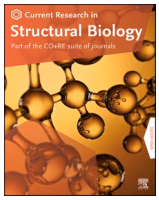
Current Research in Structural Biology 7 (2024) 100144

Contents lists available at ScienceDirect

Current Research in Structural Biology

journal homepage: www.journals.elsevier.com/current-research-in-structural-biology



Research Article

ANN based prediction of ligand binding sites outside deep cavities to facilitate drug designing

Kalpana Singh *, Yashpal Singh Malik

College of Animal Biotechnology, Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana-141004, India

ARTICLE INFO

Handling Editor: Dr A Wlodawer

Keywords:
Ligand binding site prediction
Surface roughness
Artificial neural network
Drug designing
Animal proteins
R-subspace

ABSTRACT

The ever-changing environmental conditions and pollution are the prime reasons for the onset of several emerging and re-merging diseases. This demands the faster designing of new drugs to curb the deadly diseases in less waiting time to cure the animals and humans. Drug molecules interact with only protein surface on specific locations termed as ligand binding sites (LBS). Therefore, the knowledge of LBS is required for rational drug designing. Existing geometrical LBS prediction methods rely on search of cavities based on the fact that 83% of the LBS found in deep cavities, however, these methods usually fail where LBS localize outside deep cavities. To overcome this challenge, the present work provides an artificial neural network (ANN) based method to predict LBS outside deep cavities in animal proteins including human to facilitate drug designing. In the present work a feed-forward backpropagation neural network was trained by utilizing 38 structural, atomic, physiochemical, and evolutionary discriminant features of LBS and non-LBS residues localized in the extracted roughest patch on protein surface. The performance of this ANN based prediction method was found 76% better for those proteins where cavity subspace (extracted by MetaPocket 2.0, a consensus method) failed to predict LBS due to their localization outside the deep cavities. The prediction of LBS outside deep cavities will facilitate in drug designing for the proteins where it is not possible due to lack of LBS information as the geometrical LBS prediction methods rely on extraction of deep cavities.

1. Introduction

The ever changing environment as well as pollution might be reasons for the onset of emerging diseases quite often, demands the faster designing of new drugs to curb the deadly diseases in animals and humans. Usually, drug molecules interact with proteins and thus, knowledge of protein structures is essential for rational drug designing or structure-based drug designing (Smith & Williams, 2002; Reynolds et al., 2010). However, there are less protein structures available in comparison to available protein sequences, hence, their functions remain undetermined. Interaction of drug molecules with proteins includes only protein surface (Pettit et al., 2007) on specific locations termed as ligand binding sites (LBS) or functional sites (Via et al., 2000). Therefore, their knowledge is essential for rational drug designing (Sotriffer and Klebe, 2002). Existing geometrical LBS prediction methods (Yu et al., 2010; Weisel et al., 2007; Brady and Stouten, 2000) rely on search of cavities based on the fact that 83% of the LBS found in deep cavities (Lewis, 1991; Laskowsk et al., 1996). There are various cavity based LBS prediction methods which are purely geometrical or geometrical with added physicochemical properties or purely energy based or evolutionary and threading based or consensus methods (which add results of other methods e.g. MetaPocket2.0 by Zhang et al. (2011)). Some of the efficient methods developed over the years are POCKET by Levitt & Banaszak (1992); SURFNET by Laskowski (1995); LIGSITE by Hendlich et al. (1997); PASS by Brady & Stouten (2000); Q-SiteFinder by Laurie & Jackson (2005); LIGSITEcs and LIGSITEcsc by Huang & Schroeder (2006); Pocketpicker by Weisel et al. (2007); FINDSITE by Brylinski & Skolnick (2008); Fpocket by Le Guilloux et al. (2009); MetaPocket by Huang (2009); MetaPocket 2.0 by Zhang et al. (2011), and Depth by Tan et al. (2013). For the final selection of the top cavities, suggested as potential LBSs, the existing methods utilize various combinations of atomic, residual, structural, and evolutionary features etc.

