HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2021 May 15.

Published in final edited form as:

J Mol Biol. 2020 May 15; 432(11): 3404–3410. doi:10.1016/j.jmb.2019.12.011.

ClusPro LigTBM: Automated Template-Based Small Molecule Docking

Andrey Alekseenko[1,2,3,*], **Sergei Kotelnikov**[1,2,4,*], **Mikhail Ignatov**[1,2,3,5,*], **Megan Egbert**[6], **Yaroslav Kholodov**[4], **Sandor Vajda**[6,7], **Dima Kozakov**[1,2,5]

[1]**Department of Applied Mathematics and Statistics**, Stony Brook University, 11794 Stony Brook, NY, USA

[2]**Laufer Center for Physical and Quantitative Biology**, Stony Brook University, 11794 Stony Brook, NY, USA

[3]**Institute of Computer Aided Design of the Russian Academy of Sciences**, 123056 Moscow, Russia

[4]**Innopolis University**, 420500 Innopolis, Russia

[5]**Institute for Advanced Computational Sciences**, Stony Brook University, 11794 Stony Brook, NY, USA

[6]**Department of Biomedical Engineering**, Boston University, 02215 Boston, MA, USA

[7]**Department of Chemistry**, Boston University, 02215 Boston, MA, USA

Abstract

The template-based approach has been essential for achieving high-quality models in the recent rounds of blind protein-protein docking competition CAPRI (Critical Assessment of Predicted Interactions). However, few such automated methods exist for protein-small molecule docking. In this paper, we present an algorithm for template-based docking of small molecules. It searches for known complexes with ligands that have partial coverage of the target ligand, performs conformational sampling and template-guided energy refinement to produce a variety of possible poses, and then scores the refined poses. The algorithm is available as the automated ClusPro LigTBM server. It allows the user to specify the target protein as a PDB file and the ligand as a SMILES string. The server then searches for templates and uses them for docking, presenting the user with top-scoring poses and their confidence scores. The method is tested on the Astex Diverse benchmark, as well as on the targets from the last round of the D3R (Drug Design Data Resource) Grand Challenge. The server is publicly available as part of the ClusPro docking server suite at https://ligtbm.cluspro.org/.

Correspondence to Dima Kozakov: Department of Applied Mathematics and Statistics, Stony Brook University, 11794 Stony Brook, NY, USA. midas@laufercenter.org.

* - These authors contributed equally to this work

Introduction

It is well known that the 3D structures of proteins are better conserved than their sequences, and hence modeling based on homology has been established as the method of choice to generate a reliable 3D model of a protein from its amino acid sequence [1]. The same template-based approach has become increasingly important for predicting the structures of protein complexes [2,3], as demonstrated by the results of the CASP (Critical Assessment of Structure Prediction) and CAPRI (Critical Assessment of Predicted Interactions) community-wide experiments [4]. It was shown that such methods can produce higher quality models than traditional direct docking if good templates are available [5]. Based on this observation, we have recently added the option of template-based modeling to our protein-protein docking server ClusPro.

In the present paper, we explore the use of the template-based approach for solving one of the most important problems in computational biophysics, the docking of small ligands to proteins, which has direct applications to drug discovery. Similar tools have been developed for the identification of ligand binding sites [6–8]. Some of these programs also perform ligand docking [9,10] and even high throughput screening [11]. While we cannot claim that the general approach we use is entirely novel, the specific algorithm we have developed performed extremely well in the last round of the D3R (Drug Design Data Resource) Grand Challenge (https://drugdesigndata.org/). Grand Challenge 4 (GC4) was a blinded prediction contest for the computational chemistry community, and it included predicting the poses of 20 ligands binding to beta secretase 1 (BACE 1). Stages 1A and 1B of GC4 were, respectively, cross-docking and self-docking challenges. In both stages our team submitted results [12] that were comparable to the excellent results produced by the groups of G. Wei [13], M. Totrov [14], and X. Zou. The fact that our method consistently was among these top performers motivated us to implement the algorithm as a server and make it available to the research community as part of the ClusPro docking tools. Such automated servers have proven to be very valuable to users. For example, the protein-protein docking server ClusPro [15] has over 5,000 new jobs submitted each month. Here we describe the steps performed by the server and its validation on the D3R Grand Challenge 4 BACE data set and the frequently used docking benchmark known as the Astex non-native set [16].

