct pose, so it is not useful to penalize one more than the other in our overall measure of performance.

Violin-style plots were selected in Figure 3 to provide a complete picture of the performance distribution of each method and facilitate a fair comparison of methods. The same data are plotted in finer detail in Figure S2 as cumulative histograms with bin sizes of 0.01 Angstrom. The actual dataset including RMSD values is available for download, as described in the Data and Software Availability section.

The number of predictions analyzed in each panel of Figure 3 is as follows, where one prediction corresponds to one protein-ligand complex, with its own RMSD: Top left - Autodock Vina: 7550, Glide: 4148, OE Fred: 8649, rDock: 8073, First External Participant: 7997, Second External Participant: 4614, Third External Participant: 1896, Fourth External Participant: 1624. Top right - Autodock Vina: 1655, Glide: 932, OE Fred: 1888, rDock: 1761, First External Participant: 1746, Second External Participant: 1007, Third External Participant: 417, Fourth External Participant: 355 Middle - LMCSS: 9761, SMCSS: 9674, hiResApo: 5610, hiResHolo: 9711, hiTanimoto: 9759 Bottom - 19435

Table S1 contains a more detailed breakdown of the number of targets analyzed in this paper, separated by workflow and candidate category.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We gratefully acknowledge the early CELPP participants, who provided feedback and developed automated workflows before the challenge was made public. We thank Torsten Schwede and Jürgen Haas for helpful discussions and for developing the inspirational CAMEO challenge. D3R is supported by NIH grant U01 GM111528 to REA and MKG. The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

10 References

. OEDOCKING (Santa Fe, NM: OpenEye Scientific Software; eyesopen.com).

. RDKit: Open-source cheminformatics (rdkit.org).

. Schrödinger Release 2015–3: Glide, Schrödinger, LLC, New York, NY, 2015.

. Schrödinger Release 2015–3: Schrödinger Suite 2015–3 Protein Preparation Wizard;

Epik, Schrödinger, LLC, New York, NY, 2015; Impact, Schrödinger, LLC, New York, NY, 2015;

Prime, Schrödinger, LLC, New York, NY, 2015.

. Schrödinger Release 2015–3: LigPrep, Schrödinger, LLC, New York, NY, 2015.

Abel R, Wang L, Mobley DL, and Friesner RA (2017). A Critical Review of Validation, Blind Testing, and Real-World Use of Alchemical Protein-Ligand Binding Free Energy Calculations. Curr Top Med Chem 17.

Alonso H, Bliznyuk Andrey A, and Gready Jill E (2006). Combining docking and molecular dynamic simulations in drug design. Medicinal Research Reviews 26, 531–568. [PubMed: 16758486]

Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ (1990). Basic local alignment search tool. Journal of Molecular Biology 215, 403–410. [PubMed: 2231712]

Amaro RE, Baron R, and Andrew McCammon J (2008). An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J Comput Aided Mol Des 22, 693–705. [PubMed: 18196463]

Amaro RE, Baudry J, Chodera J, Demir Ö, McCammon JA, Miao Y, and Smith JC (2018). Ensemble Docking in Drug Discovery. Biophysical Journal 114, 2271–2278. [PubMed: 29606412]

Anna EL, Stephan ML, and Billy Williams-Noonan SS (2017). A Practical Guide to Molecular Docking and Homology Modelling for Medicinal Chemists. Current Topics in Medicinal Chemistry 17, 2023–2040. [PubMed: 28137238]

Balasubramanian V, Bethune I, Shkurti A, Breitmoser E, Hruska E, Clementi C, Laughton C, and Jha S (2016). ExTASY: Scalable and flexible coupling of MD simulations and advanced sampling techniques. Paper presented at: 2016 IEEE 12th International Conference on e-Science (e-Science).

Bartuzi D, Kaczor AA, Targowska-Duda MK, and Matosiuk D (2017). Recent Advances and Applications of Molecular Docking to G Protein-Coupled Receptors. Molecules 22.

Beck KR, Kaserer T, Schuster D, and Odermatt A (2017). Virtual screening applications in short-chain dehydrogenase/reductase research. The Journal of Steroid Biochemistry and Molecular Biology 171, 157–177. [PubMed: 28286207]

Beisken S, Meinl T, Wiswedel B, de Figueiredo LF, Berthold M, and Steinbeck C (2013). KNIME-CDK: Workflow-driven cheminformatics. BMC Bioinformatics 14, 257. [PubMed: 24103053]

Berman H, Henrick K, and Nakamura H (2003). Announcing the worldwide Protein Data Bank. Nat Struct Biol 10, 980. [PubMed: 14634627]

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, and Bourne PE (2000). The Protein Data Bank. Nucleic Acids Res 28, 235–242. [PubMed: 10592235]

Burley SK, Berman HM, Bhikadiya C, Bi C, Chen L, Di Costanzo L, Christie C, Dalenberg K, Duarte JM, Dutta S, et al. (2018). RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. Nucleic Acids Research, gky1004-gky1004.

Carlson HA (2016). Lessons Learned over Four Benchmark Exercises from the Community Structure-Activity Resource. J Chem Inf Model 56, 951–954. [PubMed: 27345761]

Carlson HA, Smith RD, Damm-Ganamet KL, Stuckey JA, Ahmed A, Convery MA, Somers DO, Kranz M, Elkins PA, Cui G, et al. (2016). CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. J Chem Inf Model 56, 1063–1077. [PubMed: 27149958]

Chen R, Li L, and Weng Z (2003). ZDOCK: An initial-stage protein-docking algorithm. Proteins: Structure, Function, and Bioinformatics 52, 80–87.