Due to dependency of the existing LBS prediction methods on cavity search, these methods fail where LBS localize outside the deep cavities i. e. in 17% proteins as reported by Laskowsk et al. (1996) and Nisius et al. (2012). In this regard, artificial neural network (ANN) based method is proposed for the prediction of LBS that localize outside deep cavities in animal proteins including human. In the present work, a feed-forward

* Corresponding author.
E-mail address: kalpana.iiita@gmail.com (K. Singh).

https://doi.org/10.1016/j.crstbi.2024.100144
Received 25 November 2023; Received in revised form 12 April 2024; Accepted 12 April 2024
Available online 15 April 2024

back propagation ANN was trained utilizing 38 discriminant features (such as structural, atomic, physiochemical, and evolutionary etc.) of the LBS and non-LBS residues found within the roughest patch on the protein surface extracted by R-subspace (Singh & Lahiri, 2017a). The performance of this ANN based prediction method was found 76% better for those animal proteins where cavity subspace (extracted by MetaPocket 2.0, a consensus method) failed to predict LBS due to their localization outside the deep cavities. The prediction of LBS outside deep cavities will facilitate in drug designing for the proteins where it is not possible due to lack of LBS information as the geometrical LBS prediction methods rely on extraction of deep cavities.

2. Material and methods

2.1. Data utilized

The data set I of 75 animal proteins including human (Supplementary Table 1) with LBS finding on shallow surface in place of deep cavities was downloaded from RCSB-PDB to train and test the ANN. Then, the trained network was tested for cross-validation on separate data set II with test set II of 25 animal proteins including human (Supplementary Table 1) extracted from 210 protein-ligand complexes were taken from the PLD database (Puvanendrampillai and Mitchell, 2003) along with dataset of 198 drug target protein data mentioned in the work of Zhang et al. (2011) with datasets of 48 bound (Holo) and unbound (Apo) structures mentioned in work of Huang and Schroeder (2006); for which MetaPocket 2.0 failed to predict LBS as localizing outside deep cavities (08 proteins for which MetaPocket2.0 failed to provide any result, 11 proteins for which none of the predicted cavities included LBS, and 06 proteins for which top three cavities failed to include LBS, mentioned in Supplementary Table 1). By the result of MetaPocket2.0, it was confirmed that in dataset II with test set II of 25 animal proteins including human LBSs were localized not in deep cavities. Lay-out of the work plan from data collection to network testing is provided in Fig. 1.

2.2. Extraction of roughest patch on protein surface

The roughest patch on protein surface was extracted using R-subspace in all the proteins utilized in the present study based on the algorithm mentioned by Singh and Lahiri (2017a). All the residues localized within the roughest patch were grouped into LBS and non-LBS residues based on the LBS residues listed in the PDB files of the proteins used in the present study. Then, various features (Table 1) were extracted for all the LBS and non-LBS residues to feed as input to the ANN to train and test the network.

2.3. Feature extraction for network training

There were 38 features (listed in Table 1) extracted for LBS and non-LBS residues and their atoms found within the localized roughest patch on the protein surface. First 8 features were physicochemical properties of 20 amino acids. Features 1 to 7 were taken from the work of Guo et al. (2008), which they utilized for protein-protein interaction study, however, in the present study, these features were found to be useful in LBS prediction also. Feature 8 was hydropathy index given by Kyte and Doolittle (1982). Features from 9 to 13 were structure based features. Feature 9 and 10 were calculated from the information given in PDB files, both these features are mentioned in the work of Sankararaman et al. (2010). Feature 11 is the new feature added in the present study, termed as Depth of residue and calculated as average distance of all atoms of each residue from the center of gravity (CG) of the protein. Features 12 to 14 were calculated by utilizing DSSP (Kabsch & Sander, 1983) program. Feature 15 was evolutionary feature extracted in terms of conservation score of individual amino acid residues and calculated by ConSurf (Ashkenazy et al., 2016) program. Features 16 to 33 were based on various amino acid residue properties (taken in binary) mentioned in the work of Sankararaman et al. (2010) and Nelson and Cox (2008). However, features 34 to 38 were the atomic properties of atoms of amino acid residues mentioned in the work of Krivák and Hoksza (2015).

