bioRxiv preprint doi: https://doi.org/10.1101/590877; this version posted March 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

PrankWeb: web server for ligand binding-site prediction and visualization

Lukas Jendele [1,4,†], Radoslav Krivak [1], Petr Skoda [1], Marian Novotny [2] and David Hoksza [1,3*]

[1]Department of Software Engineering, Faculty of Mathematics and Physics, Charles University, Czech Republic
[2]Department of Cell Biology, Faculty of Science, Charles University, Czech Republic
[3]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg
[4]Department of Computer Science, ETH Zurich, Switzerland

Received March 18, 2019;

ABSTRACT

PrankWeb is an online resource providing an interface to P2Rank, a state-of-the-art ligand binding site prediction method. P2Rank is a template-free machine learning method which is based on the prediction of ligandability of local chemical neighborhoods centered on points placed on a solvent accessible surface of a protein. Points with high ligandability score are then clustered to form the resulting ligand binding sites. On top of that, PrankWeb then provides a web interface enabling users to easily carry out the prediction and visually inspect the predicted binding sites via an integrated sequence-structure view. Moreover, PrankWeb can determine sequence conservation for the input molecule and use it in both the prediction and results visualization steps. Alongside its online visualization options, PrankWeb also offers the possibility to export the results as a PyMOL script for offline visualization. The web frontend communicates with the serer side via a REST API. Therefore, in high-throughput scenarios users can utilize the server API directly, bypassing the need for a web-based front end or installation of the P2Rank application. PrankWeb is available at http://prankweb.cz/. The source code of the web application and the P2Rank method can be accessed at https://github.com/jendelel/PrankWebApp and https://github.com/rdk/p2rank, respectively.

INTRODUCTION

The field of structural biology has recently experienced an enormous progress in all aspects of structural determination process and as a result 3D structures of proteins are becoming increasingly available. Structural genomics consortia are even solving structures of proteins without known functions (1). The information acquired from 3D coordinates of proteins with unknown function is used to annotate these proteins. A very important clue towards function prediction of a protein can be an identification of ligands or small molecules that can bind to the protein. Ligands and other small molecules can be either directly determined within the protein's 3D structure or a 3D structure of the protein can be used to predict ligand binding sites and thus help to annotate the protein.

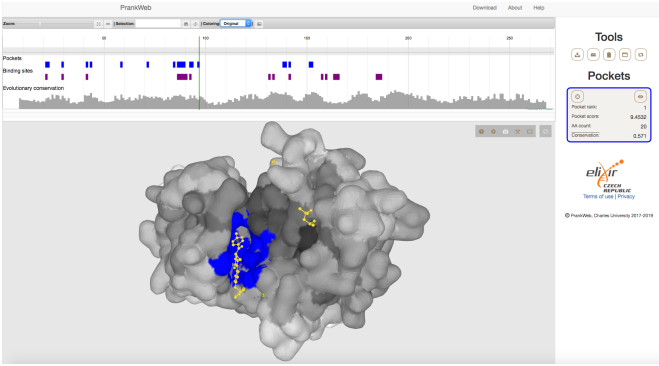

Figure 1. An example of PrankWeb output. The figure shows a prediction of ligand binding sites (blue colour) on a surface of human Lck kinase (3AD5). The actual ligand binding pose of dasatinib is shown in yellow. The second small molecule in the figure is dimethyl sulfoxide. The figure also shows a sequence view on the protein with indicated binding sites and conservation scores (top panel). The right panel shows a summary of the binding sites and provides tools to modify the view or to download the results.