Methods

Overall workflow

The input data provided by the user consists of the receptor structure (as a PDB file) and the ligand chemical structure (as a SMILES string). A similarity search is performed in the Protein Data Bank (PDB) database to find templates — highly-homologous protein chains with similar ligands. An ensemble of initial conformations is generated for the ligand. For each of the templates found, the next steps are carried out independently. The target protein structure is re-modeled (unless the user opts-out of this functionality) after the protein in the template, and the closest conformation of the ligand is added to the structure. The resulting protein-ligand structures are subjected to restrained all-atom energy minimization (RM) to remove possible clashes and “relax” the ligand. The poses are ranked based on their similarity to the template. Users can supply a list of PDB codes to exclude from the template

search. The pipeline is shown in Fig. 1, and a more detailed description of each stage follows.

Template search

The protocol starts with a search for known structures of closely related complexes. We use BLAST+ [17] to search for the sequence-similar (e-value = 10^{-20}, sequence identity ≥ 30%) chain structures in the Protein Data Bank (PDB). For each ligand in the found structures we calculate two versions of Maximum Common Substructure ("weak" and "strict" MCS) and Tanimoto score based on Daylight molecular fingerprint [18] as implemented in RDKit [19]. "Weak" version of MCS requires a match of atoms, valences, and bond types (with the only exception that single bond can match aromatic bond); "strict" version has an additional requirement that only complete rings can be a part of the MCS. The ligands with Tanimoto score ≥ 0.4 and "weak" MCS coverage ≥ 50% located within 8 Å of the selected chain are retained, thus forming protein-ligand template structures. The templates are ranked based on the score $(c_w - 0.5)^2 + (i - 0.3)^2$, where c_w is the "weak" MCS coverage, i is the protein sequence identity, and 0.5 and 0.3 correspond to the chosen thresholds of 50% and 30%, respectively. Up to 20 templates with the highest score are retained and used for refinement.

Preparation of starting poses

Despite the advancements of refinement protocols, having a good starting pose is still a prerequisite for constructing a low-RMSD model. LigTBM uses the ETKDG method [20] from RDKit [19] to generate 1,000 conformers for the target ligand. For each template, we align all conformers to the template's "weak" MCS and retain only one conformer with the lowest MCS RMSD. Ligand atom partial charges are assigned using the AM1-BCC method [21] implemented in the antechamber module of the Amber software package [22]. For protonation, a pH of 7.4 is assumed. The user-submitted receptor structure is re-modeled by MODELLER [23] using the protein chain from the template structure. This step allows for "fine-tuning" of the (likely unbound) receptor structure submitted by the user to the ligand-bound template. The user has the option to skip re-modeling and use their uploaded structure without change. The resulting receptor and ligand structures are used as the starting poses for the refinement.

Refinement

We employ a basic restrained minimization protocol to refine the generated poses. The protocol is based on all-atom energy minimization using a CHARMM-based energy function with a GBSA-type solvation term (Analytical Continuum Electrostatics, ACE) [24]. During the minimization, all receptor atoms except hydrogens are fixed, while ligand atoms matching the template are restrained with a harmonic potential to the positions of the corresponding template atoms. Besides restraining the MCS, soft Gaussian potential wells are created, centered on each template atom. This gently pulls non-matching parts of the target ligand to the template if they are in the vicinity. Applying these restraints allows us to overcome the limitations of the general forcefield, and implicitly harness the details of interactions in the template X-ray structure. Minimization is performed with the L-BFGS algorithm [25], using an in-house libmol2 library (https://bitbucket.org/bu-structure/libmol2/src/master/).

Scoring and ranking

For each target, the results from all templates are ranked in three steps, first based on the ligand "weak" MCS coverage, second on the "strict" MCS coverage, and third on the receptor sequence identity. This reflects our experience in D3R challenges, where ligand having a close match to the template was essential for obtaining low-RMSD results. The ligand MCS coverage ("weak" and "strict") indicate what fraction of the target ligand is covered by the template, and thus higher values indicate that larger part of the target ligand agrees with the known binder, and smaller part of the ligand has to be modelled "ab initio," without a template. The receptor sequence identity, while serving as a tie-breaker for cases where the same or similar template ligands are bound to different receptors, is not generally indicative of the goodness of fit, since proteins as distant as having 30% sequence identity can have similar fold and function [26]. Additionally, for each model produced a confidence score is computed based on Tanimoto fingerprint similarity. As mentioned, MCS coverage measures the fraction of the target ligand "covered" by the template, and thus indicates the uncertainty in the placement of the ligand. In contrast, Tanimoto (intersection-over-union) fingerprint similarity also reflects if the template is larger than the target ligand. This symmetric similarity is not required for a good fit of the target ligand to the template, but is indicative of the closeness of the modelled interaction to the known one, and therefore it is chosen to reflect model confidence. The models with Tanimoto similarity of 90% and above are assigned "high" confidence; between 65% and 90% "medium" confidence; between 40% and 65% "low" confidence; and the templates with Tanimoto fingerprint similarity below 40% are rejected at the template search stage.