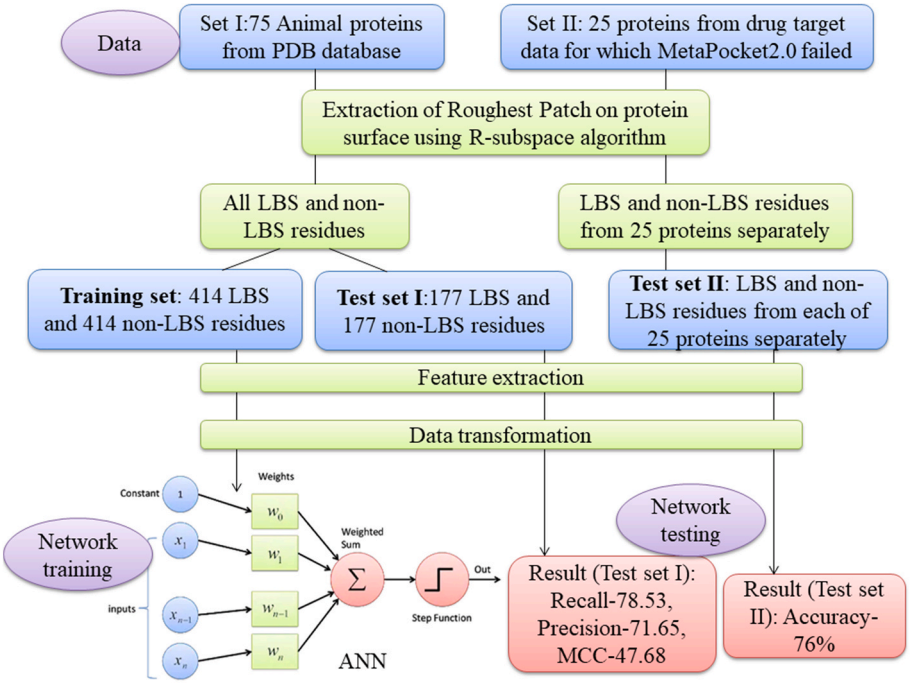


Fig. 1. Lay-out of the work plan from data collection to network testing (*LBS = ligand binding site).

Table 1
List of 38 features (based on 05 properties, denoted as A-E) extracted for LBS and non-LBS residues used to prepare input for ANN.

(A) **Physicochemical properties of amino acids**
1. Hydrophobicity - H1
2. Hydrophilicity -H2
3. Volume of side chains - V
4. Polarity – P1
5. Polarizability – P2
6. Solvent accessible surface area - SASA
7. Net charge index of side chains– NCI
8. Hydropathy index – HI

(B) **Structure based features**
9. B-factor – BF (taken from PDB file)
10. Centrality – CN (calculated as inverse average distance of each amino acid residue from all other residues of that protein)
11. Depth of residue – DR (calculated as distance of each residue from the center of gravity (CG) of the protein)
12. Secondary structure - SS (for individual residue calculated by DSSP)
13. Accessible surface – ACC (for each residue of the protein calculated by DSSP)
14. Kappa bend angle – KA (for each residue of the protein calculated by DSSP)

(C) **Evolutionary feature**
15. Conservation color – CC (conservation color scores (1–9) based on conservation scores calculated by ConSurf were taken and converted to binary by taking 9 as 1 and rest 1 to 8 as 0)

(D) **Residue properties**
16. Charged residues
17. Positive charged residues
18. Negative charged residues
19. Ionisable residues
20. Hydrophobic residues
21. Hydrophilic residues
22. Nucleophilic residues
23. Acidic residues
24. Basic residues
25. Amide residues
26. Aliphatic residues
27. Hydroxyl group containing residues
28. Cyclic side chain containing resides
29. Sulphur containing residues
30. Aromatic residues
31. Hydrogen acceptor residues
32. Hydrogen donor residues
33. Hydrogen donor acceptor residues

(E) **Atomic properties**
34. O atoms – (number of neighbouring oxygen atoms)
35. N atoms – (number of neighbouring nitrogen atoms)
36. C atoms – (number of neighbouring carbon atoms)
37. Hydrogen acceptor atoms –(number of neighbouring H-acceptor atoms)
38. Hydrogen donor atoms –(number of neighbouring H-donor atoms)

