

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

Jeffrey R. Wagner¹, Christopher P. Churas¹, Shuai Liu¹, Robert V. Swift¹, Michael Chiu¹, Chenghua Shao², Victoria A. Feher¹, Stephen K. Burley², Michael K. Gilson^{*1,3}, Rommie E. Amaro^{*1,4}

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

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

³ Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093, USA

⁴ 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

1 SUMMARY

Docking calculations can be used to accelerate drug discovery by providing predictions of the poses of candidate ligands bound to a targeted protein. However, studies in the literature use varied docking methods, and it is not clear which work best, either in general or for specific protein targets. In addition, a complete docking calculation requires components beyond the docking algorithm itself, such as preparation of the protein and ligand for calculations, and it is difficult to isolate which aspects of a method are most in need of improvement. To address such issues, we have developed the Continuous Evaluation of Ligand Protein Predictions (CELPP), a weekly blinded challenge for automated docking workflows. Participants in CELPP create a workflow to predict protein-ligand binding poses, which is then tasked with predicting 10-100 new (never before released) protein-ligand crystal structures each week. CELPP evaluates the accuracy of each workflow's predictions and posts the scores online. CELPP is a new cyberinfrastructure resource 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.

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. In the pharmaceutical industry, this step has been estimated to require over three years of work on average, at a net cost per launched drug rivaling that of clinical trials.¹ The process is perhaps most efficient when a high-resolution structure of the targeted protein is available, such as from X-ray crystallographic studies. In this setting, structure-based computational methods may be used to accelerate the discovery of high affinity ligands.²⁻¹³ The computational challenge of structure-based ligand design may be considered to comprise two main components. The first is prediction of the bound conformation, or pose, of a candidate ligand, typically by fast, ligand-protein docking algorithms.¹⁴⁻²⁸ The second involves using the predicted pose to assess the candidate ligand's binding affinity for the targeted protein, or at least to rank its affinity relative to the affinities of other compounds one contemplates purchasing or synthesizing. Both components have been the subject of intensive research and development in both academic and commercial settings.²⁹⁻⁴⁴ Nonetheless, computational methods for pose prediction and affinity ranking have yet to fulfill their perceived promise, as neither is yet fully reliable.⁴⁵⁻⁵¹ In fact, it is surprisingly difficult even to compare the reliability of various methods in a consistent manner, and this limitation makes it correspondingly difficult to make and verify technical progress. Part of the challenge of rigorously comparing methods relates to reproducibility (or lack thereof) of the complicated and highly variable end-to-end computational experiments required for pose and affinity prediction.⁵² To address these issues, the community has seen a dramatic uptick in the use and availability of automated workflows that clearly memorialize a particular experiment and provide different approaches for their execution and deployment.⁵³⁻⁵⁷

Further, although many performance comparisons have been published, the results can be difficult to interpret.^{47, 58-68} For example, new docking algorithms are frequently published along with a comparison against existing methods, but this 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, even when a study carries out careful benchmarking of multiple methods against a common dataset, the dataset often contains protein-ligand cocrystal structures that have already been published. Such retrospective studies are suboptimal, because they risk unintentional bias and because structures in the test set might have been used previously in training the docking algorithms.⁶⁹

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,⁴⁷ CSAR,^{66, 70-73} and GPCRDOCK.⁷⁴⁻⁷⁶ Similarly, in recent years, the Drug Design Resource (D3R) has run blinded prediction challenges called the Grand Challenges.^{45, 46} These efforts have led to useful benchmarking strategies, provided insight about best practices, and sometimes yielded unexpected results regarding the effectiveness of various technical approaches. However, such episodic challenges have not been large and systematic enough to afford statistically

meaningful distinctions among individual methods or to support an efficient cycle of development and evaluation that can persistently accelerate progress in the field.

Here, we introduce a new blinded prediction challenge which overcomes these limitations by taking advantage of ongoing available data streams and advances in data science.⁷⁷⁻⁷⁹ The Continuous Evaluation of Ligand Protein Predictions (CELPP) challenge uses the Protein Data Bank's (PDB)⁸⁰⁻⁸⁴ weekly publication of a list of structures slated for imminent release into the public domain as the basis for a weekly, pose-prediction challenge. This rolling challenge is akin to the Continuous Automated Model Evaluation (CAMEO) protein structure prediction challenge,⁸⁵ which served as its inspiration. The following sections detail the structure of the CELPP 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 METHODS

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 which contain 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 mother liquor from the crystallographic study. Additional scripts then search existing PDB entries for crystal structures of each target protein and extract up to five structures appropriate for docking calculations, as detailed in Section 3.2.1. These protein structures are incorporated into the weekly CELPP data package, along with the ligand identities, crystallization pH values, and additional information (Section 3.2.2). 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; i.e., 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>).

It should be evident from this description that CELPP presents what is known as a *cross-docking* challenge, i.e., the protein structure into which the ligand is docked was previously determined either with a different ligand or with no ligand at all.^{17, 86, 87} 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 used has not adapted to the ligand being docked, and indeed may be in a conformation that the ligand does not fit well. However, cross-docking is arguably 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.^{45, 46} 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 ambiguously assigned electron density, whether to remove or retain specific solvent molecules, how to account for crystallization artifacts (such as non-natural solvents, crystal contacts, and non-physiologic temperature), whether to select alternate side-chain conformations, and how to set the protonation states of titratable residues.^{62, 88-91} For the ligand, issues may include assignment of protonation and tautomer states, and the conformations of flexible rings.^{88, 92, 93}

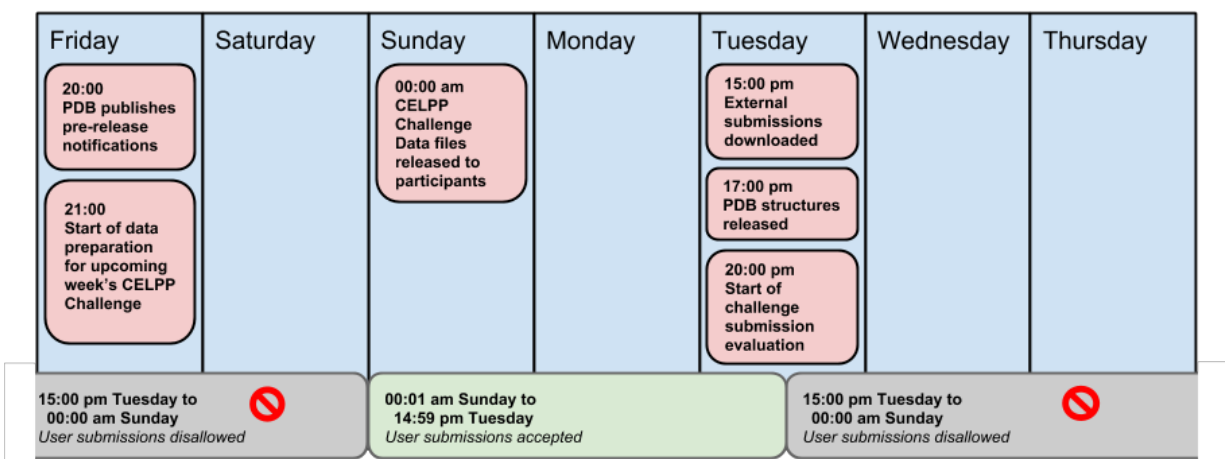


Figure 1. The CELPP week. The CELPP week begins with the publication of PDB pre-release data on Friday evening. Challenge data preparation runs Friday evening and Saturday, and the upcoming week's challenge package is made available to participants by the beginning of Sunday. Submissions are then accepted 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.