Results and discussion

Validation methodology

The LigTBM server was tested on the D3R Grand Challenge 4 BACE dataset [12] and on the Astex Non-Native Set [16,27]. For each docking case, the PDB database was searched for all ligands bound to proteins with 30% or greater sequence similarity, and all PDB IDs, where a ligand was equivalent to the docking target, were added to an exclusion list prior to submitting to LigTBM. Equivalency was determined by comparing the target ligand SMILES to the PDB ligand SMILES using RDKit [19] (Morgan Fingerprint Tanimoto Similarity) and Pybel [28] (FP2 Similarity). The list of all PDB IDs excluded for each run can be found in Table S2. RMSD of the produced models, relative to the known protein-small molecule complex, was calculated using DockRMSD [29]. The D3R Grand Challenge 4 BACE targets were run with the starting structure from PDB ID 5YGX [30], and LigTBM models were globally aligned using the PyMOL align function [31] to the native protein-ligand complexes provided by the D3R competition organizers. In the Astex Non-Native Set, the non-native receptor structures were locally aligned to the corresponding native structure around the ligand-binding site [32]. The structure with the lowest local RMSD to the native complex was selected as a starting receptor structure for the LigTBM benchmark. The performance was evaluated in terms of ligand RMSD (heavy-atoms only) after the receptor was aligned to the reference X-Ray structure.

Datasets

As mentioned, we used two different benchmark sets to evaluate the performance of the protocol:

- BACE denotes the set of 20 Beta Secretase 1 (BACE 1) inhibitors used as targets in the most recent round of the Drug Design Data Resource (D3R) blind prediction competition, Grand Challenge 4. The list of targets and their features are given in Table S1. The docking algorithm of the ClusPro LigTBM server is inspired by the semi-manual protocol our group successfully employed during this challenge [12]. The set consists of 20 compounds of the same class, of which 19 contain macrocycles. At the time of this writing crystal structures of these complexes were not deposited to the PDB, therefore no complexes were excluded from template search. This dataset served to verify that the automated version of the protocol is comparable in quality to the original one and to estimate its performance on such a difficult type of compounds as macrocycles.
- ANNS denotes the Astex Non-Native Set, containing 65 cases [16,27]. The structures of the bound complexes were excluded from the template search (see Table S2 for the full list). This dataset is used to test the real-use scenario of docking to unbound receptor structures for a wide range of complexes.

Server performance

The performance of the server on the chosen benchmarks is reported in Table 1. The first row shows the total number of targets in a given benchmark. The “Template found” row shows the number of targets for which at least one suitable template was found. If no templates were found for a target, no result was produced. The “Top-1 < 2Å” row shows the number of targets for which the model selected as top 1 had RMSD below 2 Å. The “Top-5 < 2Å” row shows the number of targets for which any of the top-5 models had RMSD below 2 Å. Detailed per-target results, including RMSD for top-5 models, are given in Tables S1 and S2.

The results show that for the BACE set, which served as the inspiration for this protocol, 19 out of 20 cases have a top-1 structure under 2 Å, consistent with the performance of the original semi-manual approach used in D3R GC4 [12]. In the ANNS set, no suitable template was found for 6 out of 65 cases (9%). Among the rest, 44 cases had a top-1 model in the sub-2Å range. This performance (68% of the total number of cases) is on par with the performance of scoring functions in many popular docking programs [33]. However, LigTBM does not require the knowledge of binding pocket, and only requires starting structures of the receptor protein and SMILES of the ligand. These results demonstrate that despite the inherent limitations of the template-based approach, LigTBM is capable of producing and correctly ranking near-native structures for a wide variety of compounds. The predicted confidence score is a good indicator of model reliability, as can be seen in Fig. 2. While some models with “low” confidence can still have low RMSD, the “medium” confidence models reliably have an RMSD score below 4 Å. “High” confidence models typically have sub-2Å RMSD.

