

## eModel-BDB: A database of comparative structure models of drug-target interactions from the Binding Database

--Manuscript Draft--

<b>Manuscript Number:</b>	GIGA-D-17-00258R1	
<b>Full Title:</b>	eModel-BDB: A database of comparative structure models of drug-target interactions from the Binding Database	
<b>Article Type:</b>	Research	
<b>Funding Information:</b>	National Institute of General Medical Sciences (R35GM119524)	Dr. Michal Brylinski
<b>Abstract:</b>	<p>Background. The structural information on proteins in their ligand-bound conformational state is invaluable for protein function studies and rational drug design. Compared to the number of available sequences, the repertoire of the experimentally determined structures of holo-proteins is not only limited, but also these structures do not always include pharmacologically relevant compounds at their binding sites. In addition, binding affinity databases provide vast quantities of information on interactions between drug-like molecules and their targets, however, often lacking structural data. On that account, there is a need for computational methods to complement existing repositories by constructing the atomic-level models of drug-protein assemblies that will not be determined experimentally in the near future.</p> <p>Results. We created eModel-BDB, a database of 200,005 comparative models of drug-bound proteins based on interaction data obtained from the Binding Database. Complex models in eModel-BDB were generated with a collection of the state-of-the-art techniques, including protein meta-threading, template-based structure modeling, refinement and binding site detection, and ligand similarity-based docking. In addition to a rigorous quality control maintained during dataset generation, a subset of weakly homologous models were selected for the retrospective validation against experimental structural data recently deposited to the Protein Data Bank. Validation results indicate that eModel-BDB contains models that are accurate not only at the global protein structure level, but also with respect to the atomic details of bound ligands.</p> <p>Conclusions. Freely available eModel-BDB can be used to support structure-based drug discovery and repositioning, drug target identification, and protein structure determination.</p>	
<b>Corresponding Author:</b>	Michal Brylinski, Ph.D. Louisiana State University Baton Rouge, Louisiana UNITED STATES	
<b>Corresponding Author Secondary Information:</b>		
<b>Corresponding Author's Institution:</b>	Louisiana State University	
<b>Corresponding Author's Secondary Institution:</b>		
<b>First Author:</b>	Michal Brylinski, Ph.D.	
<b>First Author Secondary Information:</b>		
<b>Order of Authors:</b>	Michal Brylinski, Ph.D.	
	Misagh Naderi	
	Rajiv Gandhi Govindaraj	
<b>Order of Authors Secondary Information:</b>		
<b>Response to Reviewers:</b>	Our responses are provided in a separate document "ResponseLetter.pdf" uploaded as Supplementary Material.	
<b>Additional Information:</b>		

Question	Response
<p>Are you submitting this manuscript to a special series or article collection?</p>	<p>No</p>
<p><b>Experimental design and statistics</b></p> <p>Full details of the experimental design and statistical methods used should be given in the Methods section, as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>. Information essential to interpreting the data presented should be made available in the figure legends.</p> <p>Have you included all the information requested in your manuscript?</p>	<p>Yes</p>
<p><b>Resources</b></p> <p>A description of all resources used, including antibodies, cell lines, animals and software tools, with enough information to allow them to be uniquely identified, should be included in the Methods section. Authors are strongly encouraged to cite <a href="#">Research Resource Identifiers</a> (RRIDs) for antibodies, model organisms and tools, where possible.</p> <p>Have you included the information requested as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	<p>Yes</p>
<p><b>Availability of data and materials</b></p> <p>All datasets and code on which the conclusions of the paper rely must be either included in your submission or deposited in <a href="#">publicly available repositories</a> (where available and ethically appropriate), referencing such data using a unique identifier in the references and in the “Availability of Data and Materials” section of your manuscript.</p> <p>Have you have met the above requirement as detailed in our <a href="#">Minimum Standards Reporting Checklist</a>?</p>	<p>Yes</p>

1  
2  
3  
4 **eModel-BDB: A database of comparative structure models of drug-target interactions from**  
5  
6 **the Binding Database**

7  
8  
9  
10 by

11  
12  
13  
14 Misagh Naderi<sup>1</sup>, Rajiv Gandhi Govindaraj<sup>1</sup> and Michal Brylinski<sup>1,2\*</sup>  
15  
16

17  
18 <sup>1</sup>Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA

19  
20 <sup>2</sup>Center for Computation & Technology, Louisiana State University, Baton Rouge, LA 70803, USA  
21

22  
23  
24 \* Corresponding author: Michal Brylinski  
25

26  
27 Email addresses: [mnader5@lsu.edu](mailto:mnader5@lsu.edu) (Misagh Naderi)

28  
29 [rgovinda@lsu.edu](mailto:rgovinda@lsu.edu) (Rajiv Gandhi Govindaraj)

30  
31 [michal@brylinski.org](mailto:michal@brylinski.org) (Michal Brylinski)  
32  
33

34  
35 Phone: (225) 578-2791

36  
37 Fax: (225) 578-2597  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 **Abstract**

5  
6 **Background.** The structural information on proteins in their ligand-bound conformational state  
7 is invaluable for protein function studies and rational drug design. Compared to the number of  
8 available sequences, the repertoire of the experimentally determined structures of holo-  
9 proteins is not only limited, but also these structures do not always include pharmacologically  
10 relevant compounds at their binding sites. In addition, binding affinity databases provide vast  
11 quantities of information on interactions between drug-like molecules and their targets,  
12 however, often lacking structural data. On that account, there is a need for computational  
13 methods to complement existing repositories by constructing the atomic-level models of drug-  
14 protein assemblies that will not be determined experimentally in the near future.  
15

16  
17  
18  
19  
20  
21  
22  
23 **Results.** We created eModel-BDB, a database of 200,005 comparative models of drug-bound  
24 proteins based on interaction data obtained from the Binding Database. Complex models in  
25 eModel-BDB were generated with a collection of the state-of-the-art techniques, including  
26 protein meta-threading, template-based structure modeling, refinement and binding site  
27 detection, and ligand similarity-based docking. In addition to a rigorous quality control  
28 maintained during dataset generation, a subset of weakly homologous models were selected  
29 for the retrospective validation against experimental structural data recently deposited to the  
30 Protein Data Bank. Validation results indicate that eModel-BDB contains models that are  
31 accurate not only at the global protein structure level, but also with respect to the atomic  
32 details of bound ligands.  
33

34  
35  
36  
37  
38  
39  
40  
41  
42  
43 **Conclusions.** Freely available eModel-BDB can be used to support structure-based drug  
44 discovery and repositioning, drug target identification, and protein structure determination.  
45

46  
47  
48  
49 **Keywords:** eModel-BDB, eThread, eFindSite, BindingDB, homology modeling, comparative  
50 modeling, binding pocket prediction, similarity-based docking, protein function, drug targets  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Background

Structural bioinformatics is becoming an increasingly important component of modern drug discovery. Despite significant advances in experimental methods to acquire protein structures, such as X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy, technical limitations and expensive procedures make it unlikely to have the experimental structures of all known protein sequences in the near future. For example, more than 100 million gene products are included in the Reference Sequence Database [1] as of March 2018. In contrast, the number of experimentally determined protein structures in the Protein Data Bank (PDB) [2] is 138,678, which reduces to 51,482 structures after removing similar proteins at 95% sequence identity. Genome sequencing currently produces as many as 13 million protein sequences each year, whereas only 8,872 protein structures are solved experimentally at the same time on average. Since this disparity between the number of available sequences and structures will likely continue to grow, **high-throughput** computational modeling is expected to play a significant role in biomedical sciences by generating 3D models for those proteins whose structures will not be determined in the near future.

In addition to protein sequence and structure repositories, the Binding Database (BindingDB) provides comprehensive information on interactions between small, drug-like molecules and proteins considered to be drug targets collected from affinity measurements [3]. The BindingDB can be used to identify protein targets for small molecules and bioactive compounds for new proteins, as well as to conduct virtual screening with ligand-based methods. As of March 2018, BindingDB contains 1,433,947 binding data, however, only 2,291 ligand-protein crystal structures with BindingDB affinity measurements are available in the PDB. To bridge this gap, we created eModel-BDB, a new database of 200,005 high-quality drug-protein complex models involving 108,363 unique drug-like compounds and 2,791 proteins from the BindingDB. This repository was constructed with a state-of-the-art protocol to generate protein models in their ligand-bound conformational state, employing meta-threading, pocket detection, and protein structure and ligand chemical alignment techniques. eModel-BDB significantly expands the current structural information on known drug-protein complexes.

1  
2  
3  
4 To fully appreciate the immensity of the structural data included in eModel-BDB, we  
5 estimate the time required to solve an equal number of drug-protein assemblies. Figure 1  
6 shows that at the current pace, 2,447 ligand-bound protein structures containing 607 non-  
7 redundant complexes are deposited to the PDB each month. Therefore, it would take about 329  
8 months for 200,005 unique complex structures to be determined experimentally. In contrast to  
9 other databases comprising protein models in the unbound conformational state generated  
10 through traditional structure modeling [4, 5], eModel-BDB includes annotated structure models  
11 of drug-protein complexes with known binding affinities. It provides high-quality data to  
12 support structure-based drug discovery as well as repurposing of known drugs based on binding  
13 pocket and ligand similarities. In addition, the information provided by eModel-BDB can be  
14 utilized to facilitate experimental structure determination by developing protocols to stabilize  
15 proteins with ligands. The protocol to construct eModel-BDB described in this communication is  
16 based entirely on open source software to ensure that any researcher is able to produce new  
17 holo-protein models as more data becomes available in the PDB and BindingDB.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