A wide range of approaches for protein ligand-binding site prediction have been developed. Table 1 lists those prediction tools that are provided as a web service. Fpocket (2), SiteHound (3), ConCavity (4), POCASA (5), MetaPocket 2.0 (6), FTSite (7) and bSiteFinder (8) support online visualization using Jmol (9), a Java-based molecular structure viewer. However, due to security risks, Java applets are no longer supported in modern web browsers and thus these websites can be considered outdated. A simple solution to the Jmol issue is to use JSmol (10), a JavaScript replacement for Jmol. This is the avenue taken by 3DLigandSite (11), COFACTOR (12, 13), COACH (14) ISMBLAB-LIG (15) and LIBRA (16). Although JSmol supports complex visualization options, it suffers from performance issues due to the inefficiencies introduced when migrating Jmol code from Java to JavaScript. OpenAstex (17), another Java based visualizer, is utilized by Fpocket. However, this project suffers from the

*Correspondence should be addressed to D. Hoksza. Tel: +420 951 554 406; Email: hoksza@ksi.mff.cuni.cz
† The work has been done while studying at Charles University.

bioRxiv preprint doi: https://doi.org/10.1101/590877; this version posted March 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1. Availability of existing web-based tools for structure-based ligand binding site prediction introduced since 2009.

Name	Year	Type	Stand-alone	Online Visualization	Offline Visualization	Source Code
SiteHound (3)	2009	energetic	Yes	Jmol	PyMOL†, Chimera†	Yes
ConCavity (4)	2009	conservation	Yes	Jmol	PyMOL	Yes
Fpocket (2)	2010	geometric	Yes	Jmol, OpenAstex	PyMOL, VMD	Yes
3DLigandSite (11)	2010	template	—	JSmol	PyMOL	—
POCASA (5)	2010	geometric	—	Jmol	—	—
DoGSite (22)	2010	geometric	—	NGL	—	—
MetaPocket 2.0 (6)	2011	consensus	—	Jmol	PyMOL	—
FTSite (7)	2012	energetic	—	Jmol, static	PyMOL	—
COFACTOR (12, 13)	2012, 2017	template	Yes	JSmol	—	—
COACH (14)	2013	template	Yes	JSmol	—	—
eFindSite (27)*	2014	template	Yes	—	PyMOL, VMD, Chimera	Yes
GalaxySite (24)	2014	template/docking	—	PV, static	—	—
bSiteFinder (8)	2016	template	—	Jmol	—	—
ISMBLab-LIG (15)	2016	machine learning	—	JSmol&sequence	—	—
LIBRA-WA (16)	2017	template	Yes	JSmol	—	—
DeepSite (23)	2017	machine learning	—	NGL	—	—
PrankWeb (P2Rank)	this work	machine learning	Yes	LiteMol & Proteal	PyMOL	Yes

†only data files provided
*in the process of setting up a new interface

same problem as Jmol and seems to be discontinued as we did not manage to find any active resource. Only a few of the web servers support visualization via modern WebGL-based viewers such as LiteMol (18), NGL (19, 20) or PV (21). Specifically, NGL supports visualizations in DoGSite (22) and DeepSite (23). Despite the possibility to view the 3D structures in NGL, DeepSite and DoGSite websites lack the option to customize the protein, ligand and binding sites visualization. Similarly, GalaxySite (24) offers minimal 3D cartoon visualization of the protein and its ligands via the PV viewer. Finally, we have recently developed P2Rank (25), a state-of-the-art method for protein ligand binding site prediction. This work describes PrankWeb, an online web server providing interactive interface for the P2Rank method. The purpose of PrankWeb is to serve as an intuitive tool for ligand binding site prediction and its immediate visual analysis. PrankWeb displays the predictions as a combination of the protein's 3D structure, sequence and list of binding pockets. It allows users to display protein ligand binding sites and conservation both in the structural and sequence view and customize the style of visualization. PrankWeb's visualization is based on LiteMol and Protael (26), and thus runs in all modern browsers without additional plugins.

MATERIALS AND METHODS

P2Rank

P2Rank (25), the backend of PrankWeb, is a template-free, machine learning-based, ligand binding site prediction method. It employs random forests (28) to predict ligandability of points on the solvent accessible surface of a protein. These points represent potential locations of contact atoms of binding ligands and are described by a feature vector calculated form local geometric neighbourhood. The feature vector consists of physico-chemical and geometrical properties calculated from the surrounding atoms and residues (e.g. hydrophobicity, aromaticity or surface protrusion). PrankWeb also introduces a new model that includes the information derived from sequence evolutionary conservation scores of residues (see Supplementary Information on how the conservation scores are computed). Points with high predicted ligandability are clustered and ranked according to a ranking function based on cumulative score of the cluster.