3.2 HOSTING THE CHALLENGE

3.2.1 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 at the end of the following Tuesday (Figure 1). 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,⁹⁴ 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. (See Figure 2.) 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.⁹⁵ 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 blastp⁹⁶ to find PDB entries with >95% sequence identity and >90% sequence coverage of the target sequence (Figure 2). 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. 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:

- **Largest Maximum Common Substructure (LMCSS):** This candidate complex 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 suggest the binding pocket for all five candidates for this target. In the case that two candidate complexes tie for the largest maximal common substructure, the highest-resolution candidate structure is used.
- **Smallest Maximum Common Substructure (SMCSS):** This candidate complex contains the ligand with the smallest maximal common substructure to the target ligand. In the case that two candidate complexes tie for the smallest maximal common substructure, the highest-resolution candidate is used.
- **Highest Tanimoto Similarity (hiTanimoto):** This candidate complex contains the ligand with the highest ligand Tanimoto score, using the RDKit default fingerprint method,⁹⁷ to the target ligand. In the case that two candidate complexes tie for the highest mutual Tanimoto score, the highest-resolution candidate is used. This candidate sometimes is the same as the LMCSS candidate.
- **Highest Resolution Holo (hiResHolo):** This candidate complex has the highest crystallographic resolution limit of any determined with a druglike ligand.
- **Highest Resolution Apo (hiResApo):** This candidate receptor structure has the highest crystallographic resolution limit of any for this protein determined with no druglike ligand.

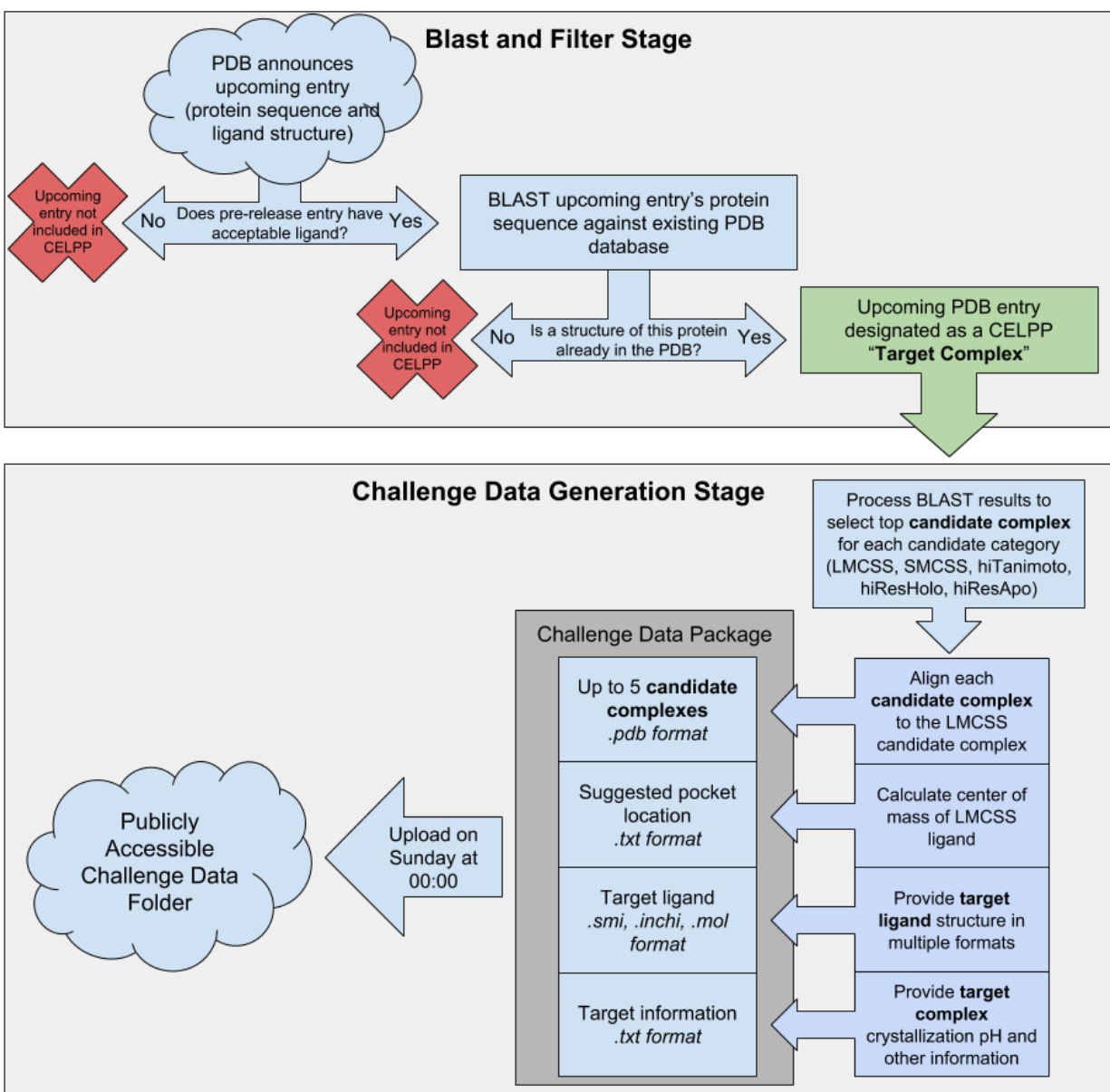


Figure 2. CELPP Target Selection and Challenge Package Generation. CELPP downloads the publicly-available PDB pre-release information and then processes the new entries to assemble the weekly challenge package. Boxes and arrows indicate processing steps, two-way arrows indicate filtering steps, clouds 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.