## 33 **Methods**

### 34 ***Protein structure modeling***

35 Drug-bound protein complexes in eModel-BDB are generated with a template-based approach.  
36 The first phase is to construct structure models for **single protein chains** 50-999 amino acids in  
37 length obtained from BindingDB with eThread [6], which supports both close and remote  
38 homology modeling. eThread employs Modeller, a commonly used comparative modeling  
39 program [7], to build apo-protein structures based on alignments produced by three fold  
40 recognition algorithms, HH-suite [8], SparksX [9], and RaptorX [10]. Subsequently, side-chain  
41 positions and hydrogen-bonding networks in the initial models are improved with ModRefiner,  
42 a program to refine protein structures at the atomic-level with a composite physics- and  
43 knowledge-based force field [11]. The quality assessment of refined models is carried out with  
44 ModelEvaluator [12] in terms of the estimated Global Distance Test score (GDT-score). Out of  
45 5,501 BindingDB proteins, 4,906 were assigned an estimated GDT-score of  $\geq 0.4$  **indicating good**  
46 **quality models [13, 14].**  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

### ***Ligand-binding site identification***

Confident structure models with a GDT-score of  $\geq 0.4$  are further annotated with binding pockets and residues by eFindSite [15], which also computes a calibrated pocket confidence score. eFindSite detected 2,922 high-, 644 moderate-, and 776 low-confidence pockets in the eThread models of BindingDB targets. At this point, BindingDB drugs can be assigned to the predicted pockets with fingerprint-based virtual screening. Specifically, for a given drug-target pair in the BindingDB, we compute a rank of the drug against pockets detected by eFindSite, where the remaining BindingDB compounds are used as the background library. eFindSite conducts virtual screening with a set of molecular fingerprints and physicochemical properties calculated for ligands extracted from weakly homologous template structures [16]. Top one, two and three pockets are considered for high-, moderate- and low-confidence targets, respectively. A drug matches the predicted pocket if it is ranked within the top 20% of the screening library. With this protocol, we matched 108,363 drugs to binding pockets identified in their target proteins.

### ***Similarity-based ligand docking***

In the next phase, drug molecules are positioned within the predicted pockets with a two-step similarity-based docking protocol. This procedure exploits a significant structural conservation of ligand binding modes across remote homologs [17]. First, globally similar ligand-bound templates from the PDB, identified by eFindSite to have a similar pocket as the BindingDB protein, are superimposed onto the apo-model. Proteins are aligned with Fr-TM-align [18] employing the Template Modeling score (TM-score) [19] to measure the global structure similarity. Subsequently, the BindingDB compound is aligned onto the template-bound ligand in order to place it in the predicted pocket of the apo-model. Here, we use chemical alignments constructed with kcombu [20], which also reports the chemical similarity between the BindingDB compound and the template-bound ligand measured by the Tanimoto coefficient (TC). Since a perfect case corresponds to both a TM-score and a TC of 1.0, we

1  
2  
3  
4 introduce a new metric, the Perfect Match Distance (PMD), combining protein structure and  
5  
6 ligand chemical similarity values:  
7  
8  
9

$$10 \quad PMD = \sqrt{(1 - TM\text{-score})^2 + (1 - TC)^2} \quad \text{Eq. 1.}$$

11  
12  
13  
14

15 PMD is simply the Cartesian distance from the perfect match in the TM-score/TC space.  
16  
17 In order to generate only high-quality holo-models, those cases with a PMD of >0.6 are  
18  
19 excluded from the modeling process. This PMD cutoff was chosen to ensure that TM-score and  
20  
21 TC for the selected templates are always above their individual significance threshold values of  
22  
23 0.4 [19, 20]. Further, for those cases having multiple ligand-bound templates satisfying the PMD  
24  
25 criterion of  $\leq 0.6$ , a template with the shortest PMD is selected to build the holo-model of the  
26  
27 BindingDB complex.  
28  
29  
30

### 31 ***Complex structure refinement and assessment***

32  
33 In the final phase, protein models are rebuilt in the presence of the docked BindingDB  
34  
35 compounds with Modeller. To make sure that the binding site is remodeled to accommodate  
36  
37 the specific ligand, binding residues identified by eFindSite are removed from the initial model  
38  
39 while enforcing the presence of secondary structure predicted by PSIPRED [21]. The resulting  
40  
41 models are further annotated with the ligand-protein interaction score according to the  
42  
43 Distance-scaled Finite Ideal-gas REference (DFIRE) potential [22]. The eModel-BDB database  
44  
45 contains atomic-level structure models of 200,005 drug-protein interactions from BindingDB,  
46  
47 comprising 2,791 non-redundant proteins and 108,363 drug-like compounds. **The list of**  
48  
49 **eModel-BDB complexes is provided as Supplementary File S1.**  
50  
51

### 52 **Analyses**

#### 53 ***Data quality control***

54  
55  
56 The quality control is pertinent to both protein structure modeling as well as binding site  
57  
58 prediction. The quality of protein models is assessed with ModelEvaluator employing various  
59  
60 structural features to compute the absolute quantitative score with a support vector  
61  
62  
63  
64  
65



1  
2  
3  
4 regression. This approach assigns the GDT-score to a model by analyzing its secondary  
5 structure, relative solvent accessibility, contact map, and  $\beta$ -sheet structure. It has been  
6 demonstrated that GDT-scores estimated by ModelEvaluator for template-based models are  
7 highly correlated with the actual values with the Pearson correlation coefficient of 0.82 [12].  
8 The first violin in Figure 2 shows that eModel-BDB contains close and remote homology models  
9 with the median target-template sequence identity of 63%. The second violin indicates that the  
10 vast majority of these structures are accurate with the median estimated GDT-score for  
11 BindingDB proteins is 0.62. Further, as many as 78% of binding sites predicted by eFindSite to  
12 match BindingDB ligands have a high confidence of  $>0.8$ . We showed previously that confidence  
13 scores of  $>0.8$  assigned by eFindSite correspond to the Matthews correlation coefficient (MCC)  
14 [23] of  $\geq 0.6$  for predicted binding residues [15]. On that account, we expect that the majority of  
15 binding sites for BindingDB drugs are correctly annotated as well. Note that in contrast to other  
16 pocket predictors, eFindSite annotations and confidence assignments are, to some extent,  
17 independent of the accuracy of protein models.  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

31 The quality of complex models is controlled by imposing thresholds on the chemical  
32 similarity between BindingDB and PDB ligands as well as the global structure similarity between  
33 eThread models and ligand-bound templates from the PDB. Figure 3A shows the distribution of  
34 both parameters across eModel-BDB models. Encouragingly, the median TM-score and TC for  
35 ligand-bound templates used to build eModel-BDB are as high as 0.81 and 0.67, respectively.  
36 Previous studies show that the probability for a protein pair to belong to the same fold is 98%  
37 when the TM-score is close to 0.8 [24]. In addition, it was demonstrated that the root-mean-  
38 square deviation (RMSD) over ligand non-hydrogen atoms for similarity-based docking  
39 conducted with the TC in the range of 0.6-0.8 is typically 2-3 Å [25]. TM-score and TC values are  
40 combined into a single assessment score, the PMD, measuring the distance from the perfect  
41 match. Therefore, selecting template proteins with a lower TM-score to BindingDB targets  
42 requires their ligands to have a high TC and vice versa, selecting PDB ligands with a lower  
43 chemical similarity to BindingDB molecules requires a high global structure similarity between  
44 proteins. Figure 3B shows that the median PMD for eModel-BDB complex models is 0.46.  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Data validation

In addition to the rigorous quality control maintained during dataset generation, eModel-BDB is validated retrospectively against experimental structures recently deposited to the PDB. The structure models of BindingDB interactions have been constructed with the PDB library as of January 31<sup>st</sup> 2017, therefore, we examined 7,012 experimental structures deposited to the PDB after February 2017 to validate eModel-BDB structures. The validation protocol is made more challenging by including only remote homology models with a template-target sequence identity of <40%. In order to maximize the validation coverage, we use the recently determined structures of eModel-BDB targets and their homologs with at least 40% sequence identity. Recently solved experimental structures selected from the PDB validate 161 eThread models and 952 BindingDB reaction set IDs, comprising 39 target proteins, 52 pockets, and 881 compounds. The list of validation pairs is given in Supplementary File S2.

## Protein structure modeling

The first violin in Figure 4 shows that the median TM-score of eModel-BDB vs. experimental structures is 0.85 with as many as 98.1% of the models having a TM-score of  $\geq 0.4$ . Clearly, the majority of structures are modeled by eThread with a high accuracy. A representative example of the correctly predicted target structure is dihydrofolate reductase (DHFR) from *Streptococcus pyogenes* build on the crystal structure of DHFR from *Streptococcus pneumoniae* (PDB-ID: 3ix9, chain B, 36% sequence identity to the target) [26]. The eThread model, whose estimated GDT-score is 0.92, was then used to construct a structure model for the BindingDB reactant set ID 00267770 consisting of DHFR complexed with BDBM50329610. This model is validated against the crystal structure of DHFR-UCP1106 from *Staphylococcus aureus* (PDB-ID: 5isp, chain A, 43% sequence identity to the target) released on 2017-06-28 [27]. Figure 5 shows the predicted weakly homologous model of DHFR-BDBM50329610 (purple) superposed on the experimental structure of DHFR-UCP1106 (gold). The eModel-BDB model is indeed highly accurate with a TM-score of 0.95 and a C $\alpha$ -RMSD of 1.23 Å over 157 aligned residues. In addition, Figure 6A shows that the estimated GDT-score employed in this study as the confidence measure to control the quality of protein models correlates with the accuracy of

1  
2  
3  
4 final models evaluated with the TM-score. On that account, the estimated GDT-score provides a  
5 robust quality assessment measure to control the quality of models in eModel-BDB.  
6  
7  
8  
9