P2Rank can use different pretrained models with varying feature vectors. PrankWeb exposes two such models: the default P2Rank model (without conservation) and a new model that uses conservation information (P2Rank+Conservation). Both models were trained on relatively small but diverse dataset of protein ligand complexes (25, 29).

As a template-free method, P2Rank does not share the limitation of template based methods which are unable

Table 2. Benchmark on COACH420 and HOLO4K datasets.

	COACH420		HOLO4K	
	Top-n	Top-(n+2)	Top-n	Top-(n+2)
Fpocket 1.0	56.4	68.9	52.4	63.1
Fpocket 3.1	42.9	56.9	54.9	64.3
SiteHound*	53.0	69.3	50.1	62.1
MetaPocket 2.0*	63.4	74.6	57.9	68.6
DeepSite*	56.4	63.4	45.6	48.2
P2Rank	72.0	**78.3**	68.6	74.0
P2Rank+Cons.†	**73.2**	77.9	**72.1**	**76.7**

Comparing identification success rate [%] measured by the DCA criterion (distance from pocket center to closest ligand atom) with 4 Å threshold considering only pockets ranked at the top of the list (n is the number of ligands in the considered structure).
*Failed to produce predictions for some of the input proteins. Here we display success rates calculated only based on subsets of proteins, on which corresponding methods finished successfully.
† P2Rank with conservation (the default prediction model of PrankWeb)

bioRxiv preprint doi: https://doi.org/10.1101/590877; this version posted March 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

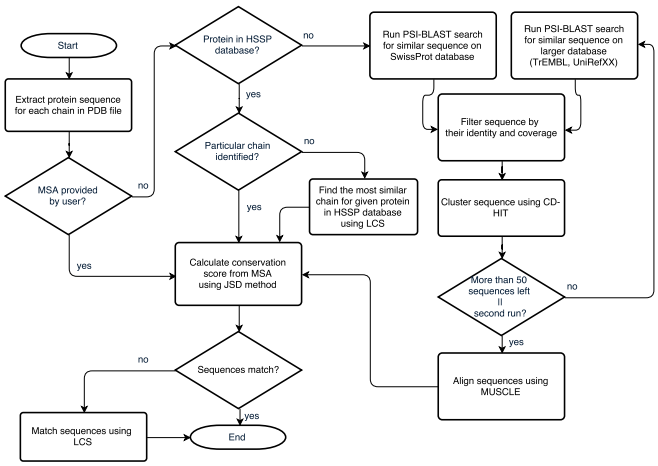

Figure 2. Diagram of conservation loading workflow and conservation pipeline.

to predict truly novel sites that have no analogues in their template libraries of known protein-ligand complexes. Therefore, P2Rank should be particularly beneficial for predicting novel allosteric sites, for which template based methods would generally be less effective (25). Another advantage of P2Rank is the fact that it is able to work directly with multi-chain structures and predict binding sites formed near the interfaces of chains.

We have compared the predictive performance of the presented tool with several competing algorithms on two datasets: COACH420 (14) containing 420 single-chain complexes and HOLO4K (25) containing 4009 multi-chain structures. Results are shown in Table 2. The default model used by PrankWeb (P2Rank+Conservation) clearly outperforms the other compared methods, and in most cases also the original P2Rank model that does not use conservation. Many of the methods listed in Table 1 are hard to compare on larger datasets as they do not, unlike PrankWeb, expose REST APIs. Thus, batch processing is further hindered by slow running times, delivering results only by email or captcha. Description of evaluation methodology and more detailed results are included in the Supplementary Material.