Web Server functionality

LigTBM has a simple user interface, which enables easy job submission and navigation. When directed to the web-site for the first time, the user will see the "Sign in" page, with options to either create an account (available for academic users only) or to use the server without an account. In the latter case, all submitted jobs will be publicly accessible. It is recommended to create an account if the data are considered confidential. Once logged in, the user is directed to the job submission page. Only three fields are mandatory for running the docking: "Protein", "Chain ID," and "Ligand SMILES". If the user fills the field "PDB exclusions," the listed PDB entries will be excluded from homology search, which can be useful for testing. If "Do not remodel" field is unchecked, the provided protein structure will be remodeled, but in this case the "MODELLER key" needs to be filled. The MODELLER license key is available from the MODELLER website (https://salilab.org/modeller/) and is free of charge for academic use. Alternatively, the "Do not remodel" field can be checked if the user does not wish to remodel the protein. Once the "Submit job" button is clicked, the job will be submitted to the queue on the computing cluster. If the input contains errors, the user will be asked to correct them. Some input errors (e.g., incorrect PDB file) will be detected prior to the submission, while others such as invalid MODELLER key or invalid SMILES string will result in failure after the job was submitted to the queue.

The status of the submitted job can be tracked on the "Queue" and "Results" pages of the server. While the job is in progress, it is shown on the "Queue" page. The job details can be accessed by clicking the job ID. When the job is completed, it is shown on the "Results" page, where the program output can be downloaded. If the docking was successful, the models of the ligand bound to the protein will be available for download and can be inspected in the 3D viewer plugin on the same page. The scores for each model can be downloaded as a CSV file. If an error occurred during docking, the job will produce a log file explaining the nature of the error. If no template structure was found, the job will not produce any models, and the status will be changed to "No templates found." In case of questions or suggestions the users are encouraged to connect us through the email address specified on the "Contact" page.

Conclusion

Protein-ligand docking is a challenging yet practically important problem of molecular biology. Recently, the template-based approach has been among the top-performing methods both in protein-protein and protein-ligand community-wide blind prediction challenges, such as CAPRI [2,3,34,35] and D3R Grand Challenges. In this paper, we present an automated online server for template-based docking of small molecules to proteins. The fully-automated pipeline allows the users to run docking without the need for specialized hardware or software and is able to produce low-RMSD structures for a variety of compounds. Results demonstrate that the method is applicable to a wide range of protein-ligand systems, with 72% of the ANNS benchmark having a sub-2Å model in top-5 results. It is clear that the template-based method is only applicable if there exists a known structure for a similar receptor-ligand pair. However, we note that docking is most frequently used in the process of drug discovery against targets that are extensively studied, generally resulting in many X-ray structures co-crystallized with a variety of ligands. In fact, docking without

any *a priori* information on the binding site is rare and tend to have a low success rate. The server features a user-friendly interface and is available for free non-commercial use at https://ligtbm.cluspro.org/.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by the National Institutes of Health grants R21 GM127952 and R35 GM118078; the National Science Foundation grants AF 1816314, AF 1645512, and DBI 1759277; the Russian Science Foundation grant 19-74-00090.

Appendix A.: Supplementary data

Supplementary Tables S1-S2 can be found online at https://docs.google.com/spreadsheets/d/1eZig8-W1MJ_uHoWhMfLkVfgBklbrkijGqo9DMDajynM/edit#gid=1944097457

References

[1]. Kryshtafovych A, Schwede T, Topf M, Fidelis K, Moult J, Critical Assessment of Methods of Protein Structure Prediction (CASP) - Round XIII, Proteins. (2019). doi:10.1002/prot.25823.

[2]. Peterson LX, Shin W-H, Kim H, Kihara D, Improved performance in CAPRI round 37 using LZerD docking and template-based modeling with combined scoring functions, Proteins. 86 Suppl 1 (2018) 311–320. doi:10.1002/prot.25376. [PubMed: 28845596]

[3]. Kundrotas PJ, Anishchenko I, Badal VD, Das M, Dauzhenka T, Vakser IA, Modeling CAPRI targets 110–120 by template-based and free docking using contact potential and combined scoring function, Proteins. 86 Suppl 1 (2018) 302–310. doi:10.1002/prot.25380. [PubMed: 28905425]

[4]. Lensink MF, Velankar S, Baek M, Heo L, Seok C, Wodak SJ, The challenge of modeling protein assemblies: the CASP12-CAPRI experiment, Proteins. 86 Suppl 1 (2018) 257–273. doi:10.1002/prot.25419. [PubMed: 29127686]