### 10 ***Binding pocket prediction***

11  
12 The accuracy of pocket prediction is validated by superposing the experimental holo  
13 structure onto the eModel-BDB model and then calculating the distance between the  
14 geometric center of a bound ligand in the experimental complex and the pocket center  
15 predicted by eFindSite in the model. The second violin in Figure 4 shows that the median pocket  
16 distance is 5.5 Å with 59.6% of pockets predicted within 6 Å, therefore, most eFindSite  
17 annotations are accurate. A binding site in the model of vitamin D receptor (VDR) is a  
18 representative example of a pocket predicted with eFindSite. This model was constructed by  
19 eThread based on human retinoic acid receptor RXR-alpha (PDB-ID: 4nqa, chain H, 38%  
20 sequence identity to the target) [28]. Although the GDT-score estimated for the VDR model is  
21 0.62 indicating a moderately accurate structure, the top-ranked binding site annotated by  
22 eFindSite is assigned a high confidence of 94.2%. Figure 7 shows the VDR model (purple  
23 ribbons) superposed onto the crystal structure of vitamin D3 receptor A (gold ribbons)  
24 complexed with a synthetic analog of 1 $\alpha$ ,25-dihydroxyvitamin D3 (PDB-ID: 5nky, chain A, 66%  
25 sequence identity to the target) released on 2017-05-24 [29]. Not only the VDR model aligns  
26 well to the experimental structure with a TM-score of 0.90 and a C $\alpha$ -RMSD of 2.13 Å over 235  
27 residues, but also the predicted pocket center (purple sphere) is only 5.5 Å away from the  
28 geometric center of vitamin D analog (gold sphere).  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

### 46 ***Ligand docking***

47  
48 Finally, we calculate the RMSD over non-hydrogen atoms between the BindingDB drug  
49 in the eModel-BDB structure and the bound ligand in the superposed experimental complex.  
50 The first violin in Figure 8 shows that the median ligand RMSD is 2.6 Å and it is  $\leq 3$  Å for 58.1% of  
51 BindingDB compounds. The model of the BindingDB reactant set ID 50103430 consisting of  
52 cytochrome P450 17A1 (CYP17A1) complexed with BDBM50061174 is selected to exemplify the  
53 accuracy of complex structures in eModel-BDB. The model of CYP17A1 built on the crystal  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 structure of human microsomal cytochrome P450 2A6 (PDB-ID: 1z11, chain A, 29% sequence  
5 identity to the target) [30] by eThread is assigned an estimated GDT-score of 0.69.  
6 Subsequently, the complex model of CYP17A1-BDBM50061174 was constructed by similarity-  
7 based docking employing the crystal structure of CYP17A1 bound to abiraterone, a steroidal  
8 prostate cancer drug (PDB-ID: 3ruk, chain D) [31]. The CYP17A1-abiraterone complex was  
9 selected as the best ligand-bound template based on the high TM-score of 0.84 and TC of 0.89,  
10 yielding the shortest PMD of 0.19. Figure 9 shows the validation of the modeled CYP17A1-  
11 BDBM50061174 by the experimental structure of CYP17A1-(R)-orterone (PDB-ID: 5irq, chain B,  
12 64% sequence identity to the target) released on 2017-03-15 [32]. Kcombu reports a significant  
13 chemical alignment between steroidal BDBM50061174 and nonsteroidal (R)-orterone with a TC  
14 of 0.54 (Figure 9A). Upon the superposition of CYP17A1 proteins, the RMSD between  
15 BDBM50061174 docked to the model and (R)-orterone bound in the experimental structure  
16 calculated over the chemical alignment reported by kcombu is 2.95 Å (Figure 9B). These results  
17 verify that the computer-generated CYP17A1-BDBM50061174 model for the BindingDB  
18 reactant set ID 50103430 is correct.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 Similarity-based docking procedure employed to construct ligand-bound structures in  
34 eModel-BDB superposes target ligands onto template molecules selected from the PDB  
35 according to the chemical alignment reported by kcombu. One may expect that superposing  
36 target compounds onto chemically similar template ligands yields more accurate binding poses  
37 than those generated from chemically less similar template molecules. Indeed, Figure 6B shows  
38 that the target-template chemical similarity measured with the TC correlates with the docking  
39 accuracy evaluated with the RMSD of ligand poses constructed based on target-template  
40 alignments. These results are in line with other studies reporting that the average RMSD values  
41 for similarity-based docking methods are generally below 2 Å when the target-template  
42 similarities are above 0.7 [33]. The performance of similarity-based docking employed to  
43 construct eModel-BDB is also compared to that of AutoDock Vina [34] and rDock [35]. In  
44 contrast to the median ligand RMSD of 2.6 Å for eModel-BDB complexes, the median RMSD  
45 values for BindingDB drugs docked to eFindSite pockets with AutoDock Vina and rDock are 6.7 Å  
46 and 7.2 Å, respectively (Figure 8). We note that similarity-based docking was demonstrated to  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 outperform traditional docking when the target-template similarity is greater than 0.4 [33],  
5 which was employed as the TC threshold to construct eModel-BDB complex models. Overall,  
6 the quality assessment as well as independently obtained validation results demonstrate that  
7 the eModel-BDB database contains high-quality models closely resembling experimentally  
8 determined structures, not only at the global structure level, but also at the level of binding  
9 pockets and bound ligands.  
10  
11  
12  
13  
14  
15  
16  
17

## 18 Discussion

19 eModel-BDB is generated to support rational drug development projects. These data can  
20 directly aid structure-based drug discovery pipelines and protein function analysis by providing  
21 atomic-level models of a large set of drug-protein interactions with known affinities curated in  
22 the BindingDB. An important application of eModel-BDB is computational drug repositioning,  
23 i.e. finding new indications for existing drugs [36]. Although drug repurposing holds a significant  
24 promise to speed up drug development, particularly for diseases considered to be unprofitable,  
25 its major bottleneck is the scarce structural information on druggable pockets. On that account,  
26 a diverse dataset of small, drug-like molecules bound to high-quality models with accurately  
27 annotated pockets provide an invaluable resource for drug repositioning employing sequence  
28 order-independent pocket matching algorithms [37-40]. It is noteworthy that computational  
29 drug repurposing has suggested new opportunities to combat tuberculosis [41, 42], malaria  
30 [43], and rare diseases [44, 45].  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

43 Binding sites in eModel-BDB can also be matched to pockets predicted in potential drug  
44 targets in order to determine whether these proteins are druggable or not. If a new pocket  
45 aligns well with drug-bound pockets in eModel-BDB then it is likely going to be druggable. That  
46 being the case, our data can be utilized right at the outset of drug discovery, in the target  
47 identification phase. Finally, ligand binding can significantly help stabilize a protein, particularly  
48 from the point of view of the conformational stability [46]. eModel-BDB can, therefore, inform  
49 crystallography efforts by suggesting possible compounds binding to certain protein targets at  
50 either the active or allosteric sites in order to increase the chances of successful crystallization.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## Availability of supporting data and materials

Structure models in eModel-BDB are named according to the BindingDB reactant set IDs, which can be obtained by searching the BindingDB at <https://www.bindingdb.org>. This procedure is illustrated in Figure 10. The BindingDB can be searched either by protein and compound names (Figure 10A) or by the target sequence (Figures 10B and 10C). Next, the complex of interest can be selected from the list of hits (Figure 10D) in order to download the corresponding SDfile of the complex (Figure 10E). The BindingDB reactant set ID, e.g. 00267770, is stored inside the SDfile (Figure 10F). The reactant set ID can then be used to find the detailed information on the BindingDB website, e.g. [https://www.bindingdb.org/jsp/dbsearch/Summary\\_ki.jsp?reactant\\_set\\_id=00267770](https://www.bindingdb.org/jsp/dbsearch/Summary_ki.jsp?reactant_set_id=00267770) (Figure 10G) as well as access the structure model in eModel-BDB, e.g. [http://www.brylinski.org/pub/eModelBDB.php?reactant\\_set\\_id=00267770](http://www.brylinski.org/pub/eModelBDB.php?reactant_set_id=00267770) (Figure 10H).

## Declarations

### *List of abbreviations*

BindingDB, Binding Database; DFIRE, Distance-scaled Finite Ideal-gas REference; GDT-score, Global Distance Test score; MCC, Matthews correlation coefficient; PMD, Perfect Match Distance; PDB, Protein Data Bank; RMSD, root-mean-square deviation; TC, Tanimoto coefficient; TM-score, Template Modeling score

### *Competing interests*

The authors declare that they have no competing interests.

### *Funding*

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM119524.

### *Authors' contributions*

1  
2  
3  
4 MB prepared protein models, annotated binding pockets, and validated protein structures and  
5  
6 pockets. MN constructed, refined, and validated drug-bound models. RGG prepared case  
7  
8 studies. MN and MB wrote the paper.  
9

## 10 11 **Acknowledgements**

12  
13  
14 Authors are grateful to Louisiana State University for providing computing resources.  
15  
16  
17

## 18 **References**

- 19  
20 1. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference  
21 sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and  
22 functional annotation. *Nucleic Acids Res.* 2016;44 D1:D733-45.  
23 doi:10.1093/nar/gkv1189.
- 24  
25 2. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data  
26 Bank. *Nucleic Acids Res.* 2000;28 1:235-42.
- 27  
28 3. Liu T, Lin Y, Wen X, Jorissen RN and Gilson MK. BindingDB: a web-accessible database of  
29 experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 2007;35  
30 Database issue:D198-201. doi:10.1093/nar/gkl999.
- 31  
32 4. Castrignano T, De Meo PD, Cozzetto D, Talamo IG and Tramontano A. The PMDB Protein  
33 Model Database. *Nucleic Acids Res.* 2006;34 Database issue:D306-9.  
34 doi:10.1093/nar/gkj105.
- 35  
36 5. Sanchez R, Pieper U, Mirkovic N, de Bakker PI, Wittenstein E and Sali A. MODBASE, a  
37 database of annotated comparative protein structure models. *Nucleic Acids Res.*  
38 2000;28 1:250-3.
- 39  
40 6. Brylinski M and Lingam D. eThread: a highly optimized machine learning-based  
41 approach to meta-threading and the modeling of protein tertiary structures. *PLoS One.*  
42 2012;7 11:e50200. doi:10.1371/journal.pone.0050200.
- 43  
44 7. Sali A and Blundell TL. Comparative protein modelling by satisfaction of spatial  
45 restraints. *J Mol Biol.* 1993;234 3:779-815. doi:10.1006/jmbi.1993.1626.
- 46  
47 8. Remmert M, Biegert A, Hauser A and Soding J. HHblits: lightning-fast iterative protein  
48 sequence searching by HMM-HMM alignment. *Nat Methods.* 2011;9 2:173-5.  
49 doi:10.1038/nmeth.1818.
- 50  
51 9. Yang Y, Faraggi E, Zhao H and Zhou Y. Improving protein fold recognition and template-  
52 based modeling by employing probabilistic-based matching between predicted one-  
53 dimensional structural properties of query and corresponding native properties of  
54 templates. *Bioinformatics.* 2011;27 15:2076-82. doi:10.1093/bioinformatics/btr350.
- 55  
56 10. Ma J, Wang S, Zhao F and Xu J. Protein threading using context-specific alignment  
57 potential. *Bioinformatics.* 2013;29 13:i257-65. doi:10.1093/bioinformatics/btt210.
- 58  
59 11. Xu D and Zhang Y. Improving the physical realism and structural accuracy of protein  
60 models by a two-step atomic-level energy minimization. *Biophys J.* 2011;101 10:2525-  
61 34. doi:10.1016/j.bpj.2011.10.024.  
62  
63  
64  
65