2.4. Classification of LBS and non-LBS residues

For the classification of amino acid residues into LBS (class 1: C1) and non-LBS residues (class 2: C2), following steps were followed:

2.4.1. Classifier utilized: ANN

In this study, ANN was utilized as classifier. Neural Network Tool of MATLAB was used with specifications such as: Feed-forward back propagation network (FFBPN) architecture, TRAINLM (Levenberg-Marquardt optimization) training function, LEARNGDM (Gradient descent with momentum weight and bias) adaptation learning function, MSE (mean of squared error) performance function, TANSIG transfer function, along with 01 hidden layers with 10 nodes, and 1000 epochs.

2.4.2. Training and test data sets

First, LBS and non-LBS residue data from the roughest patch extracted by R-subspace of 75 proteins were extracted to be taken as data set I including training set and test set (I). However, both the class data were unequal in size as the LBS residues (C1) are very less in number in comparison to non-LBS residues (C2), which could led to biased classification by network towards class of bigger data. Therefore, to nullified this class disparity, C2 data of equal size with C1 data was selected randomly from the whole C2 data using random sampling approach. Then, 70% of merged data (C1+C2) was taken as training data and rest of 30% data was taken as test set (I). A separate data set II including test set (II) was prepared from the 25 proteins (Table 2).

Table 2
Detail of data sets including animal proteins from various organisms used to train and test the network.

Protein data sets	Data sets	Number of inputs
I- 75 proteins	Training set	38 features of 414 LBS and 414 non-LBS residues
	Test set I	38 features of 177 LBS and 177 non-LBS residues)
II- 25 proteins	Test set II	38 features of LBS and non-LBS residues from individual protein at a time

2.4.3. Preparation of input and output data

Input data prepared for the ANN classifier was extracted as 38 features mentioned in Table 1 for all the residues localized within the roughest patch extracted by R-subspace. However, output data was prepared as (1, −1) for the LBS residues and (−1, 1) for the non-LBS residues.

2.4.4. Transformation of input data

In the present study, input data prepared by extracted 38 features was both in binary (e.g. residue properties) and decimals (e.g. physicochemical properties of amino acids). Therefore, the input data was needed to transform first to avoid the disparity in data ranges. The min-max normalization (Han et al., 2011) was utilized to transform input data in the range of 0–1 by using following formula:

$$X_i^{'} = \frac{X_i - min_F}{max_F - min_F}(max_{F'} - min_{F'}) + min_{F'}$$

Here, $X_i^{'}$ was transformed value of an unit X_i of feature F. min_F and max_F were minimum and maximum values of feature F, respectively, and $min_{F'}$ and $max_{F'}$ were new minimum and maximum values of feature F, respectively, which were taken as 0 and 1, respectively in the present study.

Then, the minimum and maximum values for the input data of training set were also utilized to normalize input data of both test sets I and II.

2.4.5. Parameters of performance evaluation

In the present study, following parameters (Krivák & Hoksza, 2015) were utilized to evaluate the performance of ANN classifier for the classification of LBS and non-LBS site residues:

$$Recall = \frac{tp}{tp + fn}$$

here, tp = true positive (correctly predicted LBS residue) and fn = false negative (wrongly predicted LBS residue) cases.

$$Precision = \frac{tp}{tp + fp}$$

here, fp = false positive (wrongly predicted non-LBS residues) cases

$$MCC = \frac{(tp \times tn) - (fp \times fn)}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}$$

here, tn = true negative (correctly predicted non-LBS residues)

However, success rate was utilized to evaluate the efficiency of ANN based LBS prediction.

$$Success\ rate = \frac{Number\ of\ hits}{Total\ proteins} \times 100$$

here, *hit* means a protein for which at least of one residue predicted correctly as LBS residue comes within range of 4 Å distance (Krivák & Hoksza, 2015) of ligand molecule for that protein.