3.2.2 Generation of the challenge data package

After the processing described in Section 3.2.1 has been completed, the results are incorporated into a common data package for use by CELPP participants. This typically becomes available in a public Box.com folder by 00:00 Pacific Time on Sunday (Figure 1). For each target, the challenge data package (Figure 2) contains:

- 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
- PDB structure of the LMCSS ligand, drawn from the LMCSS candidate structure
- The suggested binding pocket center (center of mass of the LMCSS ligand) in .txt format
- SMILES, InChI, and 2D MOL files of the target ligand
- 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.

3.2.3 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.

3.3 PARTICIPATING IN CELPP

3.3.1 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.

3.3.2 Developing a Prediction Workflow

Based on the typical throughput of CELPP challenges, CELPP participants should construct a pose-prediction workflow that can process up to 100 targets in the 63-hour submission window (Figure 4A). This requires executing up to 100 ligand preparation tasks, 400-500 protein preparation tasks, and 400-500 docking tasks. Although CELPP participants may choose to submit results for only a subset of the targets posed each week, doing so makes it difficult to compare different methods on an equal footing, so participants are encouraged to work with all targets each week. To help participants construct their workflows, D3R provides a workflow template, called CELPPade (Figure 3, <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 3, `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 that follows the creation of a model docking workflow based on CELPPade. This docking workflow uses Chimera DockPrep⁹⁸ to prepare both the protein and ligand, and AutoDock Vina to carry out pose prediction.^{20, 99, 100} 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.⁹⁸⁻¹⁰¹ 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.

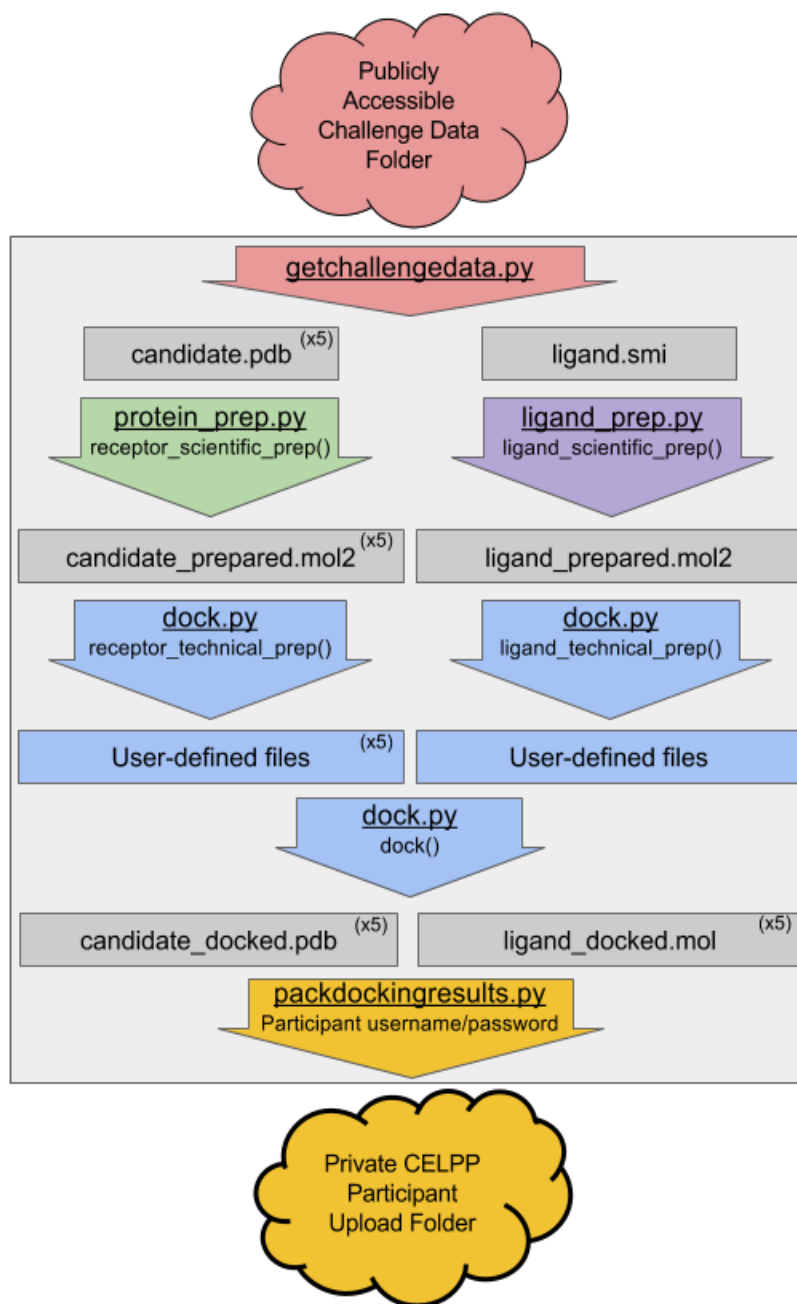


Figure 3. The CELPPade workflow template. Vertical arrows indicate functions, rectangles indicate files passed between steps, and clouds represent internet-accessible folders. The large grey box indicates the steps that are run on the participant's computer. Different colors indicate script files for 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.

3.3.3 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,^{102, 103} with ligand

coordinates in the receptor frame of reference. Participants may choose not to use any of the provided candidate receptor structures for docking but must inform us in this case. Strict adherence to these file formats is required, and deviation from them may result in improper scoring or disqualification of the submitted prediction. Using the CELPPade workflow template will ensure 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.

3.3.4 Score Reporting

Scores are emailed directly to participants and, in the near future, will 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.

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 ^{98,99} and AutoDock Tools prepare_receptor4.py ¹⁰⁴	RDKit 3D coordinate generation, ⁹⁷ Chimera DockPrep, ^{98,99} and AutoDock Tools prepare_ligand4.py ¹⁰⁴	AutoDock Vina ²⁰
Glide	Schrödinger PrepWizard ^{88,105}	Schrödinger LigPrep ¹⁰⁶	Schrödinger Glide SP ¹⁰⁷⁻¹⁰⁹
OE Fred	HETATM removal and OpenEye receptor_setup ¹¹⁰	OpenEye Omega ¹¹¹	FRED ^{68,110,112}
rDock	HETATM removal, Chimera DockPrep, ^{98,99} rbcavity ¹¹³	RDKit 3D coordinate generation ⁹⁷	rbdock ¹¹³

3.4 BASELINE DOCKING WORKFLOWS

To illustrate CELPP, test our procedures, and provide a baseline of performance, we have used the CELPPade template (Section 3.3.2) to create four in-house CELPP workflows, based on the Autodock Vina,²⁰ FRED,^{68,112} Glide,^{107,108} and rDock^{113,114} docking suites. These use the protein and ligand preparation tools that accompany the respective docking codes where possible, and open-source tools otherwise (Table 1). 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 15x15x15 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. All of 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 Section 3.3.2.