Cheng T, Li Q, Zhou Z, Wang Y, and Bryant SH (2012). Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. AAPS J 14, 133. [PubMed: 22281989]

Ciancetta A, and Jacobson AK (2017). Structural Probing and Molecular Modeling of the A3 Adenosine Receptor: A Focus on Agonist Binding. Molecules 22.

Cole JC, Murray CW, Nissink JWM, Taylor RD, and Taylor R (2005). Comparing protein–ligand docking programs is difficult. Proteins: Structure, Function, and Bioinformatics 60, 325–332.

Corbeil CR, Williams CI, and Labute P (2012). Variability in docking success rates due to dataset preparation. Journal of Computer-Aided Molecular Design 26, 775–786. [PubMed: 22566074]

Dakka J, Turilli M, Wright DW, Zasada SJ, Balasubramanian V, Wan S, Coveney PV, and Jha S (2018). High-throughput Binding Affinity Calculations at Extreme Scales arXiv:171209168v4 [csDC].

Dalby A, Nourse JG, Hounshell WD, Gushurst AKI, Grier DL, Leland BA, and Laufer J (1992). Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. Journal of Chemical Information and Computer Sciences 32, 244–255.

Damm-Ganamet KL, Smith RD, Dunbar JB Jr., Stuckey JA, and Carlson HA (2013). CSAR benchmark exercise 2011–2012: evaluation of results from docking and relative ranking of blinded congeneric series. J Chem Inf Model 53, 1853–1870. [PubMed: 23548044]

Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, and Baker NA (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. Nucleic Acids Res 35, W522–525. [PubMed: 17488841]

Doman TN, McGovern SL, Witherbee BJ, Kasten TP, Kurumbail R, Stallings WC, Connolly DT, and Shoichet BK (2002). Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. J Med Chem 45, 2213–2221. [PubMed: 12014959]

Ebejer J-P, Morris GM, and Deane CM (2012). Freely available conformer generation methods: how good are they? J Chem Inf Model 52, 1146–1158. [PubMed: 22482737]

Ewing TJA, and Kuntz ID (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem 18, 1175–1189.

Forli S (2015). Charting a Path to Success in Virtual Screening. Molecules 20.

Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, et al. (2004). Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. J Med Chem 47, 1739–1749. [PubMed: 15027865]

Gaieb Z, Liu S, Gathiaka S, Chiu M, Yang H, Shao C, Feher VA, Patrick Walters W, Kuhn B, Rudolph MG, et al. (2017). D3R Grand Challenge 2: blind prediction of protein– ligand poses, affinity rankings, and relative binding free energies. J Comput Aided Mol Des, 1–20.

Gaieb Z, Parks CD, Chiu M, Yang H, Shao C, Walters WP, Lambert MH, Nevins N, Bembenek SD, Ameriks MK, et al. (2019). D3R Grand Challenge 3: blind prediction of protein–ligand poses and affinity rankings. Journal of Computer-Aided Molecular Design 33, 1–18. [PubMed: 30632055]

Ganesan A, and Barakat K (2017). Applications of computer-aided approaches in the development of hepatitis C antiviral agents. Expert Opinion on Drug Discovery 12, 407–425. [PubMed: 28164720]

Gathiaka S, Liu S, Chiu M, Yang H, Stuckey JA, Kang YN, Delproposto J, Kubish G, Dunbar JB Jr., Carlson HA, et al. (2016). D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity predictions. J Comput Aided Mol Des 30, 651–668. [PubMed: 27696240]

Goodsell DS, Morris GM, and Olson AJ (1996). Automated docking of flexible ligands: applications of AutoDock. J Mol Recognit 9, 1–5. [PubMed: 8723313]

Goodsell DS, and Olson AJ (1990). Automated docking of substrates to proteins by simulated annealing. Proteins 8, 195–202. [PubMed: 2281083]

Grinter SZ, and Zou X (2014). Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. Molecules 19, 10150–10176. [PubMed: 25019558]

Guedes IA, de Magalhães CS, and Dardenne LE (2014). Receptor-ligand molecular docking. Biophys Rev 6, 75–87. [PubMed: 28509958]

Haas J, Barbato A, Behringer D, Studer G, Roth S, Bertoni M, Mostaguir K, Gumienny R, and Schwede T (2017). Continuous Automated Model Evaluation (CAMEO) Complementing the Critical Assessment of Structure Prediction in CASP12. Proteins

Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, and Banks JL (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 47, 1750–1759. [PubMed: 15027866]

Hawkins PCD, Skillman AG, Warren GL, Ellingson BA, and Stahl MT (2010). Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. J Chem Inf Model 50, 572–584. [PubMed: 20235588]

Heller SR, McNaught A, Pletnev I, Stein S, and Tchekhovskoi D (2015). InChI, the IUPAC International Chemical Identifier. Journal of Cheminformatics 7, 23. [PubMed: 26136848]

Hodos RA, Kidd BA, Shameer K, Readhead BP, and Dudley JT (2016). In silico methods for drug repurposing and pharmacology. Wiley Interdisciplinary Reviews: Systems Biology and Medicine 8, 186–210. [PubMed: 27080087]

Irwin JJ (2008). Community benchmarks for virtual screening. Journal of Computer-Aided Molecular Design 22, 193–199. [PubMed: 18273555]

Irwin JJ, and Shoichet BK (2005). ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. J Chem Inf Model 45, 177–182. [PubMed: 15667143]

Jain AN (2008). Bias, reporting, and sharing: computational evaluations of docking methods. J Comput Aided Mol Des 22, 201–212. [PubMed: 18075713]

Jansen JM, Cornell W, Tseng YJ, and Amaro RE (2012). Teach–Discover–Treat (TDT): Collaborative computational drug discovery for neglected diseases. Journal of Molecular Graphics and Modelling 38, 360–362. [PubMed: 23085175]