12. Wang Z, Tegge AN and Cheng J. Evaluating the absolute quality of a single protein model using structural features and support vector machines. *Proteins*. 2009;75 3:638-47. doi:10.1002/prot.22275.
13. He Z, Alazmi M, Zhang J and Xu D. Protein structural model selection by combining consensus and single scoring methods. *PLoS One*. 2013;8 9:e74006. doi:10.1371/journal.pone.0074006.
14. Mirabello C, Adelfio A and Pollastri G. Reconstructing protein structures by neural network pairwise interaction fields and iterative decoy set construction. *Biomolecules*. 2014;4 1:160-80. doi:10.3390/biom4010160.
15. Brylinski M and Feinstein WP. eFindSite: improved prediction of ligand binding sites in protein models using meta-threading, machine learning and auxiliary ligands. *J Comput Aided Mol Des*. 2013;27 6:551-67. doi:10.1007/s10822-013-9663-5.
16. Feinstein WP and Brylinski M. eFindSite: Enhanced fingerprint-based virtual screening against predicted ligand binding sites in protein models. *Mol Inform*. 2014;33 2:135-50. doi:10.1002/minf.201300143.
17. Brylinski M and Skolnick J. FINDSITE: a threading-based approach to ligand homology modeling. *PLoS Comput Biol*. 2009;5 6:e1000405. doi:10.1371/journal.pcbi.1000405.
18. Pandit SB and Skolnick J. Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score. *BMC Bioinformatics*. 2008;9:531. doi:10.1186/1471-2105-9-531.
19. Zhang Y and Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins*. 2004;57 4:702-10. doi:10.1002/prot.20264.
20. Kawabata T. Build-up algorithm for atomic correspondence between chemical structures. *Journal of chemical information and modeling*. 2011;51 8:1775-87. doi:10.1021/ci2001023.
21. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*. 1999;292 2:195-202. doi:10.1006/jmbi.1999.3091.
22. Zhang C, Liu S, Zhu Q and Zhou Y. A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes. *J Med Chem*. 2005;48 7:2325-35. doi:10.1021/jm049314d.
23. Matthews BW. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta*. 1975;405 2:442-51.
24. Xu J and Zhang Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*. 2010;26 7:889-95. doi:10.1093/bioinformatics/btq066.
25. Brylinski M. Nonlinear scoring functions for similarity-based ligand docking and binding affinity prediction. *Journal of chemical information and modeling*. 2013;53 11:3097-112. doi:10.1021/ci400510e.
26. Lee J, Yennawar NH, Gam J and Benkovic SJ. Kinetic and structural characterization of dihydrofolate reductase from *Streptococcus pneumoniae*. *Biochemistry*. 2010;49 1:195-206. doi:10.1021/bi901614m.
27. Reeve SM, Scocchera E, Ferreira JJ, N GD, Keshipeddy S, Wright DL, et al. Charged propargyl-linked antifolates reveal mechanisms of antifolate resistance and inhibit trimethoprim-resistant MRSA strains possessing clinically relevant mutations. *J Med Chem*. 2016;59 13:6493-500. doi:10.1021/acs.jmedchem.6b00688.



- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65
28. Lou X, Toresson G, Benod C, Suh JH, Philips KJ, Webb P, et al. Structure of the retinoid X receptor alpha-liver X receptor beta (RXRalpha-LXRbeta) heterodimer on DNA. *Nat Struct Mol Biol.* 2014;21 3:277-81. doi:10.1038/nsmb.2778.
  29. Belorusova AY, Martinez A, Gandara Z, Gomez G, Fall Y and Rochel N. Structure-activity relationship study of vitamin D analogs with oxolane group in their side chain. *Eur J Med Chem.* 2017;134:86-96. doi:10.1016/j.ejmech.2017.03.081.
  30. Yano JK, Hsu MH, Griffin KJ, Stout CD and Johnson EF. Structures of human microsomal cytochrome P450 2A6 complexed with coumarin and methoxsalen. *Nat Struct Mol Biol.* 2005;12 9:822-3. doi:10.1038/nsmb971.
  31. DeVore NM and Scott EE. Structures of cytochrome P450 17A1 with prostate cancer drugs abiraterone and TOK-001. *Nature.* 2012;482 7383:116-9. doi:10.1038/nature10743.
  32. Petrunak EM, Rogers SA, Aube J and Scott EE. Structural and functional evaluation of clinically relevant inhibitors of steroidogenic cytochrome P450 17A1. *Drug Metab Dispos.* 2017;45 6:635-45. doi:10.1124/dmd.117.075317.
  33. Kawabata T and Nakamura H. 3D flexible alignment using 2D maximum common substructure: dependence of prediction accuracy on target-reference chemical similarity. *Journal of chemical information and modeling.* 2014;54 7:1850-63. doi:10.1021/ci500006d.
  34. Trott O and Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31 2:455-61. doi:10.1002/jcc.21334.
  35. Ruiz-Carmona S, Alvarez-Garcia D, Foloppe N, Garmendia-Doval AB, Juhos S, Schmidtke P, et al. rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol.* 2014;10 4:e1003571. doi:10.1371/journal.pcbi.1003571.
  36. Haupt VJ and Schroeder M. Old friends in new guise: repositioning of known drugs with structural bioinformatics. *Brief Bioinform.* 2011;12 4:312-26. doi:10.1093/bib/bbr011.
  37. Brylinski M. eMatchSite: sequence order-independent structure alignments of ligand binding pockets in protein models. *PLoS Comput Biol.* 2014;10 9:e1003829. doi:10.1371/journal.pcbi.1003829.
  38. Brylinski M. Local alignment of ligand binding sites in proteins for polypharmacology and drug repositioning. *Methods Mol Biol.* 2017;1611:109-22. doi:10.1007/978-1-4939-7015-5\_9.
  39. Konc J and Janezic D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics.* 2010;26 9:1160-8. doi:10.1093/bioinformatics/btq100.
  40. Ren J, Xie L, Li WW and Bourne PE. SMAP-WS: a parallel web service for structural proteome-wide ligand-binding site comparison. *Nucleic Acids Res.* 2010;38 Web Server issue:W441-4. doi:10.1093/nar/gkq400.
  41. Kinnings SL, Liu N, Buchmeier N, Tonge PJ, Xie L and Bourne PE. Drug discovery using chemical systems biology: repositioning the safe medicine Comtan to treat multi-drug and extensively drug resistant tuberculosis. *PLoS Comput Biol.* 2009;5 7:e1000423. doi:10.1371/journal.pcbi.1000423.

- 1  
2  
3  
4 42. Anand P and Chandra N. Characterizing the pocketome of Mycobacterium tuberculosis  
5 and application in rationalizing polypharmacological target selection. Sci Rep.  
6 2014;4:6356. doi:10.1038/srep06356.  
7  
8 43. Ramakrishnan G, Chandra N and Srinivasan N. Exploring anti-malarial potential of FDA  
9 approved drugs: an in silico approach. Malar J. 2017;16 1:290. doi:10.1186/s12936-017-  
10 1937-2.  
11  
12 44. Brylinski M, Naderi M, Govindaraj RG and Lemoine J. eRepo-ORP: Exploring the  
13 opportunity space to combat orphan diseases with existing drugs. J Mol Biol. 2018;in  
14 press doi:10.1016/j.jmb.2017.12.001.  
15  
16 45. Govindaraj RG, Naderi M, Singha M, Lemoine J and Brylinski M. Large-scale  
17 computational drug repositioning to find treatments for rare diseases. NPJ Syst Biol  
18 Appl. 2018;4:13. doi:10.1038/s41540-018-0050-7.  
19  
20 46. Deller MC, Kong L and Rupp B. Protein stability: a crystallographer's perspective. Acta  
21 Crystallogr F Struct Biol Commun. 2016;72 Pt 2:72-95.  
22 doi:10.1107/S2053230X15024619.  
23  
24

## 25 Figure captions

26  
27 **Figure 1. Deposition rate of ligand-bound structures to the Protein Data Bank.** The total  
28 number of protein chains binding small molecules (light gray squares and a dashed line) is  
29 counted at any point in time. The number of unique complex structures is obtained by  
30 clustering individual chains at 80% sequence identity (dark gray circles and a solid line).  $N_t$  and  
31  $N_u$  in the linear regression equations are the total and unique number of ligand-protein  
32 complexes, respectively, and  $m$  stands for month.  
33  
34  
35  
36  
37  
38  
39  
40

41 **Figure 2. Violin and box plots for model quality control.** The distribution of the target-template  
42 sequence identity (SeqId) and the Global Distance Test (GDT) score estimated for structure  
43 models. Horizontal yellow lines represent median values.  
44  
45  
46  
47  
48

49 **Figure 3. Similarities between target and holo-template proteins.** (A) The chemical similarity  
50 between BindingDB and PDB ligands measured with the Tanimoto coefficient (TC) is plotted  
51 against the global structure similarity of eThread models and ligand-bound templates from the  
52 PDB assessed by the TM-score. The 2D contour plot is generated by smoothing the data with  
53 the kernel density estimation technique. 1D histograms show the distribution of TC (top) and  
54 TM-score (right) values across eModel-BDB models. (B) Violin and box plot for the holo-  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 template Perfect Match Distance (PMD) combining TC and TM-score. The horizontal yellow line  
5 represents the median value.  
6  
7  
8  
9