Prediction speeds vary heavily among the tools ranging from under 1 second (Fpocket, P2Rank) to more than 10 hours (COACH) for prediction on one protein of average size (2500 atoms). We have previously shown that P2Rank (without conservation) is the second fastest of the available tools (25). PrankWeb does not bring much overhead to prediction speed, however using the model with conservation may take a few minutes if conservation scores need to be calculated from scratch (see Conservation pipeline section in the Supplementary Material).

Web server

PrankWeb allows users to predict and visualize the predicted protein-ligand binding sites and contrast these with highly conserved areas as well as with actual ligand binding sites.

To carry out the prediction, users can either upload a PDB file or provide a PDB ID in which case PrankWeb will download and store the corresponding PDB file from the PDB database (30). Besides selecting what protein to analyze, users can also specify whether evolutionary conservation should be included in the prediction process, which in turn determines which of the two pretrained models will be used.

Conservation scores are calculated using Jensen-Divergence method (31) from a multiple sequence alignment (MSA) file, which can come from three sources: 1) users can specify their own alignment file, 2) if a protein's PDB code is provided, PrankWeb uses MSA from the HSSP (32) database, 3) in case no MSA is provided and no MSA is found in HSSP, the MSA is computed using PrankWeb's own conservation pipeline which utilizes UniProt (33), PSI-Blast (34), MUSCLE (35) and CD-HIT (36). The process is depicted in Figure 2 and described in detail in the Supplementary Material.

After specification of the input, the submitted data is sent via a REST API to the server which starts the prediction pipeline. The user is provided with a URL address where the progress of the prediction process can be tracked and where the results can be inspected once the process finishes.

On the results page, PrankWeb utilizes LiteMol for visualization of the 3D structure information and Protael for sequence visualization. Figure 1 displays predicted binding sites of dasatinib (a drug used for a chromic myelogenous leukemia treatment) bound to the kinase domain of human LCK (PDB ID 3AD5). The sequence and structure plugins are synchronized so that the user can easily locate a sequence position in the structure and vice versa. The sequence view comprises predicted pockets, computed conservation and binding sites (if present in the PDB file). The side panel displays information about identified pockets and a toolbar with the possibility to i) to download all inputs and calculated results, ii) share the results page link or iii) switch between visualization modes. PrankWeb comes with three predefined 3D model renderings: protein surface, cartoon and atoms. Predicted binding sites as well as conservation scores are color coded. If conservation analysis was chosen, the user can contrast the positions of putative active sites with conservation scores of the respective positions. In case the preset modes do not suffice, one can completely customize the 3D visualization using LiteMol's advanced user interface or using the PyMOL visualization script for offline inspection.

PrankWeb consists of a Java backend, REST API and a Typescript frontend. The backend is based on WildFly (37) web server and P2Rank application. The frontend utilizes Protael, LiteMol and Bootstrap.js libraries to provide interactive user interface on top of the REST API. All source code is available under the Apache License 2.0 at GitHub (https://github.com/jendelel/PrankWebApp). The GitHub website also includes documentation for other developers on how to use our REST API as well as how to deploy their own instance of the server.

DISCUSSION

PrankWeb has proven to provide correct predictions even in cases where other methods have failed. Nghan et al. (7) mentioned three cases where their method FTSite was unable to identify the ligand binding site with their best ranked prediction. PrankWeb correctly identified the binding site as the best ranked both in apo and holo structure in

bioRxiv preprint doi: https://doi.org/10.1101/590877; this version posted March 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

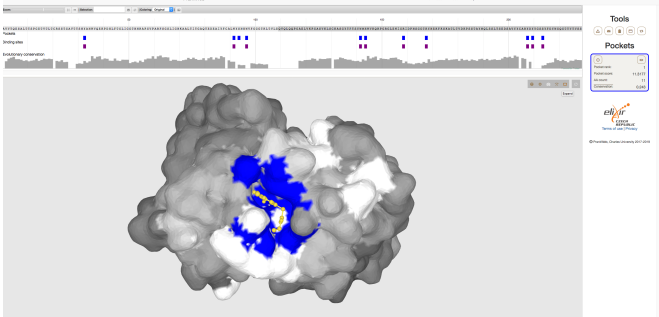

Figure 3. Figure 3: Prediction a difficult pocket. Authors of FTSite method described three structures, where their method failed. This figure shows a PrankWeb prediction for one of them - a structure of mouse immunoglobulin (1a6w). The prediction is indicated by a blue colour and the actual ligand is in yellow.