4 RESULTS

4.1 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 (Section 3.2.1) and were submitted to outside participants and our in-house workflows (Section 3.4). To permit meaningful analysis, the initial data discussed in this paper have been filtered to include only 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 4, 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 4, 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.

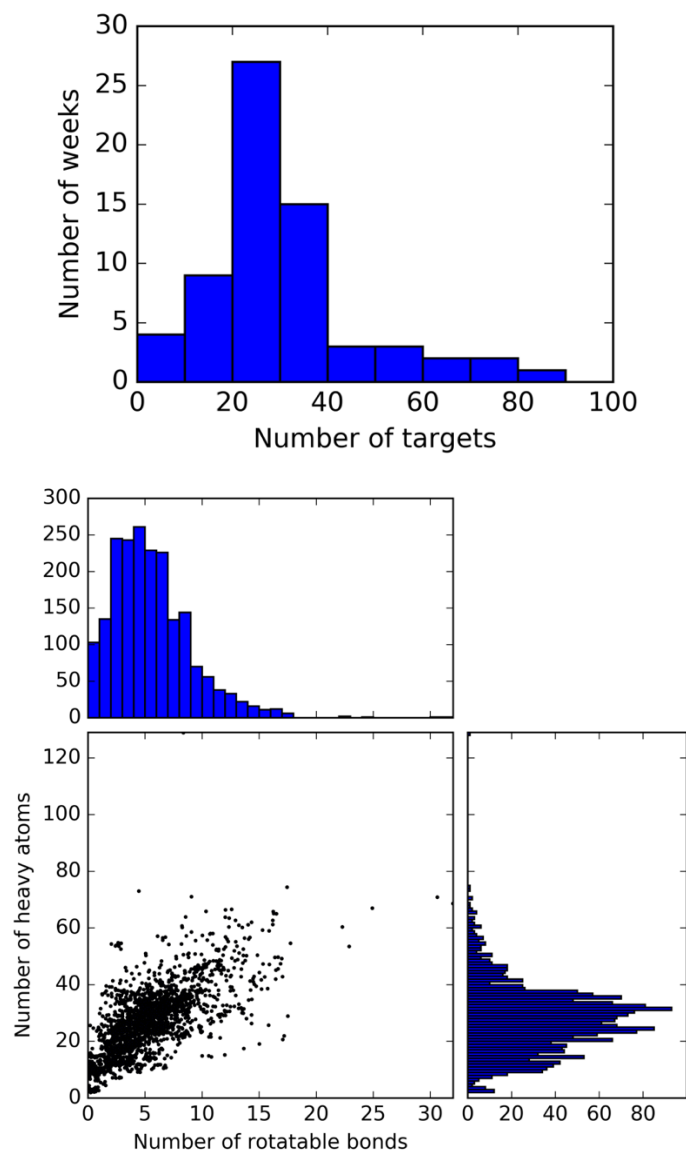


Figure 4: Characteristics of CELPP weekly challenges. Top) Number of CELPP targets per week (66 weeks total). Bottom) Characteristics of CELPP target ligands (n=1,989). Each dot represents one target ligand, and histograms above and to the right show the distribution of characteristics on each axis. 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.⁹⁷

4.2 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 already provide number of 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.^{20, 45-47} In the CELPP dataset, the median prediction RMSD for the best-case prediction categories (LMCSS and hiTanimoto) is around 5 Å (Figure 5, 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,⁴⁷ 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 found in the pose-prediction components of the recent D3R Grand Challenge 3, which yielded a corresponding success rate of 16%. Much as in CELPP, Grand Challenge participants are not provided with expertly selected and prepared receptor structures.

A finer-grained analysis reveals that most methods provide rather similar levels of accuracy (Figure 5, top), 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 extensive data set provided by CELPP also allows quantification of important trends that have been previously noted.^{45, 46} 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 5, 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.