3. Results and discussion

The present study proposed an ANN based method for the prediction of LBS localized outside deep cavities and propose its utilization as an alternative to cavity based LBS prediction methods in animal proteins. Moreover, it was interesting to check whether the cases of failure of MetaPocket 2.0 could be overcome by this new ANN based LBS prediction method. In this regard, the outcome of performance of this proposed ANN based prediction method was given in terms of efficiency of the trained network as well as efficiency of the LBS prediction on different test data sets.

3.1. Efficiency of trained ANN for LBS prediction and its basis

The recall, precision and Matthews Correlation Coefficient (MCC) were 82.37, 75.28 and 55.56, respectively for the training set and 78.53, 71.65 and 47.68, respectively for the test set A of Data set I of 75 proteins (Table 3).

There are well known facts that the LBS residues are more conserved (Panchenko et al., 2004; Liang et al., 2006) and are found more in loop or coil regions than other structures (Regad et al., 2011). Seven out of eight physiochemical properties mentioned as features 1 to 7 were taken from the work of Guo et al. (2008) where these were utilized for protein-protein interaction study with high efficiency. The present study utilized these features for the first time for the LBS prediction. Moreover, some residues, particularly Arg, His, Trp and Tyr are more frequent in LBS (Villar & Kauvar, 1994). The present study was focused to explore and utilize as many as reported discriminant features for LBS and non-LBS residues to classify them. Other features included conservation pattern and structural profiles of individual amino acid residues of proteins, physiochemical properties, atomic and residual properties of standard 20 amino acids utilized for the training of neural network for LBS prediction were the basis of the performance of this classifier. The present study added a new feature i.e. depth of residue, calculated as the average distance of all the atoms of each residue from the center of gravity (CG) of the protein. Depth of the residue was found to be an important geometric feature added for the first time in the present study. Efficiency (in terms of recall, precision and MCC) of the trained network was found to be comparable with the efficiency of classifier mentioned in the work of Krivák and Hoksza (2015) for various mentioned data sets except CHEN11 dataset, however, data set utilized in the present study was different where LBS information was available in PDB files to correctly train the network.

Table 3
Performance of trained network for the data set I.

Data set I	Recall	Precision	MCC
Training set	82.37	75.28	55.56
Test set I	78.53	71.65	47.68

3.2. Efficiency of LBS prediction

Data set II of 25 proteins as test set II (for which MetaPocket2.0 failed) was utilized for the evaluation of efficiency of LBS prediction by testing the trained network. Efficiency of this prediction is measured in terms of success rate which was 76% (as shown in Table 4) for the 25 animal proteins where MetaPocket 2.0 failed to predict LBS, though both have almost equal search subspace in terms of roughest patch from R-subspace and cavity subspace from MetaPoacket2.0. Detail of these proteins was given in Table 5 which includes 08 proteins for which MetaPocket 2.0 failed to provide any output, 11 proteins where whole cavity-subspace failed to localize LBS and 06 proteins where top three cavities failed to include LBS, where LBS localized outside the deep cavity. However, the proposed ANN + R-subspace based LBS failed only for 06 proteins out of these 25 proteins with the success rate of 76%.

It was found to be encouraging that the proposed ANN + R-subspace based LBS prediction method was successful for 76% cases in comparison to 0% success rate of MetaPocket2.0 for the set of 25 proteins. It is also much higher in comparison to results mentioned by Krivák and Hoksza (2018) as it was accessed on the proteins where LBS localized outside deep cavities. The increase of success rate of this proposed ANN based LBS prediction method might be attributed due to the inclusion of geometric roughness in the spatial distribution of atoms within the design architecture of R-subspace (Singh & Lahiri, 2017a) to get the roughest patch on the protein surface and the utilization of improved Rotating cylindrical probe method proposed by Singh and Lahiri (2017b) for the surface extraction to extract the roughest patch as the LBS is a surface phenomenon.

4. Conclusion

The present study helps to overcome the problem of occasional

Table 4
Success rate of cavity-subspace provided by MetaPocket 2.0 and ANN based LBS prediction method in Data set II.