all three cases. Namely, in the case of glucose/galactose receptor (1GCG, 1GCA), purine nucleoside phosphorylase (1ULA,1ULB) and finally also in the case of mouse FV fragment of antibody (1A6U,1A6W). Figure 3 shows the predicted ligand binding site of holo structure (1GCA) on the interface of two imunoglobulin subunits together with experimentaly solved structure of 4-HYDROXY-5-IODO-3-NITROPHENYLACETYL-EPSILON-AMINOCAPROIC ACID ANION (NIP). The 3D structure of NIP appears in the PDB just once, which makes it difficult to train its binding.

We should highlight that none of the models employed by PrankWeb was trained on metallic ion ligands and would not in general be good at finding their binding sites. Such task would be better served by a model trained on a specialized dataset (which is possible to do with P2Rank).

CONCLUSION

We presented PrankWeb, a new web interface for the state-of-the-art ligand binding prediction method P2Rank. PrankWeb allows users to quickly carry out a prediction and visually inspect the results. PrankWeb also contains a pipeline for computation of conservation scores, which are included in the ligand binding site prediction as well as the structure-sequence visualization of the results. While PrankWeb provides a user-friendly interface, it also serves a REST API, enabling developers to use PrankWeb as a service. Both PrankWeb and P2Rank are open sourced on GitHub and freely available.

ACKNOWLEDGEMENTS

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme ”Projects of Large Research, Development, and Innovations Infrastructures” (CESNET LM2015042), is greatly appreciated.

FUNDING

This work was supported by ELIXIR CZ research infrastructure project (MEYS Grant No: LM2015047) and by the Grant Agency of Charles University (project Nr. 1556217).

Conflict of interest statement. None declared.