10 **Figure 4. Violin and box plots for the distribution of validation scores.** The accuracy is assessed  
11 for remote homology complex models in the validation set. The global structure similarity is  
12 measured with the TM-score. The pocket distance is measured between the predicted pocket  
13 center and the geometric center of the ligand in the experimental structure superposed onto  
14 the eThread model. Horizontal yellow lines represent median values.  
15  
16  
17  
18  
19  
20  
21

22 **Figure 5. Representative example of a structure model constructed by eThread.** The model of  
23 dihydrofolate reductase (DHFR, purple ribbons) complexed with BDBM50329610 is superposed  
24 onto the crystal structure of homologous DHFR from *S. aureus* (gold ribbons) complexed with  
25 UCP1106. Ligands bound to target proteins are shown as solid sticks (BDBM50329610 is purple  
26 and UCP1106 is gold) with non-carbon atoms colored by atom type (O – red, N – blue).  
27  
28  
29  
30  
31  
32

33 **Figure 6. Analysis of structure modeling and ligand docking accuracy.** The accuracy is assessed  
34 for remote homology complex models in the validation set. (A) Accuracy of global structure  
35 prediction evaluated by the TM-score with respect to the estimated GDT-score. (B) Accuracy of  
36 similarity-based docking with respect to the chemical similarity between BindingDB and PDB  
37 ligands measured by the Tanimoto coefficient (TC). The ligand RMSD is calculated over non-  
38 hydrogen atoms according to the chemical alignment reported by kcombu. Solid red lines show  
39 the average prediction accuracy for binned GDT-score values in A and the chemical similarity in  
40 B. Dotted black lines mark the median TM-score in A and RMSD in B across all benchmarking  
41 cases.  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

53 **Figure 7. Representative example of a binding site detected by eFindSite.** The model of  
54 vitamin D receptor (VDR, purple ribbons) is superposed onto the crystal structure of  
55 homologous VDR from human (gold ribbons) complexed with a synthetic analog of vitamin D  
56 (gold and red sticks). C $\alpha$  atoms of binding residues predicted in the VDR model by eFindSite are  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4 shown as small spheres. Large spheres connected by a dashed black line are placed at the  
5 location of the predicted pocket center (purple) and the geometric center of vitamin D analog  
6 (gold).  
7  
8  
9

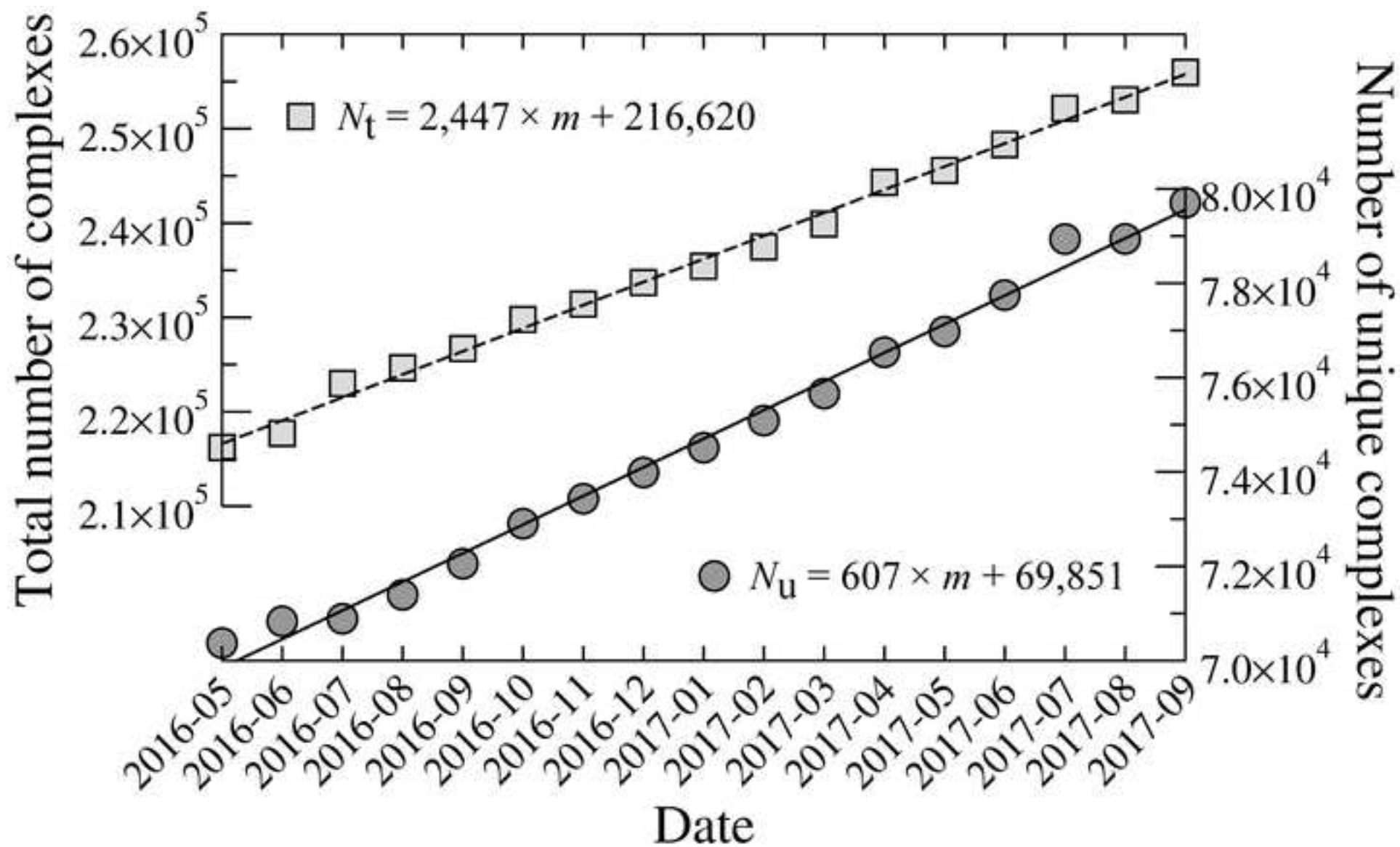
10  
11  
12 **Figure 8. Violin and box plots for the docking accuracy.** The accuracy is assessed for remote  
13 homology complex models in the validation set. The ligand RMSD is calculated over non-  
14 hydrogen atoms according to the chemical alignment reported by kcombu. The performance of  
15 similarity-based docking employed to construct eModel-BDB is compared to that of AutoDock  
16 Vina and rDock. Horizontal yellow lines represent median values.  
17  
18  
19  
20  
21  
22

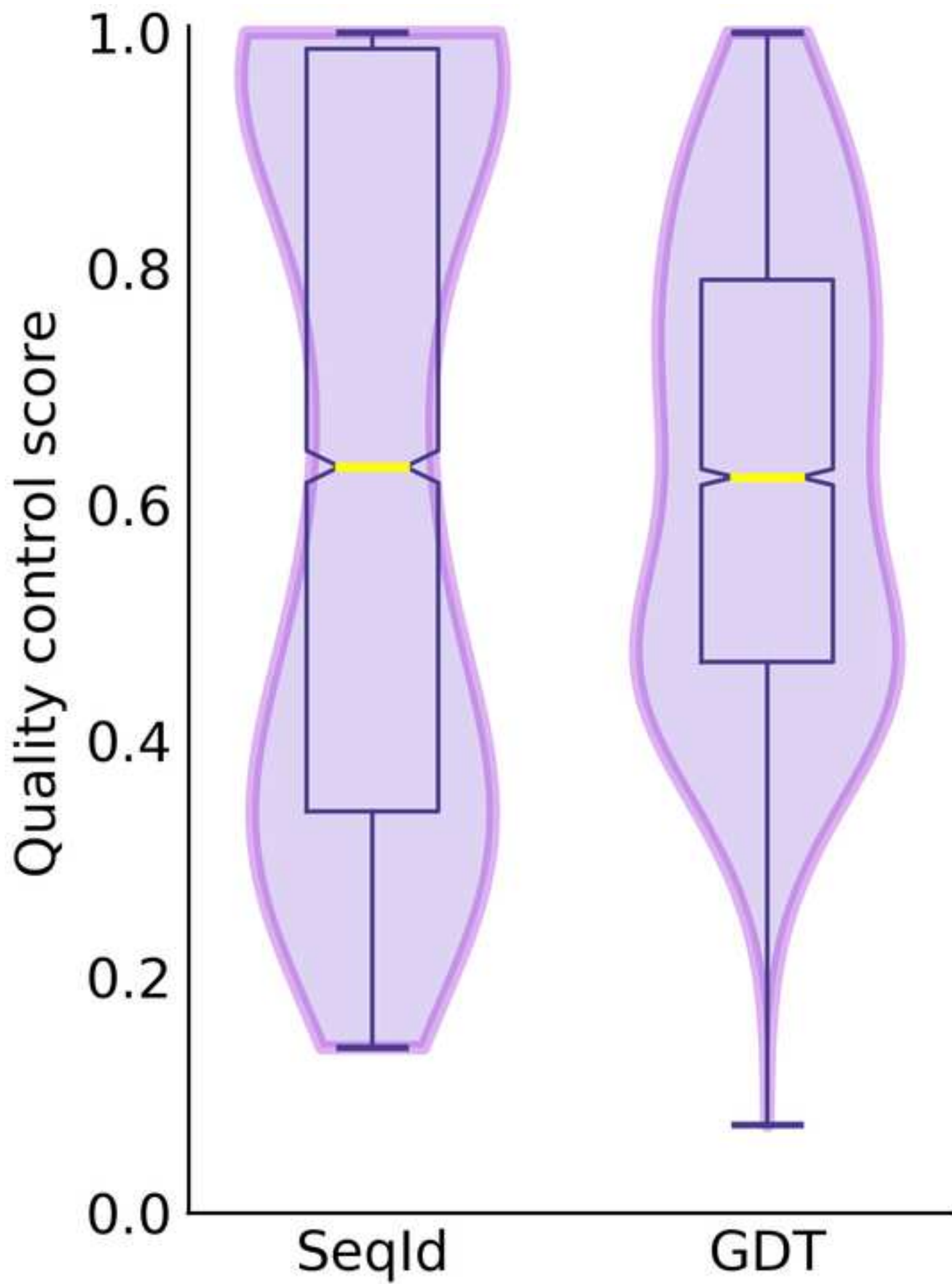
23  
24 **Figure 9. Representative example of a complex structure constructed by similarity-based**  
25 **docking.** (A) Chemical alignment between steroidal BDBM50061174 (left) and non-steroidal (R)-  
26 orteronel (right) reported by kcombu. 17 equivalent atom pairs constituting the maximum  
27 common substructure are numbered and outlined in purple in BDBM50061174 and in gold in  
28 (R)-orteronel. (B) The model of cytochrome P450 17A1 (CYP17A1, purple ribbons) is superposed  
29 onto the crystal structure of CYP17A1 from human (gold ribbons) complexed with (R)-orteronel  
30 (gold sticks).  $\alpha$  atoms of binding residues identified in the CYP17A1 model by eFindSite are  
31 shown as purple spheres, whereas the target compound, BDBM50061174, docked into the  
32 predicted pocket is represented by purple sticks. Non-carbon atoms in BDBM50061174 and (R)-  
33 orteronel are colored by atom type (O – red, N – blue).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