Kim J, Yang G, and Ha J (2017). Targeting of AMP-activated protein kinase: prospects for computer-aided drug design. Expert Opinion on Drug Discovery 12, 47–59. [PubMed: 27797589]

Knegtel RM, Kuntz ID, and Oshiro CM (1997). Molecular docking to ensembles of protein structures. J Mol Biol 266, 424–440. [PubMed: 9047373]

Kufareva I, Katritch V, Stevens RC, and Abagyan R (2014). Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. Structure 22, 1120–1139. [PubMed: 25066135]

Kufareva I, Rueda M, Katritch V, Stevens RC, and Abagyan R (2011). Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment. Structure 19, 1108–1126. [PubMed: 21827947]

Kuhn B, Guba W, Hert J, Banner D, Bissantz C, Ceccarelli S, Haap W, Körner M, Kuglstatter A, Lerner C, et al. (2016). A Real-World Perspective on Molecular Design. J Med Chem 59, 4087–4102. [PubMed: 26878596]

Kumar A, and Zhang KYJ (2017). A cross docking pipeline for improving pose prediction and virtual screening performance. J Comput Aided Mol Des

Kuntz ID (1992). Structure-based strategies for drug design and discovery. Science 257, 1078–1082. [PubMed: 1509259]

Kurtzer GM, Sochat V, and Bauer MW (2017). Singularity: Scientific containers for mobility of compute. PLOS ONE 12, e0177459. [PubMed: 28494014]

Lang PT, Brozell SR, Mukherjee S, Pettersen EF, Meng EC, Thomas V, Rizzo RC, Case DA, James TL, and Kuntz ID (2009). DOCK 6: Combining techniques to model RNA–small molecule complexes. RNA 15, 1219–1230. [PubMed: 19369428]

Leelananda SP, and Lindert S (2016). Computational methods in drug discovery. Beilstein Journal of Organic Chemistry 12, 2694–2718. [PubMed: 28144341]

Li L, Jiang S, Li X, Liu Y, Su J, and Chen J (2018). Recent advances in trimethoxyphenyl (TMP) based tubulin inhibitors targeting the colchicine binding site. European Journal of Medicinal Chemistry 151, 482–494. [PubMed: 29649743]

Lybrand TP (1995). Ligand—protein docking and rational drug design. Current Opinion in Structural Biology 5, 224–228. [PubMed: 7648325]

Mafud AC, Ferreira LG, Mascarenhas YP, Andricopulo AD, and de Moraes J (2016). Discovery of Novel Antischistosomal Agents by Molecular Modeling Approaches. Trends in Parasitology 32, 874–886. [PubMed: 27593339]

Martinez-Mayorga K, Byler KG, Ramirez-Hernandez AI, and Terrazas-Alvares DE (2015). Cruzain inhibitors: efforts made, current leads and a structural outlook of new hits. Drug Discovery Today 20, 890–898. [PubMed: 25697479]

McGann M (2011). FRED Pose Prediction and Virtual Screening Accuracy. J Chem Inf Model 51, 578–596. [PubMed: 21323318]

McGann M (2012). FRED and HYBRID docking performance on standardized datasets. J Comput Aided Mol Des 26, 897–906. [PubMed: 22669221]

McGann MR, Almond HR, Nicholls A, Grant JA, and Brown FK (2002). Gaussian docking functions. Biopolymers 68, 76–90.

McNaught AD, and Wilkinson A (1997). IUPAC. Compendium of Chemical Terminology, 2nd ed. (the “Gold Book”). In XML on-line corrected version: http://goldbookiupacorg (2006-) created by M Nic, J Jirat, B Kosata; updates compiled by A Jenkins (Oxford: Blackwell Scientific Publications).

Medina-Franco JL, Méndez-Lucio O, Dueñas-González A, and Yoo J (2015). Discovery and development of DNA methyltransferase inhibitors using in silico approaches. Drug Discovery Today 20, 569–577. [PubMed: 25526932]

Merkel D (2014). Docker: lightweight Linux containers for consistent development and deployment. Linux J 2014, 2.

Michino M, Abola E, Brooks CL, Scott Dixon J, Moult J, and Stevens RC (2009). Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. Nat Rev Drug Discov 8, 455–463. [PubMed: 19461661]

Mobley DL, and Dill KA (2009). Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”. Structure 17, 489–498. [PubMed: 19368882]

Morgnanesi D, Heinrichs EJ, Mele AR, Wilkinson S, Zhou S, and Kulp JL (2015). A computational chemistry perspective on the current status and future direction of hepatitis B antiviral drug discovery. Antiviral Research 123, 204–215. [PubMed: 26477294]

Morley SD, David Morley S, and Afshar M (2004). Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock®. J Comput Aided Mol Des 18, 189–208. [PubMed: 15368919]

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, and Olson AJ (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry 30, 2785–2791. [PubMed: 19399780]

Moustakas DT, Lang PT, Pegg S, Pettersen E, Kuntz ID, Brooijmans N, and Rizzo RC (2006). Development and validation of a modular, extensible docking program: DOCK 5. Journal of Computer-Aided Molecular Design 20, 601–619. [PubMed: 17149653]

Muddana HS, Daniel Varnado C, Bielawski CW, Urbach AR, Isaacs L, Geballe MT, and Gilson MK (2012). Blind prediction of host–guest binding affinities: a new SAMPL3 challenge. Journal of Computer-Aided Molecular Design 26, 475–487. [PubMed: 22366955]

Muddana HS, Fenley AT, Mobley DL, and Gilson MK (2014). The SAMPL4 host–guest blind prediction challenge: an overview. Journal of Computer-Aided Molecular Design 28, 305–317. [PubMed: 24599514]

Neves MAC, Totrov M, and Abagyan R (2012). Docking and scoring with ICM: the benchmarking results and strategies for improvement. Journal of Computer-Aided Molecular Design 26, 675–686. [PubMed: 22569591]

O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, and Hutchison GR (2011). Open Babel: An open chemical toolbox. Journal of Cheminformatics 3, 33. [PubMed: 21982300]

Ogungbe VI, and Setzer NW (2016). The Potential of Secondary Metabolites from Plants as Drugs or Leads against Protozoan Neglected Diseases—Part III: In-Silico Molecular Docking Investigations. Molecules 21.