REFERENCES

1. Skolnick, J., Fetrow, J. S., and Kolinski, A. (2000) Structural genomics and its importance for gene function analysis. Nature Biotechnol, **18**(3), 283 [PubMed:10700142] [doi:doi:10.1038/73723].
2. Schmidtke, P., Le Guilloux, V., Maupetit, J., and Tufféry, P. (2010) Fpocket: online tools for protein ensemble pocket detection and tracking. Nucleic Acids Res., **38**(suppl_2), W582–W589 [PubMed:20478829] [PubMed Central:PMC2896101] [doi:10.1093/nar/gkq383].
3. Hernandez, M., Ghersi, D., and Sanchez, R. (2009) SITEHOUND-web: a server for ligand binding site identification in protein structures. Nucleic Acids Res., **37**(suppl_2), W413–W416 [PubMed:19398430] [PubMed Central:PMC2703923] [doi:10.1093/nar/gkp281].
4. Capra, J. A., Laskowski, R. A., Thornton, J. M., Singh, M., and Funkhouser, T. A. (2009) Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3D structure. PLoS Comput Biol, **5**(12), e1000585 [PubMed:19997483] [PubMed Central:PMC2777313] [doi:10.1371/journal.pcbi.1000585].
5. Yu, J., Zhou, Y., Tanaka, I., and Yao, M. (2010) Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. Bioinformatics, **26 1**, 46–52 [PubMed:19846440] [doi:10.1093/bioinformatics/btp599].
6. Zhang, Z., Li, Y., Lin, B., Schroeder, M., and Huang, B. (2011) Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics, **27**(15), 2083 [PubMed:21636590] [doi:10.1093/bioinformatics/btr331].
7. Ngan, C.-H., Hall, D. R., Zerbe, B. S., Grove, L. E., Kozakov, D., and Vajda, S. (2012) FTSite: high accuracy detection of ligand binding sites on unbound protein structures. Bioinformatics, **28 2**, 286–7 [PubMed:22113084] [PubMed Central:PMC3259439] [doi:10.1093/bioinformatics/btr651].
8. Gao, J., Zhang, Q., Liu, M., Zhu, L., Wu, D., Cao, Z., and Zhu, R. (2016) bSiteFinder, an improved protein-binding sites prediction server based on structural alignment: more accurate and less time-consuming. J. Cheminf., **8**(1), 38 [PubMed:27403208] [PubMed Central:PMC4939519] [doi:10.1186/s13321-016-0149-z].
9. Tully, S. P., Stitt, T. M., Caldwell, R. D., Hardock, B. J., Hanson, R. M., and Maslak, P. (2013) Interactive Web-Based Pointillist Visualization of Hydrogenic Orbitals Using Jmol. J. Chem. Educ., **90**(1), 129–131 [doi:10.1021/ed300393s].
10. Hanson, R. M., Prilusky, J., Renjian, Z., Nakane, T., and Sussman, J. L. (2013) JSmol and the Next-Generation Web-Based Representation of 3 D Molecular Structure as Applied to Proteopedia. Isr. J. Chem., **53**(3-4), 207–216 [doi:10.1002/ijch.201300024].
11. Wass, M. N., Kelley, L. A., and Sternberg, M. J. (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. Nucleic Acids Res., **38**(suppl_2), W469–W473 [PubMed:20513649] [PubMed Central:PMC2896164] [doi:10.1093/nar/gkq406].
12. Roy, A., Yang, J., and Zhang, Y. (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. Nucleic Acids Res., **40**(W1), W471–W477 [PubMed:22570420] [PubMed Central:PMC3394312] [doi:10.1093/nar/gks372].
13. Zhang, C., Freddolino, P. L., and Zhang, Y. (2017) COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. Nucleic Acids Res., **45**(W1), W291–W299 [PubMed:28472402] [PubMed Central:PMC5793808] [doi:10.1093/nar/gkx366].
14. Yang, J., Roy, A., and Zhang, Y. (2013) Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics, **29 20**, 2588–95 [PubMed:23975762] [PubMed Central:PMC3789548] [doi:10.1093/bioinformatics/btt447].
15. Jian, J.-W., Elumalai, P., Pitti, T., Wu, C. Y., Tsai, K.-C., Chang, J.-Y., Peng, H.-P., and Yang, A.-S. (2016) Predicting ligand binding sites on protein surfaces by 3-Dimensional probability density distributions of interacting atoms. PLoS One, **11**(8), e0160315 [PubMed:27513851] [PubMed Central:PMC4981321] [doi:10.1371/journal.pone.0160315].

bioRxiv preprint doi: https://doi.org/10.1101/590877; this version posted March 28, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