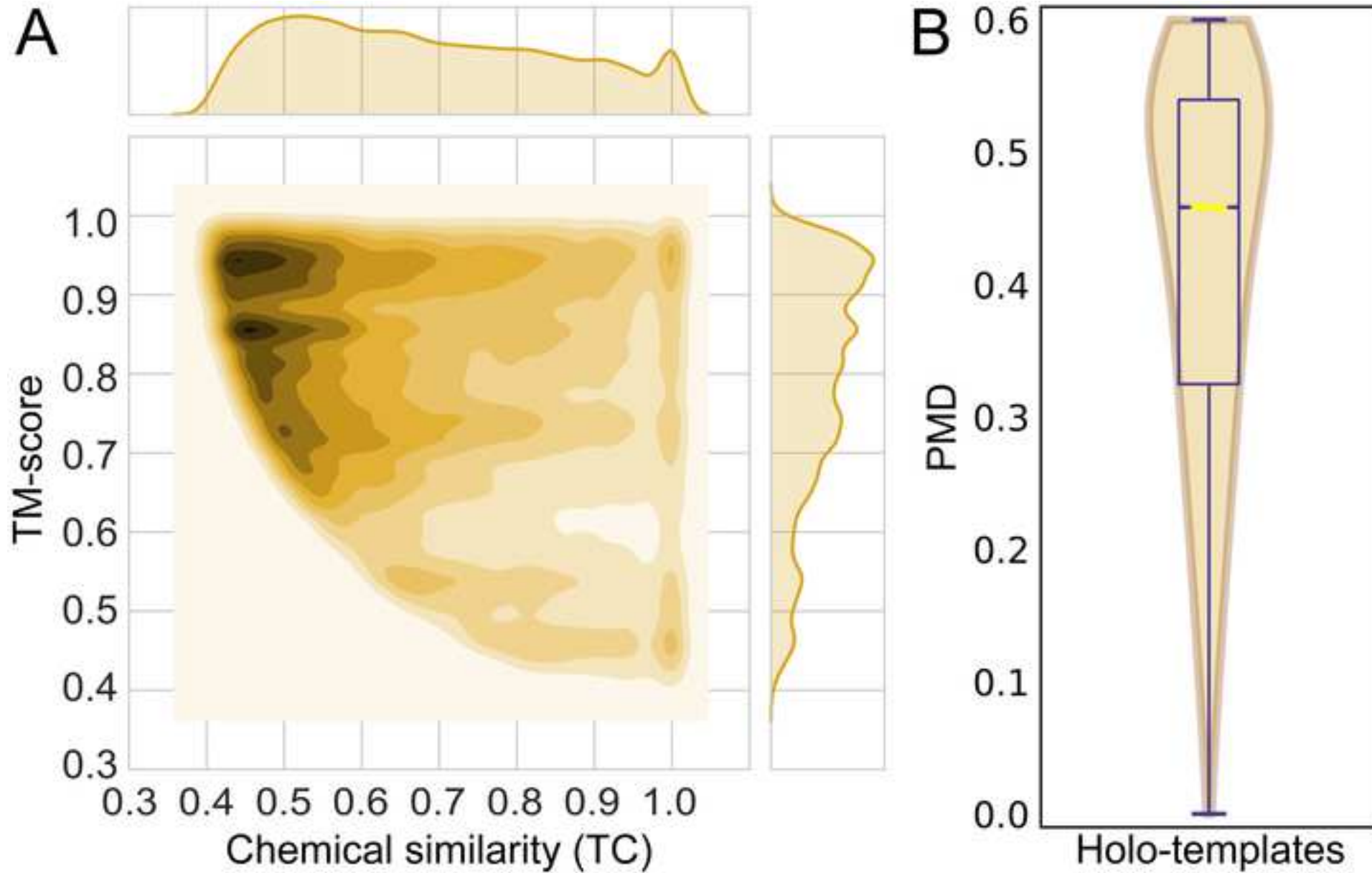
45 **Figure 10. Procedure to obtain eModel-BDB complexes via the BindingDB website.** Target  
46 complex can be identified based on either the protein (red arrows and boxes) or the ligand of  
47 interest (blue arrows and boxes). Common actions that a user needs to perform are colored in  
48 green. (A) Specific ligands and proteins can directly be searched for on the BindingDB website.  
49 (B, C) Alternatively, target proteins can be found with the blast search. (D) A complex of interest  
50 can then be selected in order to (E) generate and download a SDfile. (F) The BindingDB reactant  
51 set ID stored inside the SDfile is used to (G) view a web page containing detailed information  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

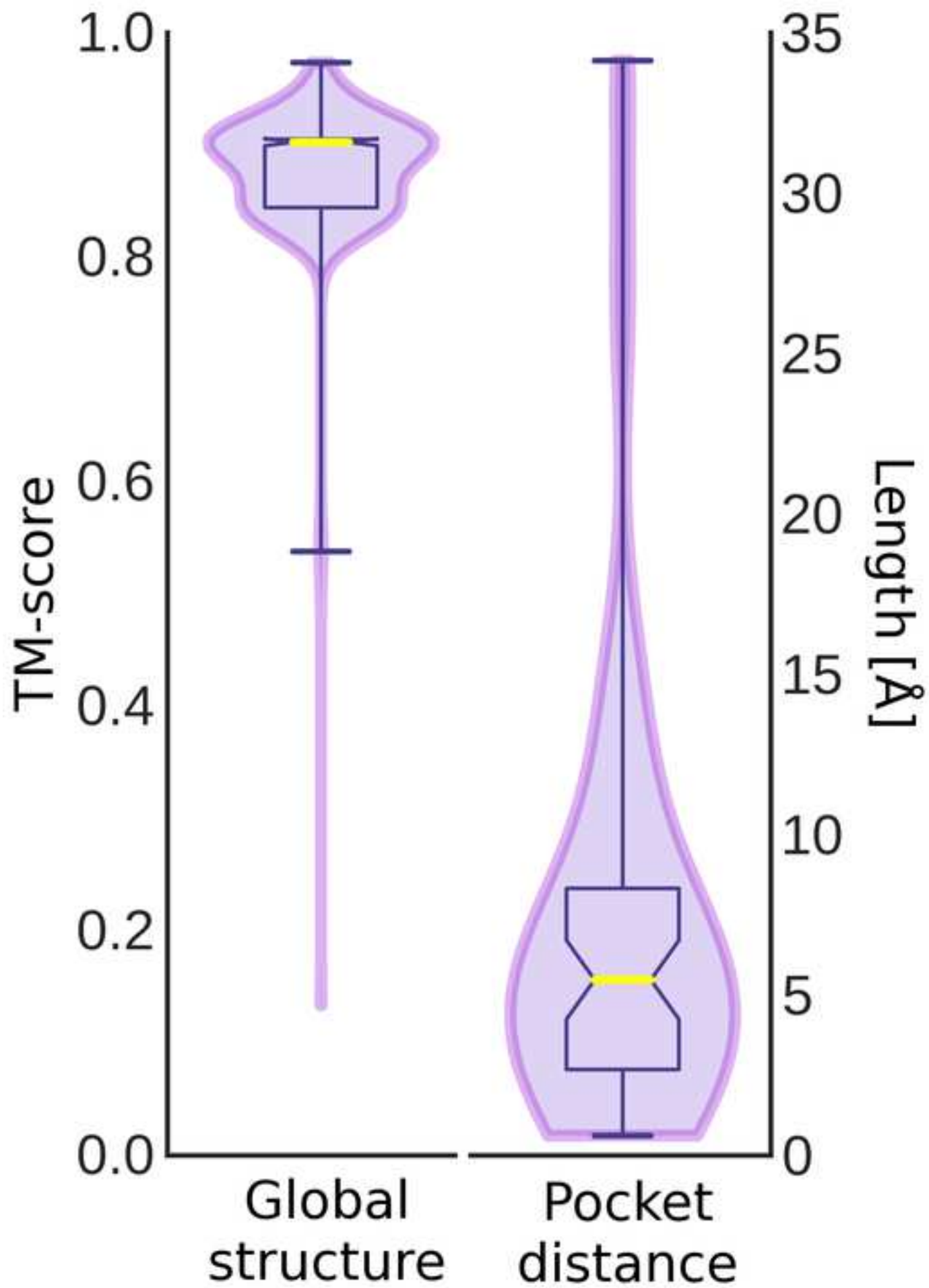
about the target complex as well as (H) access the corresponding eModel-BDB structure model  
named according to the BindingDB reactant set ID.