OpenEye Scientific Software, I. OERMSD -- Toolkits -- Python.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera—A visualization system for exploratory research and analysis. Journal of Computational Chemistry 25, 1605–1612. [PubMed: 15264254]

Pettersen Eric F, Goddard Thomas D, Huang Conrad C, Couch Gregory S, Greenblatt Daniel M, Meng Elaine C, and Ferrin Thomas E (2004). UCSF Chimera—A visualization system for exploratory research and analysis. Journal of Computational Chemistry 25, 1605–1612. [PubMed: 15264254]

Pronk S, Pouya I, Lundborg M, Rotskoff G, Wesén B, Kasson PM, and Lindahl E (2015). Molecular Simulation Workflows as Parallel Algorithms: The Execution Engine of Copernicus, a Distributed High-Performance Computing Platform. Journal of Chemical Theory and Computation 11, 2600–2608. [PubMed: 26575558]

Purawat S, Ieong PU, Malmstrom RD, Chan GJ, Yeung AK, Walker RC, Altintas I, and Amaro RE (2017). A Kepler Workflow Tool for Reproducible AMBER GPU Molecular Dynamics. Biophysical Journal 112, 2469–2474. [PubMed: 28636905]

Raghavendra NM, Pingili D, Kadasi S, Mettu A, and Prasad SVUM (2018). Dual or multi-targeting inhibitors: The next generation anticancer agents. European Journal of Medicinal Chemistry 143, 1277–1300. [PubMed: 29126724]

Rosano C, Ponassi M, Santolla MF, Pisano A, Felli L, Vivacqua A, Maggiolini M, and Lappano R (2016). Macromolecular Modelling and Docking Simulations for the Discovery of Selective GPER Ligands. The AAPS Journal 18, 41–46. [PubMed: 26573009]

Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z, et al. (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res 45, D271–D281. [PubMed: 27794042]

Rosenfeld R, Vajda S, and DeLisi C (1995). Flexible Docking and Design. Annual Review of Biophysics and Biomolecular Structure 24, 677–700.

Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Beatriz Garmendia-Doval A, Juhos S, Schmidtke P, Barril X, Hubbard RE, and David Morley S (2014). rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. PLoS Comput Biol 10, e1003571. [PubMed: 24722481]

Santiago V, Eduardo S-S, and Lourdes Santana and Eugenio, U. (2017). Molecular Docking and Drug Discovery in β-Adrenergic Receptors. Current Medicinal Chemistry 24, 4340–4359. [PubMed: 28738772]

Sastry GM, Adzhigirey M, Day T, Annabhimoju R, and Sherman W (2013). Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J Comput Aided Mol Des 27, 221–234. [PubMed: 23579614]

Shamsara J (2016). CrossDocker: a tool for performing cross-docking using Autodock Vina. Springerplus 5.

Shunmugam L, Ramharack P, and Soliman MES (2017). Road Map for the Structure-Based Design of Selective Covalent HCV NS¾A Protease Inhibitors. The Protein Journal 36, 397–406. [PubMed: 28815420]

Singh N, and Ecker G (2018). Insights into the Structure, Function, and Ligand Discovery of the Large Neutral Amino Acid Transporter 1, LAT1. International Journal of Molecular Sciences 19.

Sinko W, Lindert S, and McCammon JA (2013). Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. Chem Biol Drug Des 81, 41–49. [PubMed: 23253130]

Sledz P, and Caflisch A (2017). Protein structure-based drug design: from docking to molecular dynamics. Curr Opin Struct Biol 48, 93–102. [PubMed: 29149726]

Sliwoski G, Kothiwale S, Meiler J, and Lowe EW (2014). Computational Methods in Drug Discovery. Pharmacol Rev 66, 334–395. [PubMed: 24381236]

Smith RD, Damm-Ganamet KL, Dunbar JB Jr., Ahmed A, Chinnaswamy K, Delproposto JE, Kubish GM, Tinberg CE, Khare SD, Dou J, et al. (2016). CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. J Chem Inf Model 56, 1022–1031. [PubMed: 26419257]

Smith RD, Dunbar JB Jr., Ung PM-U, Esposito EX, Yang C-Y, Wang S, and Carlson HA (2011). CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. J Chem Inf Model 51, 2115–2131. [PubMed: 21809884]

Sousa SF, Fernandes PA, and Ramos MJ (2006). Protein–ligand docking: Current status and future challenges. Proteins: Structure, Function, and Bioinformatics 65, 15–26.

Spitzer R, and Jain AN (2012). Surflex-Dock: Docking benchmarks and real-world application. Journal of Computer-Aided Molecular Design 26, 687–699. [PubMed: 22569590]

Tan Z, Chaudhai R, and Zhang S (2016). Polypharmacology in Drug Development: A Minireview of Current Technologies. ChemMedChem 11, 1211–1218. [PubMed: 27154144]

Topiol S, and Sabio M (2015). The role of experimental and computational structural approaches in 7TM drug discovery. Expert Opinion on Drug Discovery 10, 1071–1084. [PubMed: 26211671]

Trott O, and Olson AJ (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem, NA-NA.