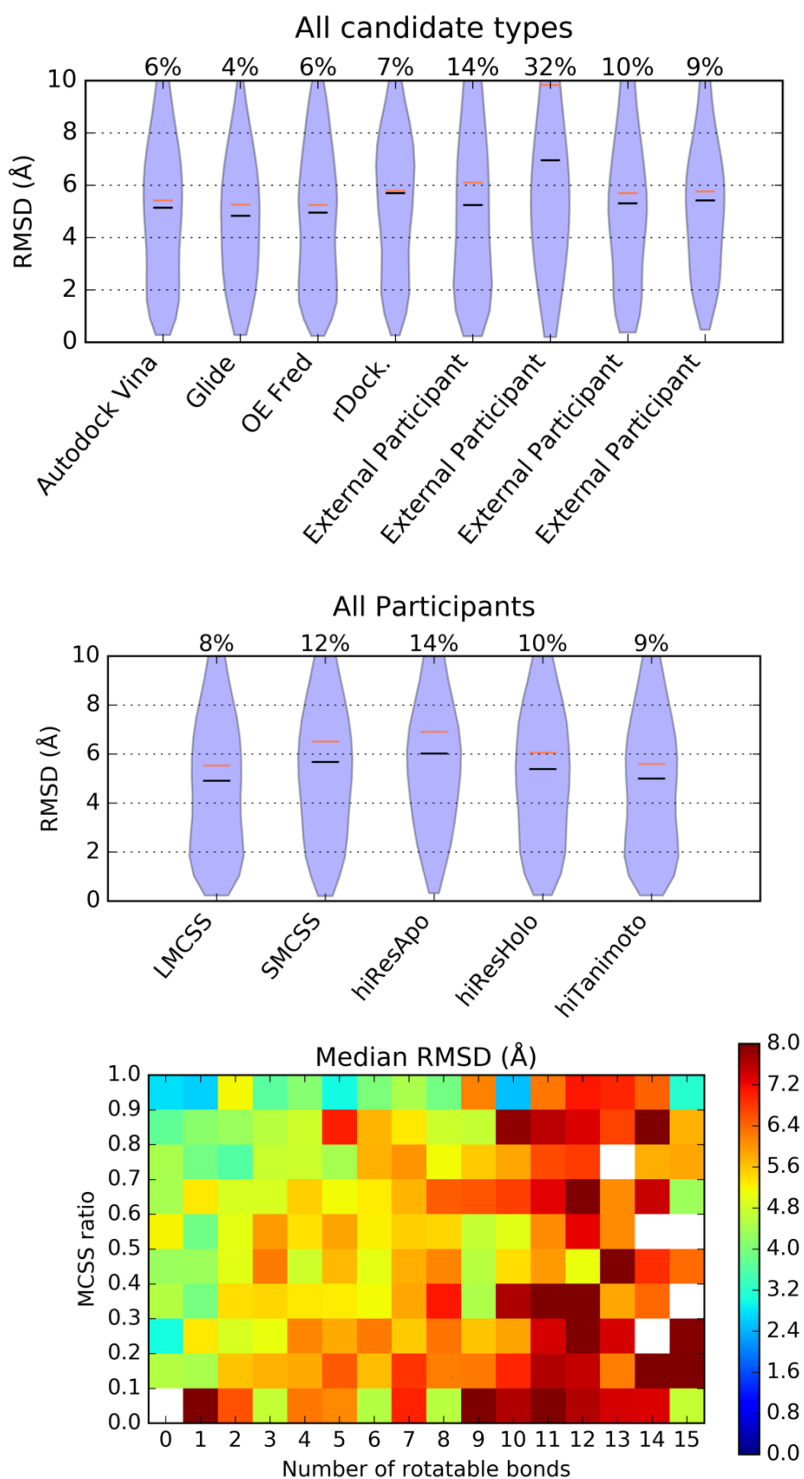


Figure 5. Results of CELPP. Top) Performance by participant or in-house method, combining predictions from all candidate categories. Middle) Performance by candidate category, combining predictions from all participants and in-house methods. Black line indicates median, orange line indicates mean. The 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.

5 DISCUSSION

The CELPP challenge introduced here is a powerful new cyberinfrastructure 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.

The high throughput nature of CELPP provides a dramatic increase in the statistical power of analyses for docking, and thus enables sharper distinctions among methods. We anticipate new insights into not only the core docking algorithms, but also key 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 help practitioners choose methods suited for their specific applications and set meaningful expectations for the quality of predictions on new systems. In parallel, we will 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.

It is 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. Perhaps the best way to address this will ultimately be for participants to package their stable methods into shareable workflows, using technologies such as Docker¹¹⁵ and Singularity¹¹⁶, which can then be executed automatically on machines hosted by the CELPP project. Participants also would benefit by not having to manage the processes or allocate computer time for the calculations. Ideally, such workflows would produce consistent results and be structured into modular steps with standardized I/O, as this would enable 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 put to use.

6 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.

7 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.

8 REFERENCES

1. Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L., How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery* **2010**, *9*, 203.
2. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W., Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66* (1), 334-395.
3. Kuhn, B.; Guba, W.; Hert, J.; Banner, D.; Bissantz, C.; Ceccarelli, S.; Haap, W.; Körner, M.; Kuglstatter, A.; Lerner, C.; Mattei, P.; Neidhart, W.; Pinard, E.; Rudolph, M. G.; Schulz-Gasch, T.; Woltering, T.; Stahl, M., A Real-World Perspective on Molecular Design. *J. Med. Chem.* **2016**, *59* (9), 4087-4102.
4. Anna, E. L.; Stephan, M. L.; Billy Williams-Noonan and Stephanie, S. S., A Practical Guide to Molecular Docking and Homology Modelling for Medicinal Chemists. *Current Topics in Medicinal Chemistry* **2017**, *17* (18), 2023-2040.
5. Sledz, P.; Cafilisch, A., Protein structure-based drug design: from docking to molecular dynamics. *Curr Opin Struct Biol* **2017**, *48*, 93-102.
6. Raghavendra, N. M.; Pingili, D.; Kadasi, S.; Mettu, A.; Prasad, S. V. U. M., Dual or multi-targeting inhibitors: The next generation anticancer agents. *European Journal of Medicinal Chemistry* **2018**, *143*, 1277-1300.
7. Santiago, V.; Eduardo, S.-S.; Lourdes Santana and Eugenio, U., Molecular Docking and Drug Discovery in β -Adrenergic Receptors. *Current Medicinal Chemistry* **2017**, *24* (39), 4340-4359.
8. Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C., Ensemble Docking in Drug Discovery. *Biophysical Journal* **2018**, *114* (10), 2271-2278.
9. Rosenfeld, R.; Vajda, S.; DeLisi, C., Flexible Docking and Design. *Annual Review of Biophysics and Biomolecular Structure* **1995**, *24* (1), 677-700.
10. Lybrand, T. P., Ligand—protein docking and rational drug design. *Current Opinion in Structural Biology* **1995**, *5* (2), 224-228.
11. Leelananda, S. P.; Lindert, S., Computational methods in drug discovery. *Beilstein Journal of Organic Chemistry* **2016**, *12*, 2694-2718.
12. Ganesan, A.; Barakat, K., Applications of computer-aided approaches in the development of hepatitis C antiviral agents. *Expert Opinion on Drug Discovery* **2017**, *12* (4), 407-425.
13. Bartuzi, D.; Kaczor, A. A.; Targowska-Duda, M. K.; Matosiuk, D., Recent Advances and Applications of Molecular Docking to G Protein-Coupled Receptors. *Molecules* **2017**, *22* (2).
14. Sinko, W.; Lindert, S.; McCammon, J. A., Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. *Chem. Biol. Drug Des.* **2013**, *81* (1), 41-49.
15. Amaro, R. E.; Baron, R.; Andrew McCammon, J., An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* **2008**, *22* (9), 693-705.
16. Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K., Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.* **2002**, *45* (11), 2213-2221.
17. Yuriev, E.; Holien, J.; Ramsland, P. A., Improvements, trends, and new ideas in molecular docking: 2012-2013 in review. *J. Mol. Recognit.* **2015**, *28* (10), 581-604.
18. Guedes, I. A.; de Magalhães, C. S.; Dardenne, L. E., Receptor-ligand molecular docking. *Biophys. Rev.* **2014**, *6* (1), 75-87.
19. Cheng, T.; Li, Q.; Zhou, Z.; Wang, Y.; Bryant, S. H., Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. *AAPS J.* **2012**, *14* (1), 133.

20. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, NA-NA.
21. Goodsell, D. S.; Morris, G. M.; Olson, A. J., Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **1996**, 9 (1), 1-5.
22. Goodsell, D. S.; Olson, A. J., Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, 8 (3), 195-202.
23. Kuntz, I. D., Structure-based strategies for drug design and discovery. *Science* **1992**, 257 (5073), 1078-1082.
24. Ewing, T. J. A.; Kuntz, I. D., Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, 18 (9), 1175-1189.
25. Knegtel, R. M.; Kuntz, I. D.; Oshiro, C. M., Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, 266 (2), 424-440.
26. Chen, R.; Li, L.; Weng, Z., ZDOCK: An initial - stage protein - docking algorithm. *Proteins: Structure, Function, and Bioinformatics* **2003**, 52 (1), 80-87.
27. McGann Mark, R.; Almond Harold, R.; Nicholls, A.; Grant, J. A.; Brown Frank, K., Gaussian docking functions. *Biopolymers* **2002**, 68 (1), 76-90.
28. Sousa Sérgio, F.; Fernandes Pedro, A.; Ramos Maria, J., Protein–ligand docking: Current status and future challenges. *Proteins: Structure, Function, and Bioinformatics* **2006**, 65 (1), 15-26.
29. Ogungbe, V. I.; Setzer, N. W., The Potential of Secondary Metabolites from Plants as Drugs or Leads against Protozoan Neglected Diseases—Part III: In-Silico Molecular Docking Investigations. *Molecules* **2016**, 21 (10).
30. Mafud, A. C.; Ferreira, L. G.; Mascarenhas, Y. P.; Andricopulo, A. D.; de Moraes, J., Discovery of Novel Antischistosomal Agents by Molecular Modeling Approaches. *Trends in Parasitology* **2016**, 32 (11), 874-886.
31. Beck, K. R.; Kaserer, T.; Schuster, D.; Odermatt, A., Virtual screening applications in short-chain dehydrogenase/reductase research. *The Journal of Steroid Biochemistry and Molecular Biology* **2017**, 171, 157-177.
32. Kim, J.; Yang, G.; Ha, J., Targeting of AMP-activated protein kinase: prospects for computer-aided drug design. *Expert Opinion on Drug Discovery* **2017**, 12 (1), 47-59.
33. Shunmugam, L.; Ramharack, P.; Soliman, M. E. S., Road Map for the Structure-Based Design of Selective Covalent HCV NS3/4A Protease Inhibitors. *The Protein Journal* **2017**, 36 (5), 397-406.
34. Ciancetta, A.; Jacobson, A. K., Structural Probing and Molecular Modeling of the A3 Adenosine Receptor: A Focus on Agonist Binding. *Molecules* **2017**, 22 (3).
35. Singh, N.; Ecker, G., Insights into the Structure, Function, and Ligand Discovery of the Large Neutral Amino Acid Transporter 1, LAT1. *International Journal of Molecular Sciences* **2018**, 19 (5).
36. Li, L.; Jiang, S.; Li, X.; Liu, Y.; Su, J.; Chen, J., Recent advances in trimethoxyphenyl (TMP) based tubulin inhibitors targeting the colchicine binding site. *European Journal of Medicinal Chemistry* **2018**, 151, 482-494.
37. Tan, Z.; Chaudhai, R.; Zhang, S., Polypharmacology in Drug Development: A Minireview of Current Technologies. *ChemMedChem* **2016**, 11 (12), 1211-1218.
38. Hodos Rachel, A.; Kidd Brian, A.; Shameer, K.; Readhead Ben, P.; Dudley Joel, T., In silico methods for drug repurposing and pharmacology. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **2016**, 8 (3), 186-210.
39. Rosano, C.; Ponassi, M.; Santolla, M. F.; Pisano, A.; Felli, L.; Vivacqua, A.; Maggiolini, M.; Lappano, R., Macromolecular Modelling and Docking Simulations for the Discovery of Selective GPER Ligands. *The AAPS Journal* **2016**, 18 (1), 41-46.

40. Morgnanesi, D.; Heinrichs, E. J.; Mele, A. R.; Wilkinson, S.; Zhou, S.; Kulp, J. L., A computational chemistry perspective on the current status and future direction of hepatitis B antiviral drug discovery. *Antiviral Research* **2015**, *123*, 204-215.
41. Topiol, S.; Sabio, M., The role of experimental and computational structural approaches in 7TM drug discovery. *Expert Opinion on Drug Discovery* **2015**, *10* (10), 1071-1084.
42. Zhu, J.; Hou, T.; Mao, X., Discovery of selective phosphatidylinositol 3-kinase inhibitors to treat hematological malignancies. *Drug Discovery Today* **2015**, *20* (8), 988-994.
43. Martinez-Mayorga, K.; Byler, K. G.; Ramirez-Hernandez, A. I.; Terrazas-Alvares, D. E., Cruzain inhibitors: efforts made, current leads and a structural outlook of new hits. *Drug Discovery Today* **2015**, *20* (7), 890-898.
44. Medina-Franco, J. L.; Méndez-Lucio, O.; Dueñas-González, A.; Yoo, J., Discovery and development of DNA methyltransferase inhibitors using in silico approaches. *Drug Discovery Today* **2015**, *20* (5), 569-577.
45. Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J. A.; Kang, Y. N.; Delproposto, J.; Kubish, G.; Dunbar, J. B., Jr.; Carlson, H. A.; Burley, S. K.; Walters, W. P.; Amaro, R. E.; Feher, V. A.; Gilson, M. K., D3R grand challenge 2015: Evaluation of protein-ligand pose and affinity predictions. *J. Comput. Aided Mol. Des.* **2016**, *30* (9), 651-668.
46. Gaieb, Z.; Liu, S.; Gathiaka, S.; Chiu, M.; Yang, H.; Shao, C.; Feher, V. A.; Patrick Walters, W.; Kuhn, B.; Rudolph, M. G.; Burley, S. K.; Gilson, M. K.; Amaro, R. E., D3R Grand Challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies. *J. Comput. Aided Mol. Des.* **2017**, 1-20.
47. Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S., A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49* (20), 5912-5931.
48. Muddana, H. S.; Daniel Varnado, C.; Bielawski, C. W.; Urbach, A. R.; Isaacs, L.; Geballe, M. T.; Gilson, M. K., Blind prediction of host–guest binding affinities: a new SAMPL3 challenge. *Journal of Computer-Aided Molecular Design* **2012**, *26* (5), 475-487.
49. Muddana, H. S.; Fenley, A. T.; Mobley, D. L.; Gilson, M. K., The SAMPL4 host–guest blind prediction challenge: an overview. *Journal of Computer-Aided Molecular Design* **2014**, *28* (4), 305-317.
50. Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K., Overview of the SAMPL5 host-guest challenge: Are we doing better? *J Comput Aided Mol Des* **2017**, *31* (1), 1-19.
51. Zied, G.; Conor, P.; Michael, C.; Huanwang, Y.; Chenghua, S.; Patrick, W.; Millard, L.; Neysa, N.; Scott D., B.; Stephen K., B.; Rommie E., A.; Michael K., G., D3R Grand Challenge 3: Blind Prediction of Protein-Ligand Poses and Affinity Rankings. *ChemRxiv* **2018**.
52. Jansen, J. M.; Cornell, W.; Tseng, Y. J.; Amaro, R. E., Teach–Discover–Treat (TDT): Collaborative computational drug discovery for neglected diseases. *Journal of Molecular Graphics and Modelling* **2012**, *38*, 360-362.
53. Pronk, S.; Pouya, I.; Lundborg, M.; Rotskoff, G.; Wesén, B.; Kasson, P. M.; Lindahl, E., Molecular Simulation Workflows as Parallel Algorithms: The Execution Engine of Copernicus, a Distributed High-Performance Computing Platform. *Journal of Chemical Theory and Computation* **2015**, *11* (6), 2600-2608.
54. Purawat, S.; leong, P. U.; Malmstrom, R. D.; Chan, G. J.; Yeung, A. K.; Walker, R. C.; Altintas, I.; Amaro, R. E., A Kepler Workflow Tool for Reproducible AMBER GPU Molecular Dynamics. *Biophysical Journal* **2017**, *112* (12), 2469-2474.
55. Beisken, S.; Meinl, T.; Wiswedel, B.; de Figueiredo, L. F.; Berthold, M.; Steinbeck, C., KNIME-CDK: Workflow-driven cheminformatics. *BMC Bioinformatics* **2013**, *14* (1), 257.