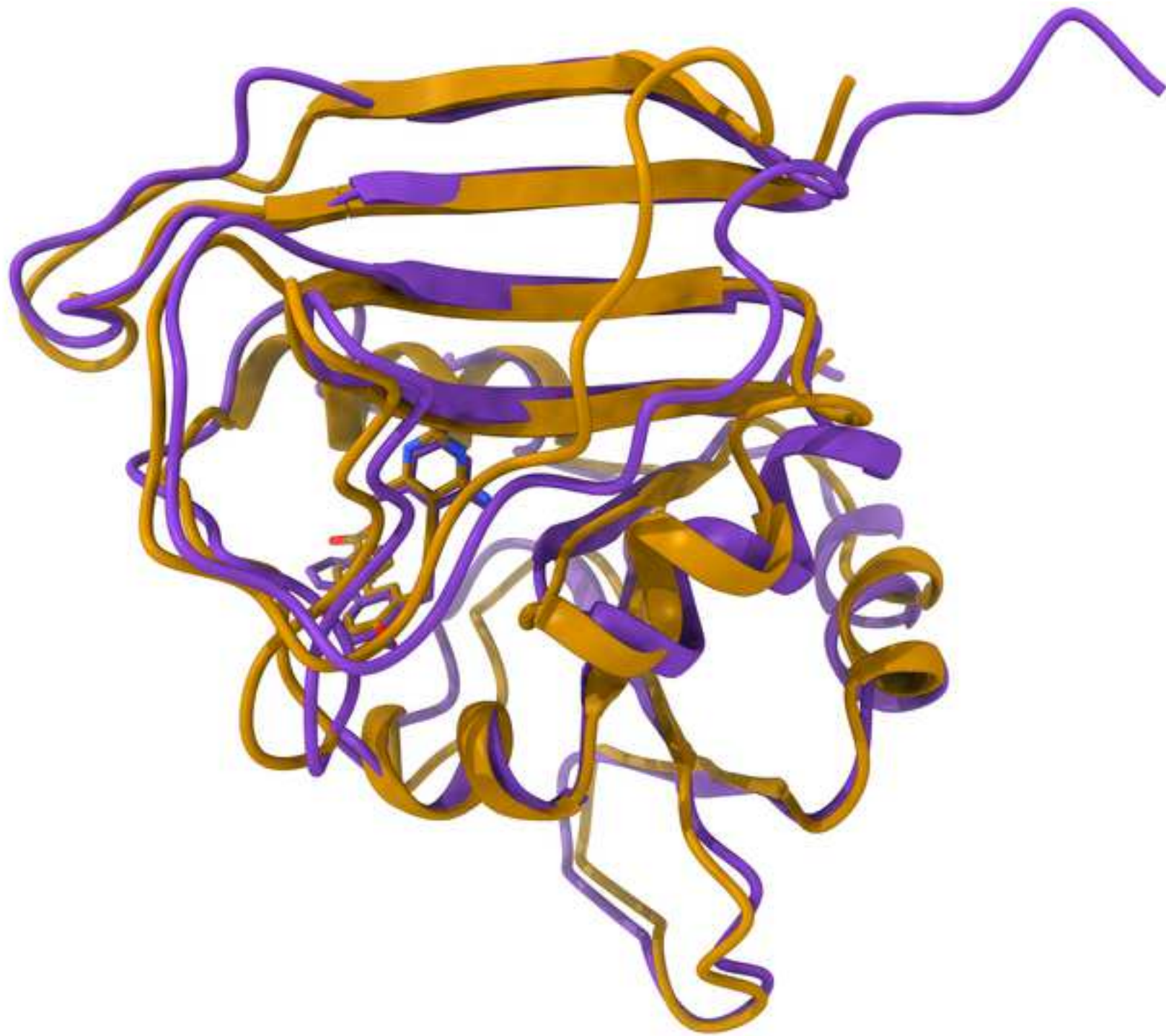


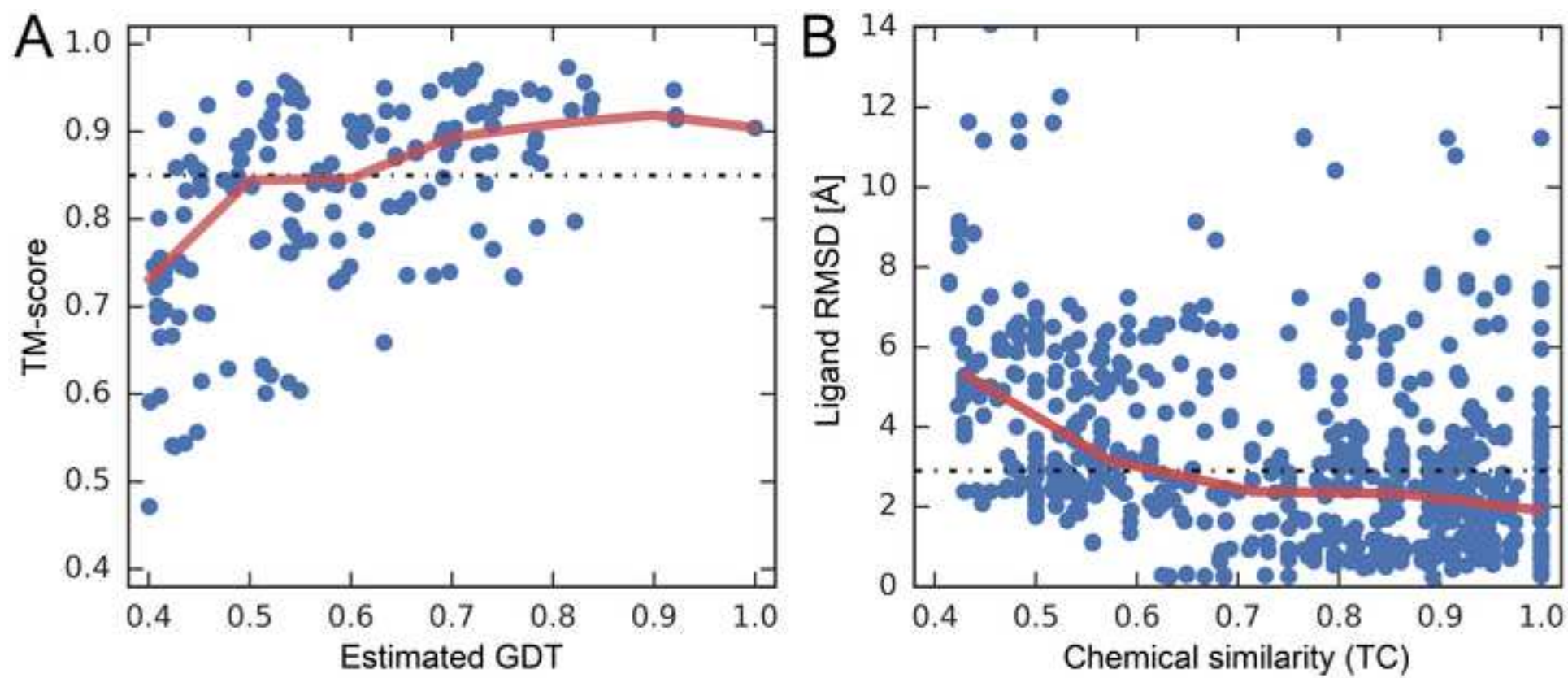


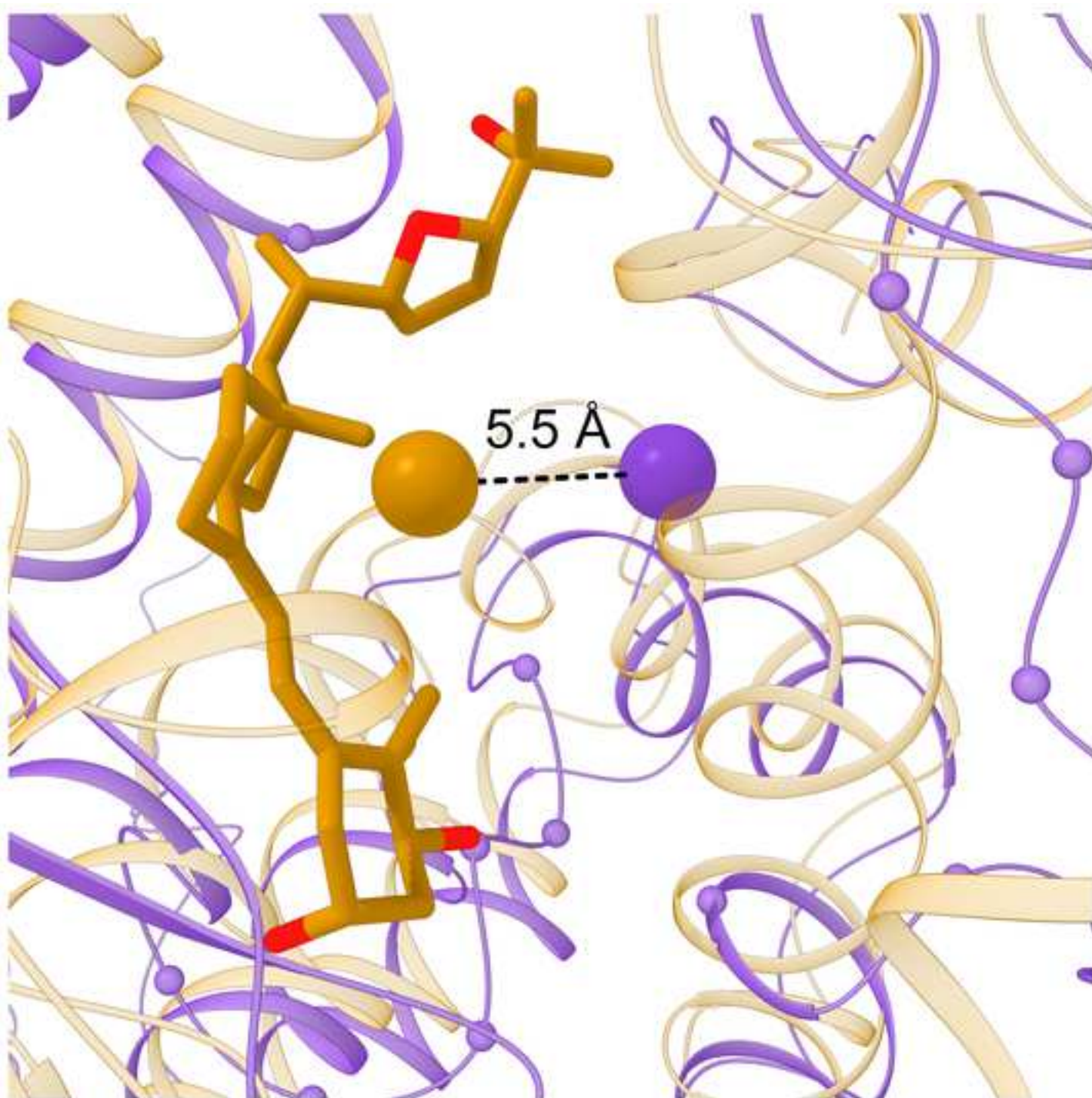


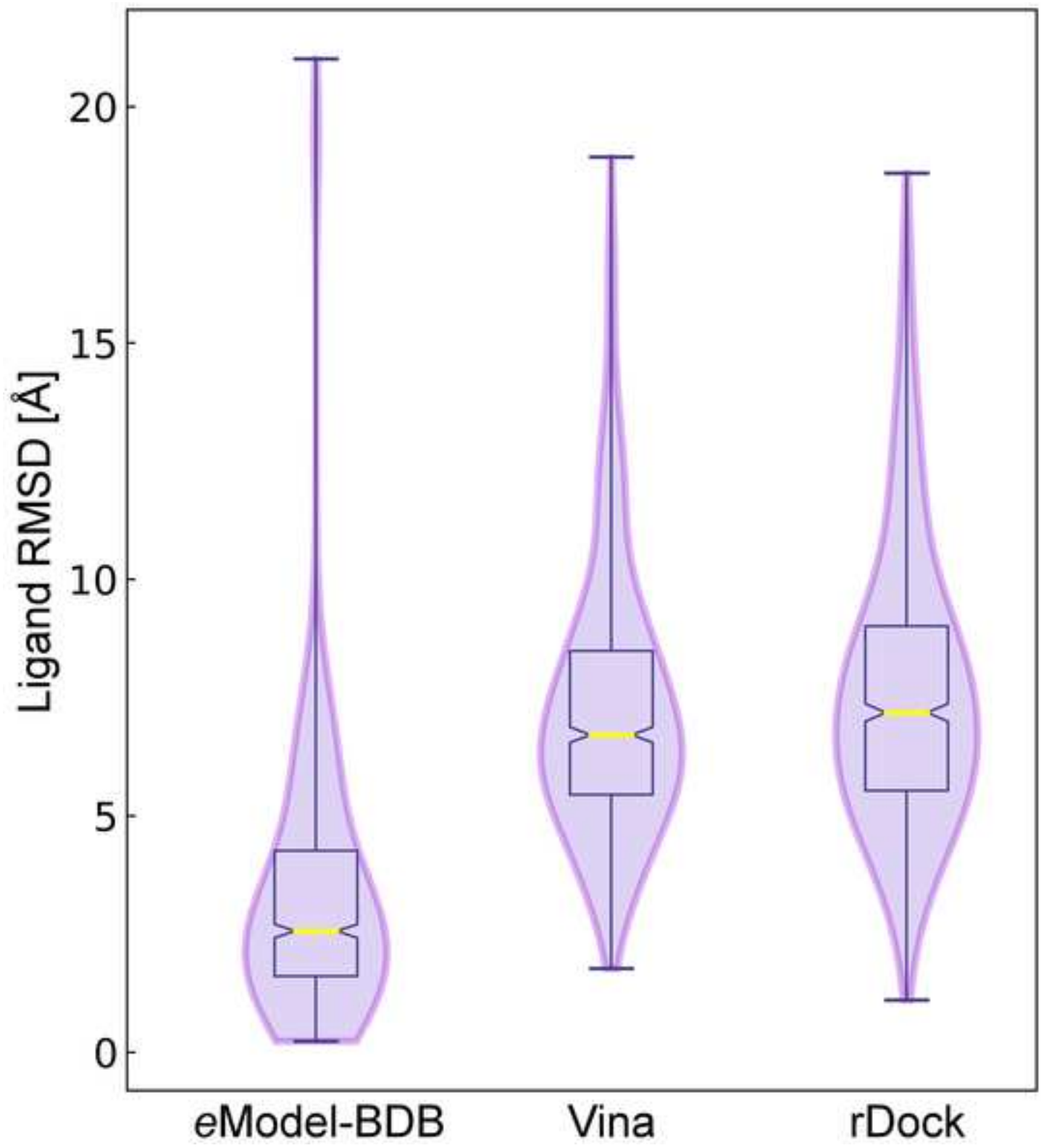


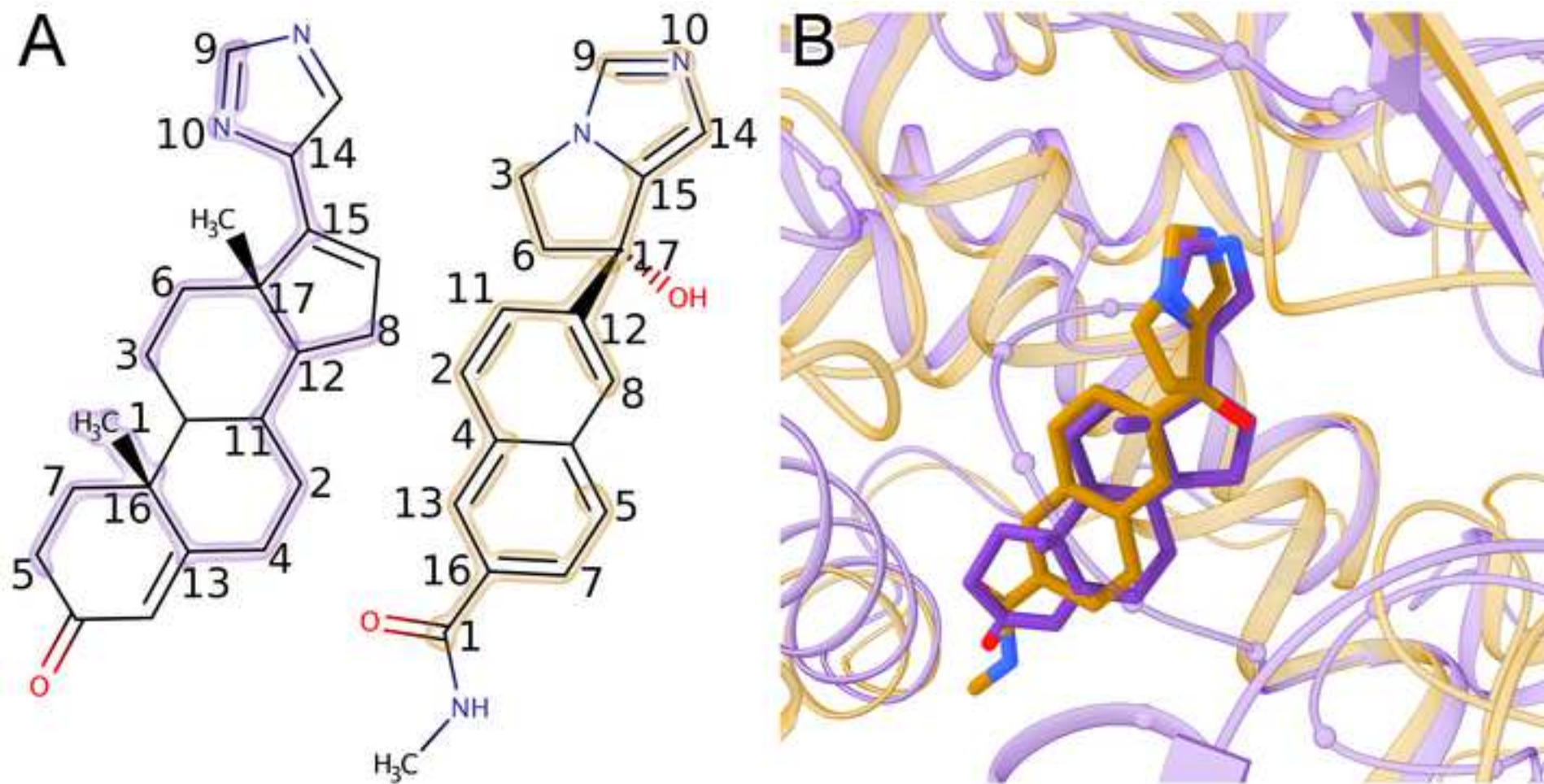












**A** The Binding Database

BindingDB is a public, web accessible database of measured binding affinities, between molecules and sequences of DNA, for small molecules.

Search and Browse

By target protein sequence

By ligand or target protein name

Simple Search

CHEMBL1270633

Advanced Search

**B** BindingDB Data by Sequence

Search BindingDB by sequence

**C** BLAST Output

Chromoblasta salicaria (2478)

**D** The Binding Database

Target: Chromoblasta salicaria (2478)

Ligand: CHEMBL1270633

**E** The Binding Database

Work with Selected Data

Download SDF file

```
><From>
www.bindingDB.org
><BindingDB Reactant_set_id>
267770
><ligand InChI>
InChI=1S/C22H22N4O/c1-3-19-18(21)
><ligand InChI Key>
LBQMRSNKEUNXMT-LBIFFFADYSA-N
```

**F**

**G** The Binding Database

Reaction Details

Target: Chromoblasta salicaria (2478)

Ligand: CHEMBL1270633

**H** eModelBDB

00267770.pdb



Click here to access/download  
**Supplementary Material**  
SupplementaryFileS1.csv







Click here to access/download  
**Supplementary Material**  
SupplementaryFileS2.csv



Click here to access/download  
**Supplementary Material**  
ResponseLetter.pdf