Wagner JR, Churas CP, Liu S, Swift R, V, Chiu M, Shao C, Crawl D, Feher VA, Burley SK, et al. (2019). Data From: Continuous Evaluation of Ligand Protein Predictions: A Weekly Community Challenge for Drug Docking (UC San Diego Library Digital Collections 10.6075/J0610XPS).

Warren GL, Andrews CW, Capelli A-M, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, et al. (2006). A critical assessment of docking programs and scoring functions. J Med Chem 49, 5912–5931. [PubMed: 17004707]

Weiss DR, Bortolato A, Tehan B, and Mason JS (2016). GPCR-Bench: A Benchmarking Set and Practitioners' Guide for G Protein-Coupled Receptor Docking. Journal of Chemical Information and Modeling 56, 642–651. [PubMed: 26958710]

Yin J, Henriksen NM, Slochower DR, Shirts MR, Chiu MW, Mobley DL, and Gilson MK (2017). Overview of the SAMPL5 host-guest challenge: Are we doing better? J Comput Aided Mol Des 31, 1–19. [PubMed: 27658802]

Young JY, Westbrook JD, Feng Z, Sala R, Peisach E, Oldfield TJ, Sen S, Gutmanas A, Armstrong DR, Berrisford JM, et al. (2017). OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. Structure 25, 536–545. [PubMed: 28190782]

Yuriev E, Holien J, and Ramsland PA (2015). Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. J Mol Recognit 28, 581–604. [PubMed: 25808539]

Zhu J, Hou T, and Mao X (2015). Discovery of selective phosphatidylinositol 3-kinase inhibitors to treat hematological malignancies. Drug Discovery Today 20, 988–994. [PubMed: 25857437]

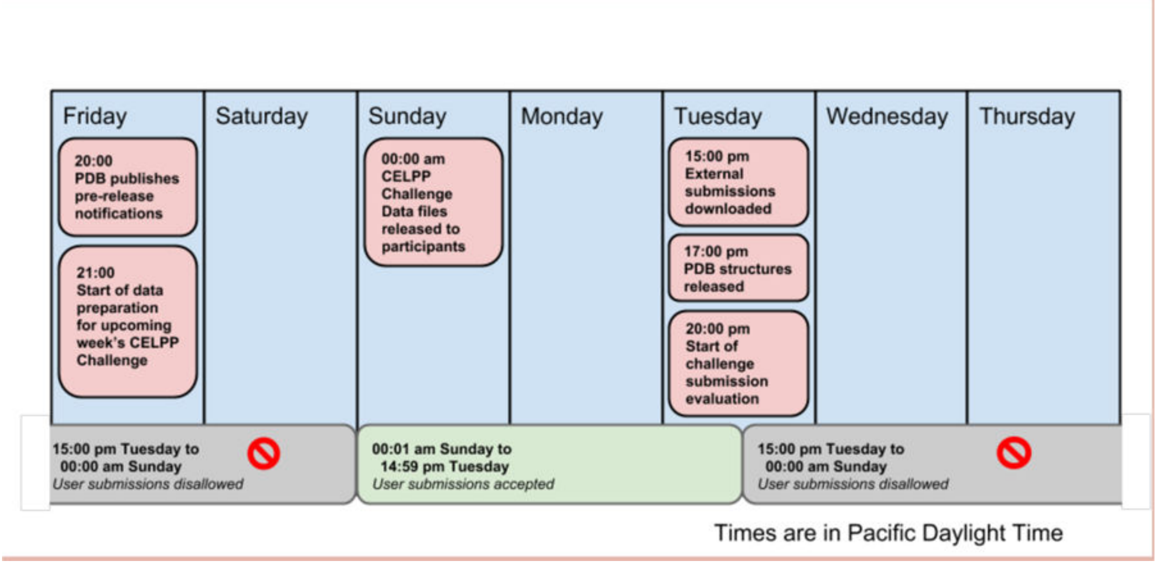

Friday
Saturday
Sunday
Monday
Tuesday
Wednesday
Thursday
20:00 PDB publishes pre-release notifications
21:00 Start of data preparation for upcoming week's CELPP Challenge
00:00 am CELPP Challenge Data files released to participants
15:00 pm External submissions downloaded
17:00 pm PDB structures released
20:00 pm Start of challenge submission evaluation
15:00 pm Tuesday to 00:00 am Sunday User submissions disallowed
00:01 am Sunday to 14:59 pm Tuesday User submissions accepted
15:00 pm Tuesday to 00:00 am Sunday User submissions disallowed
Times are in Pacific Daylight Time

Figure 1. The CELPP week.

The CELPP week begins with the publication of PDB pre-release data on Friday evening. Challenge data preparation runs Friday evening and Saturday, and the upcoming week's challenge package is made available to participants by the beginning of Sunday. Submissions are then accepted until Tuesday at 3:01 pm. Evaluation of the predictions begins on Tuesday evening, following release of the new PDB entries used in the challenge. Times are in the Pacific time zone.