56. Dakka, J.; Turilli, M.; Wright, D. W.; Zasada, S. J.; Balasubramanian, V.; Wan, S.; Coveney, P. V.; Jha, S., High-throughput Binding Affinity Calculations at Extreme Scales. *arXiv:1712.09168v4 [cs.DC]* **2018**.
57. Balasubramanian, V.; Bethune, I.; Shkurti, A.; Breitmoser, E.; Hruska, E.; Clementi, C.; Laughton, C.; Jha, S. In *ExTASY: Scalable and flexible coupling of MD simulations and advanced sampling techniques*, 2016 IEEE 12th International Conference on e-Science (e-Science), 23-27 Oct. 2016; 2016; pp 361-370.
58. Abel, R.; Wang, L.; Mobley, D. L.; Friesner, R. A., A Critical Review of Validation, Blind Testing, and Real- World Use of Alchemical Protein-Ligand Binding Free Energy Calculations. *Curr. Top. Med. Chem.* **2017**, *17* (23).
59. Grinter, S. Z.; Zou, X., Challenges, Applications, and Recent Advances of Protein-Ligand Docking in Structure-Based Drug Design. *Molecules* **2014**, *19* (7), 10150-10176.
60. Jain, A. N., Bias, reporting, and sharing: computational evaluations of docking methods. *J. Comput. Aided Mol. Des.* **2008**, *22* (3-4), 201-212.
61. Mobley, D. L.; Dill, K. A., Binding of small-molecule ligands to proteins: "what you see" is not always "what you get". *Structure* **2009**, *17* (4), 489-498.
62. Corbeil, C. R.; Williams, C. I.; Labute, P., Variability in docking success rates due to dataset preparation. *Journal of Computer-Aided Molecular Design* **2012**, *26* (6), 775-786.
63. Cole Jason, C.; Murray Christopher, W.; Nissink, J. W. M.; Taylor Richard, D.; Taylor, R., Comparing protein–ligand docking programs is difficult. *Proteins: Structure, Function, and Bioinformatics* **2005**, *60* (3), 325-332.
64. Irwin, J. J., Community benchmarks for virtual screening. *Journal of Computer-Aided Molecular Design* **2008**, *22* (3), 193-199.
65. Spitzer, R.; Jain, A. N., Surflex-Dock: Docking benchmarks and real-world application. *Journal of Computer-Aided Molecular Design* **2012**, *26* (6), 687-699.
66. Damm-Ganamet, K. L.; Smith, R. D.; Dunbar, J. B., Jr.; Stuckey, J. A.; Carlson, H. A., CSAR benchmark exercise 2011-2012: evaluation of results from docking and relative ranking of blinded congeneric series. *J. Chem. Inf. Model.* **2013**, *53* (8), 1853-1870.
67. Neves, M. A. C.; Totrov, M.; Abagyan, R., Docking and scoring with ICM: the benchmarking results and strategies for improvement. *Journal of Computer-Aided Molecular Design* **2012**, *26* (6), 675-686.
68. McGann, M., FRED and HYBRID docking performance on standardized datasets. *J. Comput. Aided Mol. Des.* **2012**, *26* (8), 897-906.
69. Weiss, D. R.; Bortolato, A.; Tehan, B.; Mason, J. S., GPCR-Bench: A Benchmarking Set and Practitioners' Guide for G Protein-Coupled Receptor Docking. *Journal of Chemical Information and Modeling* **2016**, *56* (4), 642-651.
70. Carlson, H. A., Lessons Learned over Four Benchmark Exercises from the Community Structure-Activity Resource. *J. Chem. Inf. Model.* **2016**, *56* (6), 951-954.
71. Carlson, H. A.; Smith, R. D.; Damm-Ganamet, K. L.; Stuckey, J. A.; Ahmed, A.; Convery, M. A.; Somers, D. O.; Kranz, M.; Elkins, P. A.; Cui, G.; Peishoff, C. E.; Lambert, M. H.; Dunbar, J. B., Jr., CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. *J. Chem. Inf. Model.* **2016**, *56* (6), 1063-1077.
72. Smith, R. D.; Damm-Ganamet, K. L.; Dunbar, J. B., Jr.; Ahmed, A.; Chinnaswamy, K.; Delproposto, J. E.; Kubish, G. M.; Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Stuckey, J. A.; Baker, D.; Carlson, H. A., CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. *J. Chem. Inf. Model.* **2016**, *56* (6), 1022-1031.
73. Smith, R. D.; Dunbar, J. B., Jr.; Ung, P. M.-U.; Esposito, E. X.; Yang, C.-Y.; Wang, S.; Carlson, H. A., CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *J. Chem. Inf. Model.* **2011**, *51* (9), 2115-2131.