[5]. Porter KA, Desta I, Kozakov D, Vajda S, What method to use for protein-protein docking?, Curr. Opin. Struct. Biol 55 (2019) 1–7. doi:10.1016/j.sbi.2018.12.010. [PubMed: 30711743]

[6]. Wass MN, Kelley LA, Sternberg MJE, 3DLigandSite: predicting ligand-binding sites using similar structures, Nucleic Acids Res. 38 (2010) W469–73. doi:10.1093/nar/gkq406. [PubMed: 20513649]

[7]. Roy A, Zhang Y, Recognizing protein-ligand binding sites by global structural alignment and local geometry refinement, Structure. 20 (2012) 987–997. doi:10.1016/j.str.2012.03.009. [PubMed: 22560732]

[8]. Hwang H, Dey F, Petrey D, Honig B, Structure-based prediction of ligand-protein interactions on a genome-wide scale, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 13685–13690. doi:10.1073/pnas.1705381114. [PubMed: 29229851]

[9]. Brylinski M, Skolnick J, FINDSITE: a threading-based approach to ligand homology modeling, PLoS Comput. Biol. 5 (2009) e1000405. doi:10.1371/journal.pcbi.1000405. [PubMed: 19503616]

[10]. Pons J-L, Labesse G, @TOME-2: a new pipeline for comparative modeling of protein-ligand complexes, Nucleic Acids Res. 37 (2009) W485–91. doi:10.1093/nar/gkp368. [PubMed: 19443448]

[11]. Zhou H, Skolnick J, FINDSITE(comb): a threading/structure-based, proteomic-scale virtual ligand screening approach, J. Chem. Inf. Model 53 (2013) 230–240. doi:10.1021/ci300510n. [PubMed: 23240691]

[12]. Kotelnikov S, Alekseenko A, Liu C, Ignatov M, Padhorny D, Brini E, Lukin M, Coutsias E, Dill KA, Kozakov D, Sampling and refinement protocols for template-based macrocycle docking: 2018 D3R Grand Challenge 4, Journal of Computer-Aided Molecular Design (accepted). (n.d.).

[13]. Nguyen DD, Gao K, Wang M, Wei G-W, MathDL: Mathematical deep learning for D3R Grand Challenge 4, arXiv [q-bio.QM]. (2019). http://arxiv.org/abs/1909.07784.

[14]. Lam PC-H, Abagyan R, Totrov M, Macrocycle modeling in ICM: benchmarking and evaluation in D3R Grand Challenge 4, J. Comput. Aided Mol. Des. (2019). doi:10.1007/s10822-019-00225-9.

[15]. Kozakov D, Hall DR, Xia B, Porter KA, Padhorny D, Yueh C, Beglov D, Vajda S, The ClusPro web server for protein-protein docking, Nat. Protoc 12 (2017) 255–278. doi:10.1038/nprot.2016.169. [PubMed: 28079879]

[16]. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW, Protein-ligand docking against non-native protein conformers, J. Chem. Inf. Model 48 (2008) 2214–2225. doi:10.1021/ci8002254. [PubMed: 18954138]

[17]. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL, BLAST+: architecture and applications, BMC Bioinformatics. 10 (2009) 421. doi:10.1186/1471-2105-10-421. [PubMed: 20003500]

[18]. Daylight Chemical Information Systems, DAYLIGHT Software Manual, (2011). https://www.daylight.com/dayhtml/doc/theory/index.html.

[19]. Landrum G, RDKit: Open-source cheminformatics https://rdkit.org/, 2019 https://rdkit.org/.

[20]. Riniker S, Landrum GA, Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation, J. Chem. Inf. Model 55 (2015) 2562–2574. doi:10.1021/acs.jcim.5b00654. [PubMed: 26575315]

[21]. Jakalian A, Bush BL, Jack DB, Bayly CI, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method, J. Comput. Chem. 21 (2000) 132–146.

[22]. Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Izadi S, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein DJ, Merz KM, Miao Y, Monard G, Nguyen C, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Smith J, Salomon-Ferrer R, Swails J, Walker RC, Wang J, Wei H, Wolf RM, Wu X, Xiao L, York DM, Kollman PA, AMBER 2018, (2018). http://ambermd.org/.

[23]. Webb B, Sali A, Comparative Protein Structure Modeling Using MODELLER, Curr. Protoc. Bioinformatics 54 (2016) 5.6.1–5.6.37. doi:10.1002/cpbi.3. [PubMed: 27322406]

[24]. Moghadasi M, Mirzaei H, Mamonov A, Vakili P, Vajda S, Paschalidis IC, Kozakov D, The impact of side-chain packing on protein docking refinement, J. Chem. Inf. Model. 55 (2015) 872–881. doi:10.1021/ci500380a. [PubMed: 25714358]

[25]. Liu DC, Nocedal J, On the limited memory BFGS method for large scale optimization, Math. Program 45 (1989) 503–528. doi:10.1007/BF01589116.