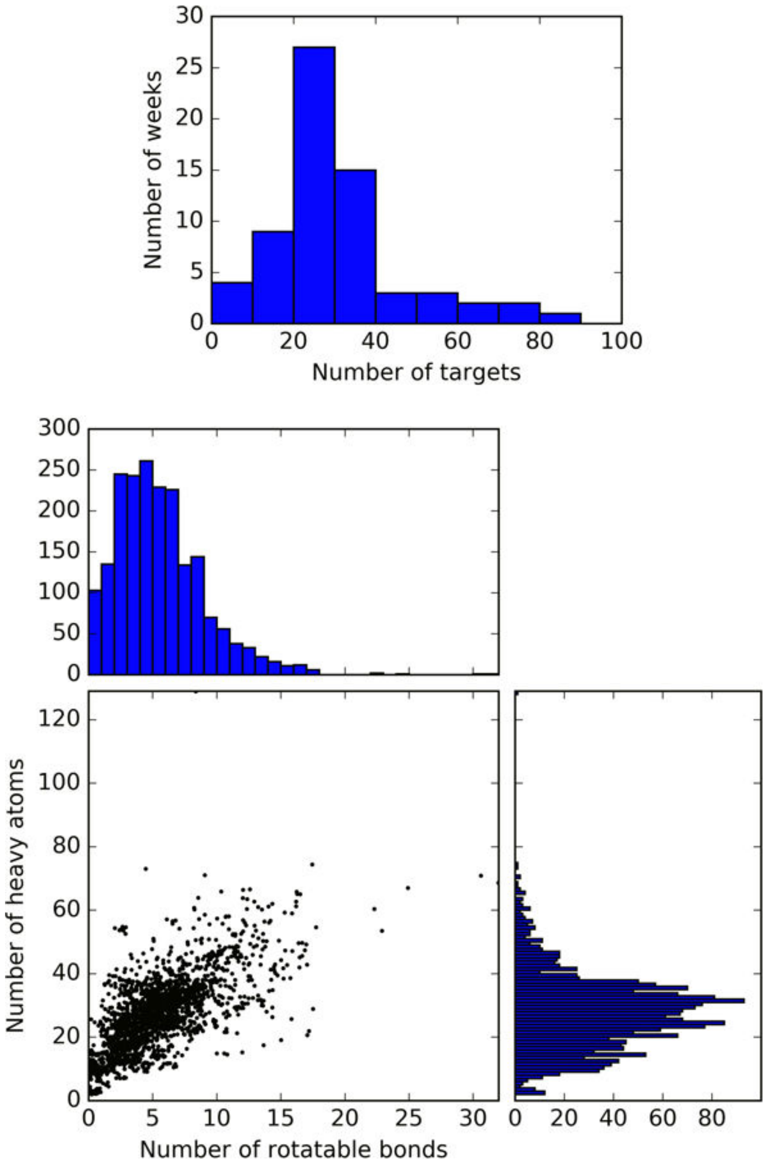


Figure 2. Characteristics of CELPP weekly challenges.

Top) Number of CELPP targets per week (66 weeks total). Bottom) Characteristics of CELPP target ligands (n=1,989). Each dot represents one target ligand, and histograms above and to the right show the distribution of characteristics on each axis. Uniformly distributed random values in the range [−0.5, 0.5] were added to X and Y coordinates to better show point density. Numbers of rotatable bonds and heavy atoms were calculated from InChI strings using RDKit.

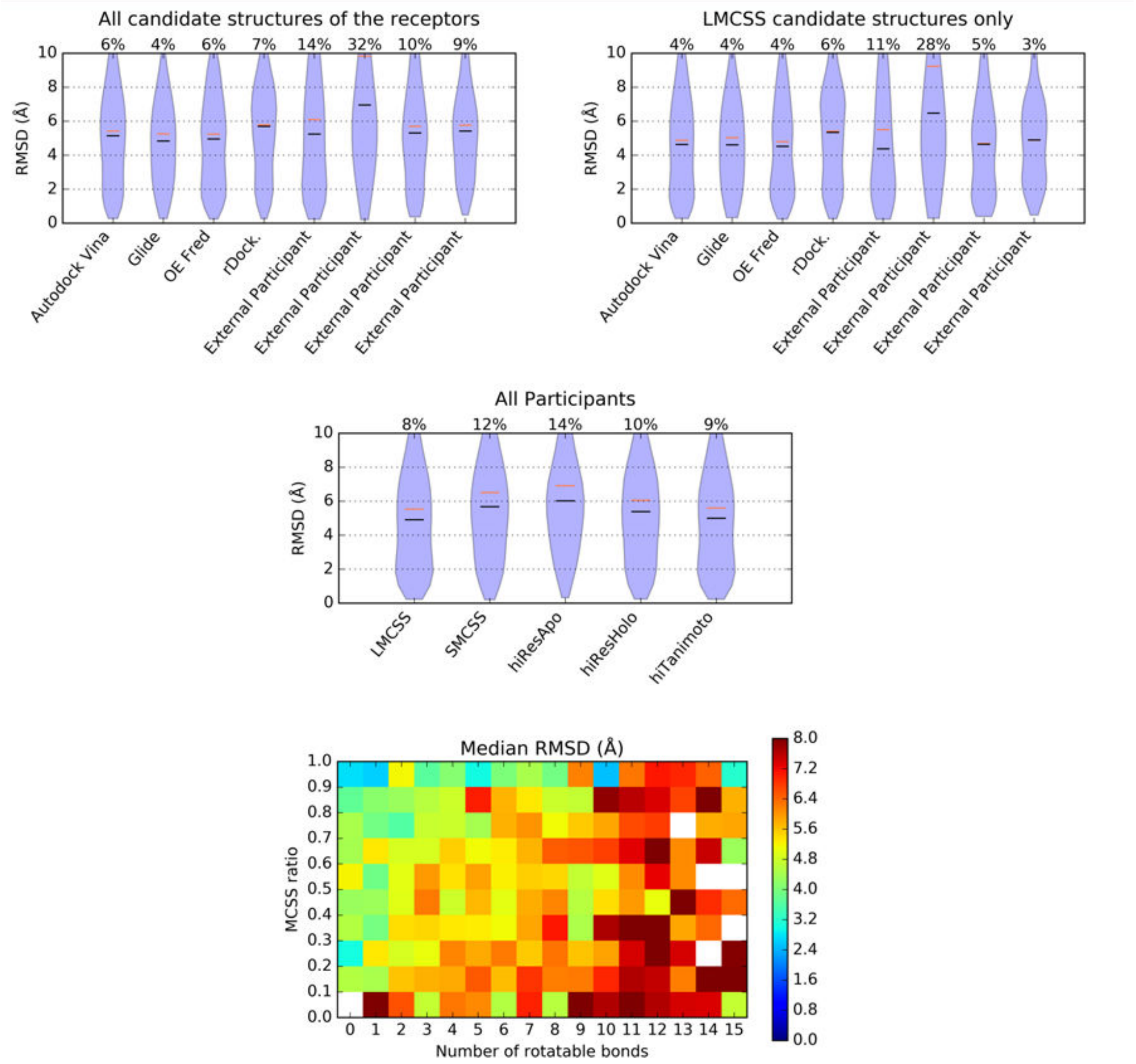


Figure 3. Results of CELPP.