74. Kufareva, I.; Rueda, M.; Katritch, V.; Stevens, R. C.; Abagyan, R., Status of GPCR Modeling and Docking as Reflected by Community-wide GPCR Dock 2010 Assessment. *Structure* **2011**, *19* (8), 1108-1126.
75. Kufareva, I.; Katritch, V.; Stevens, R. C.; Abagyan, R., Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure* **2014**, *22* (8), 1120-1139.
76. Michino, M.; Abola, E.; Brooks, C. L.; Scott Dixon, J.; Moulton, J.; Stevens, R. C., Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* **2009**, *8* (6), 455-463.
77. Amaro, Rommie E., Drug Discovery Gets a Boost from Data Science. *Structure* **2016**, *24* (8), 1225-1226.
78. Mura, C.; Draizen, E. J.; Bourne, P. E., Structural biology meets data science: does anything change? *Current Opinion in Structural Biology* **2018**, *52*, 95-102.
79. Amaro, R. E.; Mulholland, A. J., Multiscale methods in drug design bridge chemical and biological complexity in the search for cures. *Nature Reviews Chemistry* **2018**, *2*, 0148.
80. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235-242.
81. Rose, P. W.; Prlić, A.; Altunkaya, A.; Bi, C.; Bradley, A. R.; Christie, C. H.; Costanzo, L. D.; Duarte, J. M.; Dutta, S.; Feng, Z.; Green, R. K.; Goodsell, D. S.; Hudson, B.; Kalro, T.; Lowe, R.; Peisach, E.; Randle, C.; Rose, A. S.; Shao, C.; Tao, Y.-P.; Valasatava, Y.; Voigt, M.; Westbrook, J. D.; Woo, J.; Yang, H.; Young, J. Y.; Zardecki, C.; Berman, H. M.; Burley, S. K., The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2017**, *45* (D1), D271-D281.
82. Berman, H.; Henrick, K.; Nakamura, H., Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.* **2003**, *10* (12), 980.
83. Burley, S. K.; Berman, H. M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L.; Christie, C.; Dalenberg, K.; Duarte, J. M.; Dutta, S.; Feng, Z.; Ghosh, S.; Goodsell, D. S.; Green, R. K.; Guranović, V.; Guzenko, D.; Hudson, B. P.; Kalro, T.; Liang, Y.; Lowe, R.; Namkoong, H.; Peisach, E.; Periskova, I.; Prlić, A.; Randle, C.; Rose, A.; Rose, P.; Sala, R.; Sekharan, M.; Shao, C.; Tan, L.; Tao, Y.-P.; Valasatava, Y.; Voigt, M.; Westbrook, J.; Woo, J.; Yang, H.; Young, J.; Zhuravleva, M.; Zardecki, C., RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Research* **2018**, gky1004-gky1004.
84. Young, J. Y.; Westbrook, J. D.; Feng, Z.; Sala, R.; Peisach, E.; Oldfield, T. J.; Sen, S.; Gutmanas, A.; Armstrong, D. R.; Berrisford, J. M.; Chen, L.; Chen, M.; Di Costanzo, L.; Dimitropoulos, D.; Gao, G.; Ghosh, S.; Gore, S.; Guranovic, V.; Hendrickx, P. M. S.; Hudson, B. P.; Igarashi, R.; Ikegawa, Y.; Kobayashi, N.; Lawson, C. L.; Liang, Y.; Mading, S.; Mak, L.; Mir, M. S.; Mukhopadhyay, A.; Patwardhan, A.; Persikova, I.; Rinaldi, L.; Sanz-Garcia, E.; Sekharan, M. R.; Shao, C.; Swaminathan, G. J.; Tan, L.; Ulrich, E. L.; van Ginkel, G.; Yamashita, R.; Yang, H.; Zhuravleva, M. A.; Quesada, M.; Kleywegt, G. J.; Berman, H. M.; Markley, J. L.; Nakamura, H.; Velankar, S.; Burley, S. K., OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. *Structure* **2017**, *25* (3), 536-545.
85. Haas, J.; Barbato, A.; Behringer, D.; Studer, G.; Roth, S.; Bertoni, M.; Mostaguir, K.; Gumienny, R.; Schwede, T., Continuous Automated Model Evaluation (CAMEO) Complementing the Critical Assessment of Structure Prediction in CASP12. *Proteins* **2017**.
86. Shamsara, J., CrossDock: a tool for performing cross-docking using Autodock Vina. *Springerplus* **2016**, *5*.
87. Kumar, A.; Zhang, K. Y. J., A cross docking pipeline for improving pose prediction and virtual screening performance. *J. Comput. Aided Mol. Des.* **2017**.

88. Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **2013**, 27 (3), 221-234.
89. Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A., PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, 35 (Web Server issue), W522-5.
90. Forli, S., Charting a Path to Success in Virtual Screening. *Molecules* **2015**, 20 (10).
91. Alonso, H.; Bliznyuk Andrey, A.; Gready Jill, E., Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews* **2006**, 26 (5), 531-568.
92. Irwin, J. J.; Shoichet, B. K., ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, 45 (1), 177-182.
93. Ebejer, J.-P.; Morris, G. M.; Deane, C. M., Freely available conformer generation methods: how good are they? *J. Chem. Inf. Model.* **2012**, 52 (5), 1146-1158.
94. Heller, S. R.; McNaught, A.; Pletnev, I.; Stein, S.; Tchekhovskoi, D., InChI, the IUPAC International Chemical Identifier. *Journal of Cheminformatics* **2015**, 7 (1), 23.
95. OpenEye Scientific Software, I. OERMSD -- Toolkits -- Python. <https://docs.eyesopen.com/toolkits/python/oechemtk/OEChemFunctions/OERMSD.html> (accessed June 10 2018).
96. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *Journal of Molecular Biology* **1990**, 215 (3), 403-410.
97. RDKit: Open-source cheminformatics. rdkit.org.
98. Pettersen Eric, F.; Goddard Thomas, D.; Huang Conrad, C.; Couch Gregory, S.; Greenblatt Daniel, M.; Meng Elaine, C.; Ferrin Thomas, E., UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* **2004**, 25 (13), 1605-1612.
99. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D., DOCK 6: Combining techniques to model RNA–small molecule complexes. *RNA* **2009**, 15 (6), 1219-1230.
100. Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C., Development and validation of a modular, extensible docking program: DOCK 5. *Journal of Computer-Aided Molecular Design* **2006**, 20 (10), 601-619.
101. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011**, 3 (1), 33.
102. Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J., Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *Journal of Chemical Information and Computer Sciences* **1992**, 32 (3), 244-255.
103. McNaught, A. D.; Wilkinson, A., IUPAC. Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). In *XML on-line corrected version: <http://goldbook.iupac.org> (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins.*, Blackwell Scientific Publications: Oxford, 1997.
104. Morris Garrett, M.; Huey, R.; Lindstrom, W.; Sanner Michel, F.; Belew Richard, K.; Goodsell David, S.; Olson Arthur, J., AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **2009**, 30 (16), 2785-2791.
105. 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.
106. Schrödinger Release 2015-3: LigPrep, Schrödinger, LLC, New York, NY, 2015.
107. Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S.,

Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739-1749.

108. Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L., Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47* (7), 1750-1759.

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

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

111. Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T., Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572-584.

112. McGann, M., FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2011**, *51* (3), 578-596.

113. Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Beatriz Garmendia-Doval, A.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; David Morley, S., rDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **2014**, *10* (4), e1003571.

114. Morley, S. D.; David Morley, S.; Afshar, M., Validation of an empirical RNA-ligand scoring function for fast flexible docking using RiboDock®. *J. Comput. Aided Mol. Des.* **2004**, *18* (3), 189-208.

115. Merkel, D., Docker: lightweight Linux containers for consistent development and deployment. *Linux J.* **2014**, *2014* (239), 2.

116. Kurtzer, G. M.; Sochat, V.; Bauer, M. W., Singularity: Scientific containers for mobility of compute. *PLOS ONE* **2017**, *12* (5), e0177459.

9 TABLE OF CONTENTS IMAGE