[26]. Sánchez R, Sali A, Large-scale protein structure modeling of the Saccharomyces cerevisiae genome, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 13597–13602. doi:10.1073/pnas.95.23.13597. [PubMed: 9811845]

[27]. Hartshorn MJ, Verdonk ML, Chessari G, Brewerton SC, Mooij WTM, Mortenson PN, Murray CW, Diverse, high-quality test set for the validation of protein-ligand docking performance, J. Med. Chem 50 (2007) 726–741. doi:10.1021/jm061277y. [PubMed: 17300160]

[28]. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR, Open Babel: An open chemical toolbox, J. Cheminform 3 (2011) 33. doi:10.1186/1758-2946-3-33. [PubMed: 21982300]

[29]. Bell EW, Zhang Y, DockRMSD: an open-source tool for atom mapping and RMSD calculation of symmetric molecules through graph isomorphism, J. Cheminform. 11 (2019) 40. doi:10.1186/s13321-019-0362-7. [PubMed: 31175455]

[30]. Nakahara K, Fuchino K, Komano K, Asada N, Tadano G, Hasegawa T, Yamamoto T, Sako Y, Ogawa M, Unemura C, Hosono M, Ito H, Sakaguchi G, Ando S, Ohnishi S, Kido Y, Fukushima T, Dhuyvetter D, Borghys H, Gijsen HJM, Yamano Y, Iso Y, Kusakabe K-I, Discovery of Potent

and Centrally Active 6-Substituted 5-Fluoro-1,3-dihydro-oxazine β-Secretase (BACE1) Inhibitors via Active Conformation Stabilization, J. Med. Chem. 61 (2018) 5525–5546. doi:10.1021/acs.jmedchem.8b00011. [PubMed: 29775538]

[31]. Schrödinger LLC, The PyMOL Molecular Graphics System, Version 2.0, (2019). https://pymol.org/2/.

[32]. Beglov D, Hall DR, Wakefield AE, Luo L, Allen KN, Kozakov D, Whitty A, Vajda S, Exploring the structural origins of cryptic sites on proteins, Proc. Natl. Acad. Sci. U. S. A 115 (2018) E3416–E3425. doi:10.1073/pnas.1711490115. [PubMed: 29581267]

[33]. Verdonk ML, Ludlow RF, Giangreco I, Rathi PC, Protein-Ligand Informatics Force Field (PLIff): Toward a Fully Knowledge Driven "Force Field" for Biomolecular Interactions, J. Med. Chem. 59 (2016) 6891–6902. doi:10.1021/acs.jmedchem.6b00716. [PubMed: 27353137]

[34]. Lensink MF, Velankar S, Kryshtafovych A, Huang S-Y, Schneidman-Duhovny D, Sali A, Segura J, Fernandez-Fuentes N, Viswanath S, Elber R, Grudinin S, Popov P, Neveu E, Lee H, Baek M, Park S, Heo L, Rie Lee G, Seok C, Qin S, Zhou H-X, Ritchie DW, Maigret B, Devignes M-D, Ghoorah A, Torchala M, Chaleil RAG, Bates PA, Ben-Zeev E, Eisenstein M, Negi SS, Weng Z, Vreven T, Pierce BG, Borrman TM, Yu J, Ochsenbein F, Guerois R, Vangone A, Rodrigues JPGLM, van Zundert G, Nellen M, Xue L, Karaca E, Melquiond ASJ, Visscher K, Kastritis PL, Bonvin AMJJ, Xu X, Qiu L, Yan C, Li J, Ma Z, Cheng J, Zou X, Shen Y, Peterson LX, Kim H-R, Roy A, Han X, Esquivel-Rodriguez J, Kihara D, Yu X, Bruce NJ, Fuller JC, Wade RC, Anishchenko I, Kundrotas PJ, Vakser IA, Imai K, Yamada K, Oda T, Nakamura T, Tomii K, Pallara C, Romero-Durana M, Jiménez-García B, Moal IH, Férnandez-Recio J, Joung JY, Kim JY, Joo K, Lee J, Kozakov D, Vajda S, Mottarella S, Hall DR, Beglov D, Mamonov A, Xia B, Bohnuud T, Del Carpio CA, Ichiishi E, Marze N, Kuroda D, Roy Burman SS, Gray JJ, Chermak E, Cavallo L, Oliva R, Tovchigrechko A, Wodak SJ, Prediction of homoprotein and heteroprotein complexes by protein docking and template-based modeling: A CASP-CAPRI experiment, Proteins. 84 Suppl 1 (2016) 323–348. doi:10.1002/prot.25007. [PubMed: 27122118]