Top Left) Performance by participant or in-house method, combining predictions from all candidate categories. Top Right) Performance by participant or in-house method for the LMCSS category only. Middle) Performance by candidate category, combining predictions from all participants and in-house methods. Black line indicates median, orange line indicates mean. The number above each plot indicates the fraction of predictions above 10 Å. Bottom) Median prediction RMSD as a function of number of rotatable bonds and MCSS ratio. The MCSS ratio is defined as the fraction of the heavy atoms in the target ligand that are in its maximal common substructure with the candidate ligand. Data are taken from all participant and in-house method predictions in LMCSS and SMCSS categories. White indicates no data.

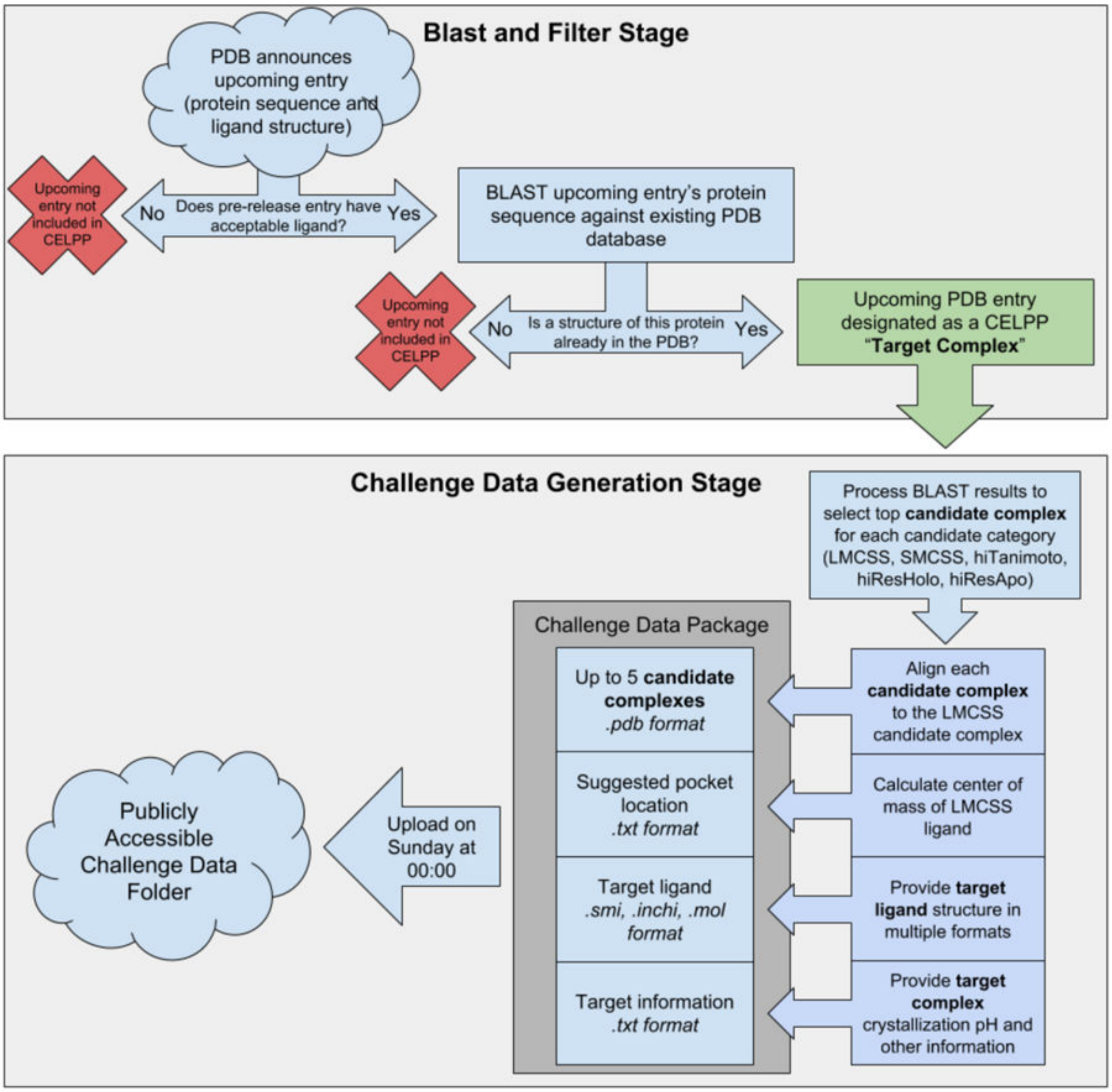

Blast and Filter Stage
PDB announces upcoming entry (protein sequence and ligand structure)
Upcoming entry not included in CELPP
No
Does pre-release entry have acceptable ligand?
Yes
BLAST upcoming entry's protein sequence against existing PDB database
Upcoming entry not included in CELPP
No
Is a structure of this protein already in the PDB?
Yes
Upcoming PDB entry designated as a CELPP "Target Complex"
Challenge Data Generation Stage
Process BLAST results to select top candidate complex for each candidate category (LMCSS, SMCSS, hiTanimoto, hiResHolo, hiResApo)
Challenge Data Package
Up to 5 candidate complexes .pdb format
Suggested pocket location .txt format
Target ligand .smi, .inchi, .mol format
Target information .txt format
Align each candidate complex to the LMCSS candidate complex
Calculate center of mass of LMCSS ligand
Provide target ligand structure in multiple formats
Provide target complex crystallization pH and other information
Upload on Sunday at 00:00
Publicly Accessible Challenge Data Folder

Figure 4. CELPP Target Selection and Challenge Package Generation.