16. Toti, D., Viet Hung, L., Tortosa, V., Brandi, V., and Polticelli, F. (2017) LIBRA-WA: a web application for ligand binding site detection and protein function recognition. Bioinformatics, **34**(5), 878–880 [PubMed:29126218] [PubMed Central:PMC6192203] [doi:10.1093/bioinformatics/btx715].
17. Hartshorn, M. J. (2002) AstexViewer TM: a visualisation aid for structure-based drug design. J. Comput. Aid. Mol. Des., **16**(12), 871–881 [PubMed:12825620] [doi:10.1023/A:1023813504011].
18. Sehnal, D., Deshpande, M., Vareková, R. S., Mir, S., Berka, K., Midlik, A., Pravda, L., Velankar, S., and Koča, J. (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. Nat. Methods, **14**, 1121–1122 [PubMed:29190272] [doi:10.1038/nmeth.4499].
19. Rose, A. S. and Hildebrand, P. W. (2015) NGL Viewer: a web application for molecular visualization. Nucleic Acids Res., **43**(W1), W576 [PubMed:25925569] [PubMed Central:PMC4489237] [doi:10.1093/nar/gkv402].
20. Rose, A. S., Bradley, A. R., Valasatava, Y., Duarte, J. M., Prlić, A., and Rose, P. W. (2016) Web-based Molecular Graphics for Large Complexes. In Proc. 21st Int. Conf. Web3D Technology New York, NY, USA: ACM [PubMed:29850778] [PubMed Central:PMC6198858] [doi:10.1093/bioinformatics/bty419] Web3D '16 pp. 185–186.
21. Biasini, M. pv: v1.8.1. (2015).
22. Volkamer, A., Griewel, A., Grombacher, T., and Rarey, M. (2010) Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets. J. Chem. Inf. Model., **50 11**, 2041–52 [PubMed:20945875] [doi:10.1021/ci100241y].
23. Jiménez, J., Doerr, S., Martínez-Rosell, G., Rose, A. S., and Fabritiis, G. D. (2017) DeepSite: protein-binding site predictor using 3D-convolutional neural networks. Bioinformatics, **33 19**, 3036–3042 [PubMed:28575181] [doi:10.1093/bioinformatics/btx350].
24. Heo, L., Shin, W.-H., Lee, M. S., and Seok, C. (2014) GalaxySite: ligand-binding-site prediction by using molecular docking. Nucleic Acids Res., **42**(W1), W210–W214 [PubMed:24753427] [PubMed Central:PMC4086128] [doi:10.1093/nar/gku321].
25. Krivák, R. and Hoksza, D. (2018) P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. J. Cheminf., **10**(1), 39 [PubMed:30109435] [PubMed Central:PMC6091426] [doi:10.1186/s13321-018-0285-8].
26. Sedova, M., Jaroszewski, L., and Godzik, A. (2016) Protael: protein data visualization library for the web. Bioinformatics, **32**(4), 602–604 [PubMed:26515826] [PubMed Central:PMC5963358] [doi:10.1093/bioinformatics/btv605].
27. Feinstein, W. P. and Brylinski, M. (2014) eFindSite: Enhanced Fingerprint-Based Virtual Screening Against Predicted Ligand Binding Sites in Protein Models. Mol. Inf., **33**(2), 135–150 [PubMed:27485570] [doi:10.1002/minf.201300143].
28. Ho, T. K. (1995) Random decision forests. In Proc. 3rd Int. Conf. Document Analysis and Recognition IEEE [doi:10.1109/ICDAR.1995.598994] Vol. 1, pp. 278–282.
29. Chen, K., Mizianty, M., Gao, J., and Kurgan, L. (2011) A critical comparative assessment of predictions of protein-binding sites for biologically relevant organic compounds.. Structure, **19**(5), 613–621 [PubMed:21565696] [doi:10.1016/j.str.2011.02.015].
30. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The protein data bank. Nucleic Acids Res., **28**(1), 235–242 [PubMed:10592235] [PubMed Central:PMC102472] [doi:doi:10.1093/nar/28.1.235].
31. Capra, J. A. and Singh, M. (2007) Predicting functionally important residues from sequence conservation. Bioinformatics, **23**(15), 1875–1882 [PubMed:17519246] [doi:10.1093/bioinformatics/btm270].
32. Joosten, R. P., te Beek, T. A., Krieger, E., Hekkelman, M. L., Hooft, R. W., Schneider, R., Sander, C., and Vriend, G. (2011) A series of PDB related databases for everyday needs. Nucleic Acids Research, **39**(suppl_1), D411 [PubMed:21071423] [PubMed Central:PMC3013697] [doi:10.1093/nar/gkq1105].
33. The UniProt Consortium (2018) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res., **47**(D1), D506–D515 [PubMed:30395287] [PubMed Central:PMC6323992] [doi:10.1093/nar/gky1049].
34. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.. Nucleic Acids Res., **25**, 3389–3402 [PubMed:9254694] [PubMed Central:PMC146917] [doi:10.1093/nar/25.17.3389].
35. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res., **32**(5), 1792 [PubMed:15034147] [PubMed Central:PMC390337] [doi:10.1093/nar/gkh340].
36. Li, W. and Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, **22**(13), 1658 [PubMed:16731699] [doi:10.1093/bioinformatics/btl158].
37. WildFly Homepage WildFly. http://wildfly.org/.