[35]. Lensink MF, Velankar S, Wodak SJ, Modeling protein-protein and protein-peptide complexes: CAPRI 6th edition, Proteins. 85 (2017) 359–377. doi:10.1002/prot.25215. [PubMed: 27865038]

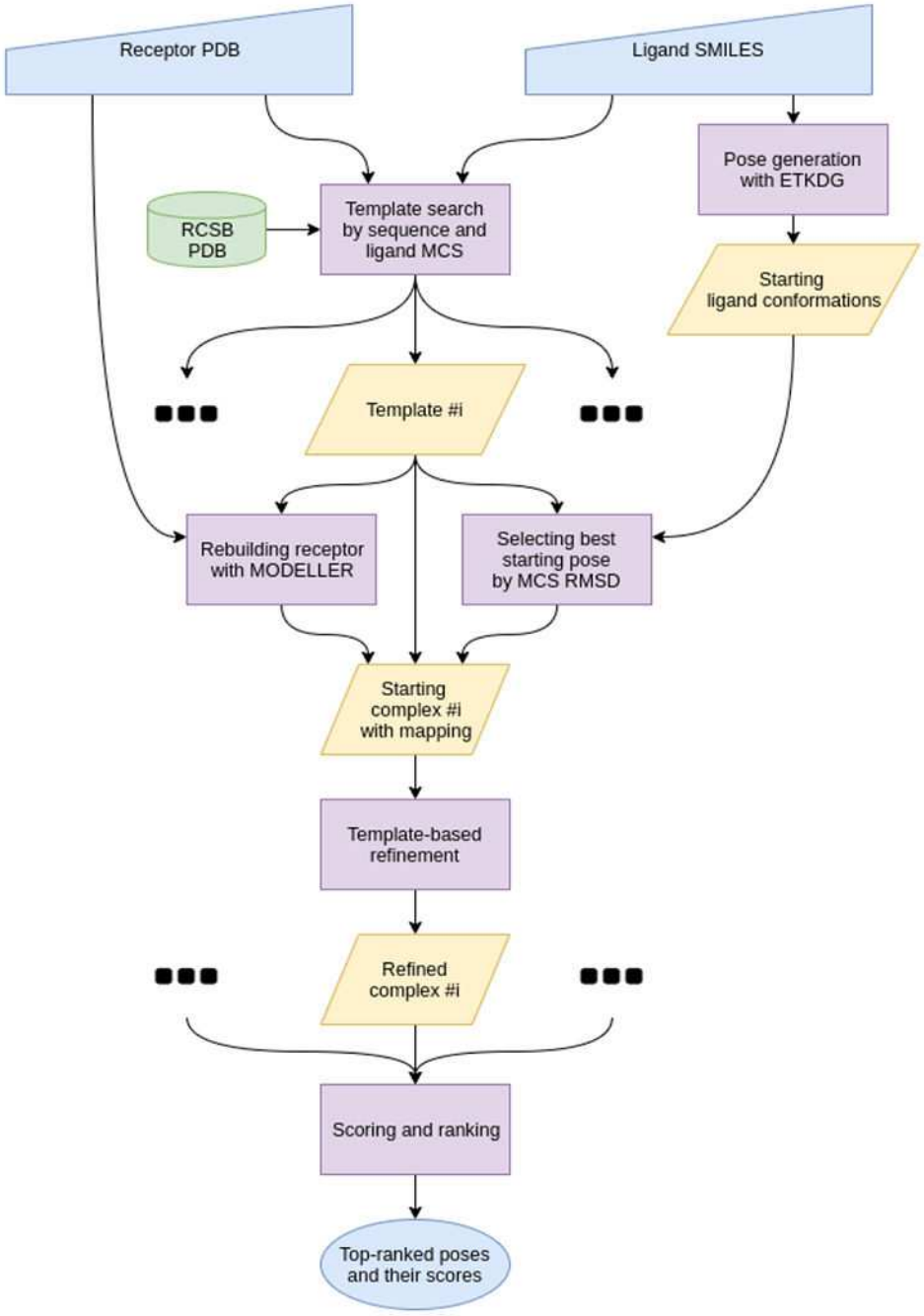

Fig. 1.