CELPP downloads the publicly-available PDB pre-release information and then processes the new entries to assemble the weekly challenge package. Boxes and arrows indicate processing steps, two-way arrows indicate filtering steps, clouds indicate internet-accessible file platforms, and the dark grey box indicates the weekly challenge data package, in which each target is one subdirectory. See main text for details.

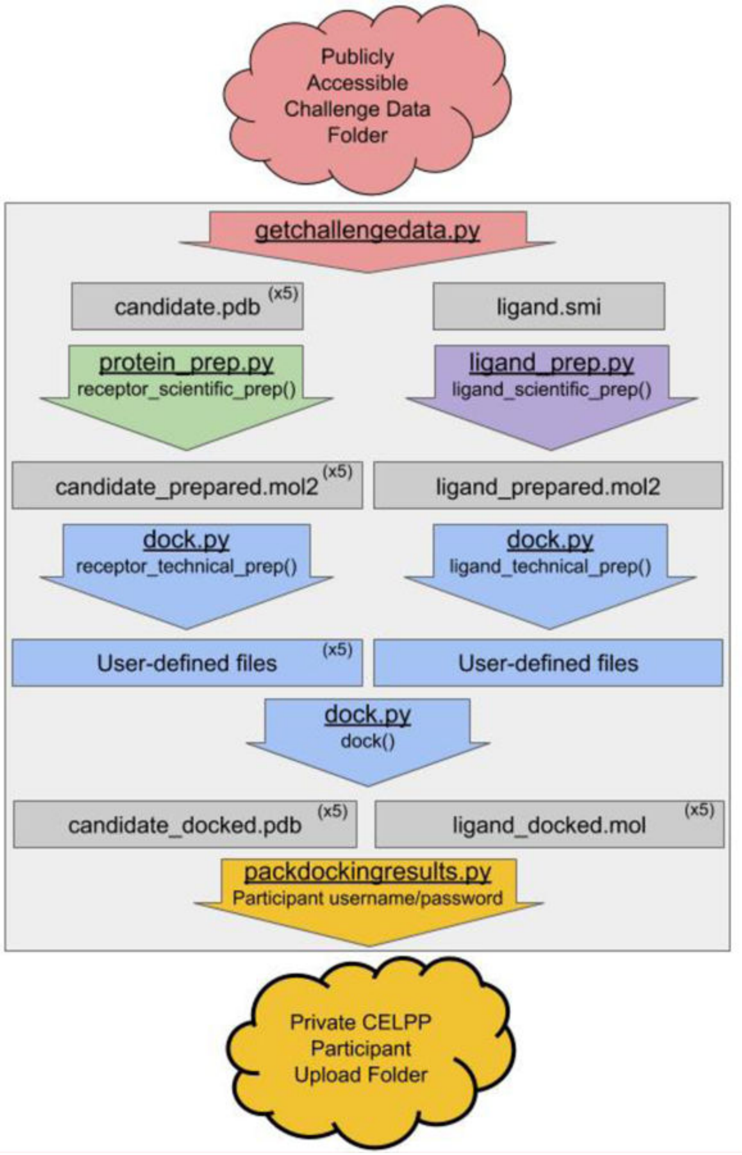


Figure 5. The CELPPade workflow template.

Vertical arrows indicate functions, rectangles indicate files passed between steps, and clouds represent internet-accessible folders. The large grey box indicates the steps that are run on the participant's computer. Different colors indicate script files for different steps of pose prediction, and names ending in () indicate python functions that are implemented by participants. The output files from protein_prep.py and ligand_prep.py is not strictly required to be in mol2 format but adopting this format will improve interoperability of steps from diverse workflows.

Table 1.

Baseline docking workflows.

Methods used for protein preparation, ligand preparation, and docking in the D3R in-house workflows. (Versions: Chimera 1.10.1, RDKit 2016.3.3, MGLTools 1.5.7, AutoDock Vina 1.1.2, Schrodinger 2015–3 release, Omega 2.5.1.4, FRED 3.0.1, RBDock 2013.1/901)

Workflow Name	Protein Prep Method	Ligand Prep Method	Docking Algorithm
Autodock Vina	Chimera DockPrep(Lang et al., 2009; Pettersen et al., 2004) and AutoDock Tools prepare_receptor4.py(Morris et al., 2009)	RDKit 3D coordinate generation, Chimera DockPrep,(Lang et al., 2009; Pettersen et al., 2004) and AutoDock Tools prepare_ligand4.py(Morris et al., 2009)	AutoDock Vina(Trott and Olson, 2009)
Glide	Schrödinger PrepWizard(Sastry et al., 2013)	Schrödinger LigPrep	Schrödinger Glide SP(Friesner et al., 2004; Halgren et al., 2004)
OE Fred	HETATM removal and OpenEye receptor_setup	OpenEye Omega(Hawkins et al., 2010)	FRED(McGann, 2011, 2012)
rDock	HETATM removal, Chimera DockPrep,(Lang et al., 2009; Pettersen et al., 2004) rbcavity(Ruiz-Carmona et al., 2014)	RDKit 3D coordinate generation	rbdock(Ruiz-Carmona et al., 2014)