The general overview of the pipeline of LigTBM server. The steps between initial template search and the final scoring are performed independently for each template found, indicated by ellipses.

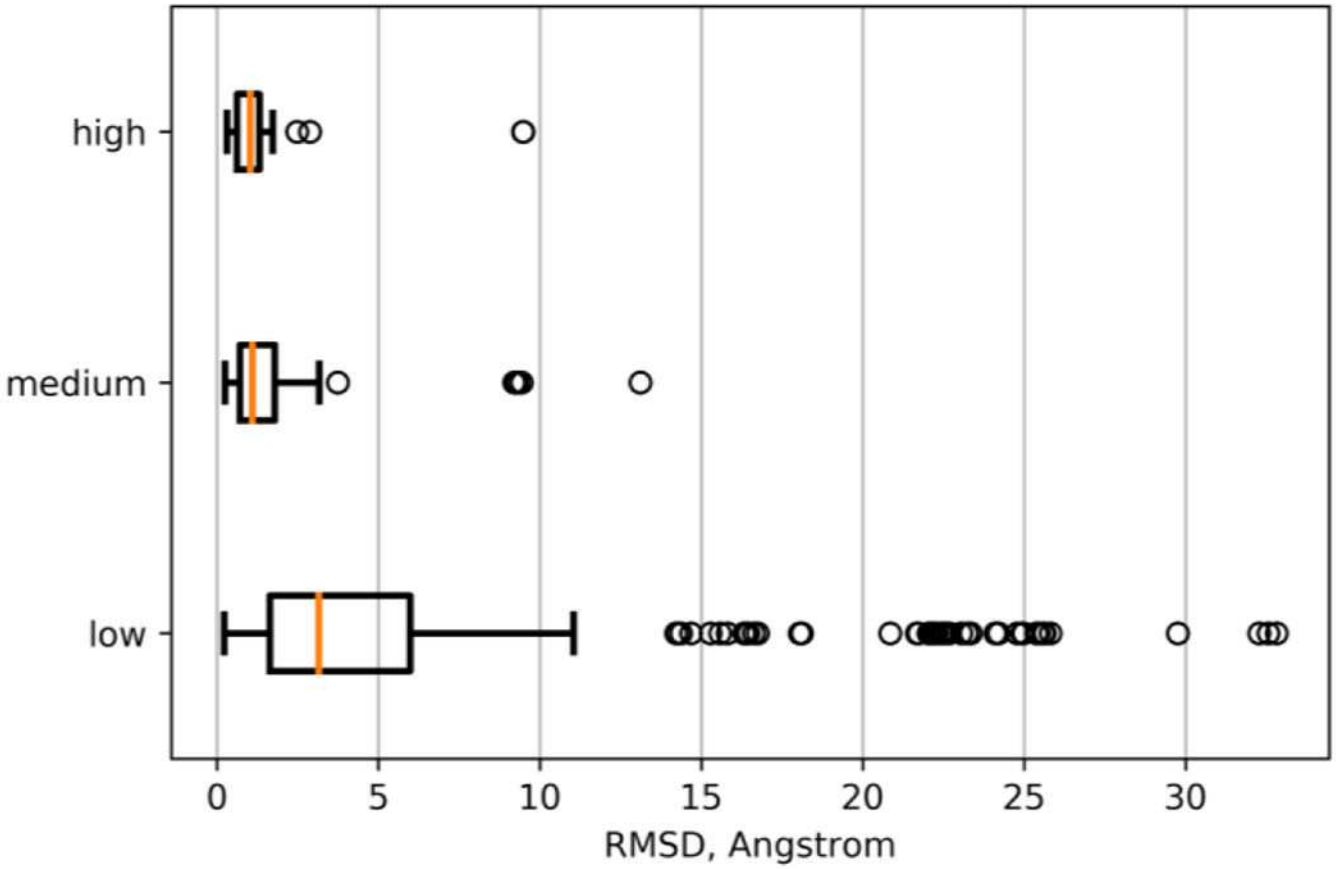

Fig. 2.
The boxplot of RMSD values for all produced models for the ANNS set.

Table 1.

Summary of the server performance on selected benchmarks

	BACE	ANNS
Total # of targets	20	65
Templates found	20 (100%)	59 (91%)
Top-1 <2Å	17 (85%)	44 (68%)
Top-5 <2Å	19 (95%)	47 (72%)