Data set II	Success rate of LBS prediction	
	MetaPocket 2.0	ANN + R-subspace based method
Test set II	0/25 proteins	19/25 proteins (76%)

Table 5
Details of proteins where LBS prediction failed by both of the compared approaches MetaPocket 2.0 & ANN based method.

Data set II (Test set II) where MetaPocket 2.0 failed to predict LBS		ANN + R-subspace based method failed
No output obtained	1h7x_A, 2bdm_A, 3iyt_A, 2gsk_A, 2nvu_B, 2bxg_A, 2c6n_A, 3b9m_A	1h7x_A
Whole cavity-subspace failed	2bxn_A, 2f2q_A, 1jxm_A, 1ltq_A, 1od2_B, 1qgj_A, 1r6n_A, 1tz8_A, 3h6t_A, 1usq_A, 1lxf_C	1jxm_A, 1ltq_A, 1od2_B
Top 3 cavities failed	2vdm_B, 1tt6_A, 2hzp_A, 3ba0_A, 2xh1_A, 3cfq_A	2vdm_B, 3ba0_A

failure of cavity based LBS prediction in animal proteins where they localize outside deep cavities by providing an ANN + R-subspace based LBS prediction method. It was found that the proposed method could serve as a better alternative with 76% better success rate than the cavity based LBS prediction methods where these methods failed due the localization of LBS outside deep cavity in case of about 17% of the proteins. The result was interpreted in the light of two properties, utilization of the roughest patch extracted on the protein surface using R-subspace capable of localizing LBS where cavity-subspace fails and utilization of more available discriminant features between LBS and non-LBS residues for the training of artificial neural network classifier. Therefore, this approach could be utilized as a better alternative to MetaPocket 2.0, a consensus method for LBS prediction, to enhance the success rate of LBS prediction, where LBS localizes outside deep cavities. The prediction of LBS outside deep cavities will facilitate in drug designing for the proteins where it is not possible due to lack of LBS information as the geometrical LBS prediction methods rely mainly on extraction of deep cavities.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

CRediT authorship contribution statement

Conceptualization, draft manuscript writing and rewriting, data analysis, KS; reviewing, finalizing and rewriting the manuscript, YPSM. All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We shared the PDB ids of the proteins of data set I in the Supplementary Table 1

Acknowledgements

The authors gratefully acknowledge Guru Angad Dev Veterinary and Animal Sciences University, Ludhiana, India for funding scheme RKVY-13 (2020-21) component B.1: establishment of centralized research laboratory and Prof. T. Lahiri, Dean (Academics & Research), Indian Institute of Information Technology, Allahabad to pursue this research work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crstbi.2024.100144.

References

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., Ben-Tal, N., 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucl Acids Res 44, W344–W350. https://doi.org/10.1093/nar/gkw408.

Brady Jr., G.P., Stouten, P.F., 2000. Fast prediction and visualization of protein binding pockets with PASS. J. Comput. Aided Mol. Des. 14, 383–401. https://doi.org/10.1023/A:1008124202956.

Brylinski, M., Skolnick, J., 2008. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. Proc. Natl. Acad. Sci. U.S.A. 105, 129–134. https://doi.org/10.1073/pnas.0707684105.

Guo, Y., Yu, L., Wen, Z., Li, M., 2008. Using support vector machine combined with auto covariance to predict protein-protein interactions from protein sequences. Nucl Acids Res 36 (9), 3025–3030. https://doi.org/10.1093/nar/gkn159.

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques, third ed. Morgan Kaufmann. ISBN: 9780123814791.

Hendlich, M., Rippmann, F., Barnickel, G., 1997. LIGSITE: automatic and efficient detection of 1potential small molecule-binding sites in proteins. J. Mol. Graph. Model. 15, 359–363. https://doi.org/10.1016/s1093-3263(98)00002-3.

Huang, B., 2009. MetaPocket: a meta approach to improve protein ligand binding site prediction. OMICS 13, 325–333. https://doi.org/10.1089/omi.2009.0045.

Huang, B., Schroeder, M., 2006. LIGSITEcsc: predicting protein binding sites using the Connolly surface and degree of conservation. BMC Struct. Biol. 6, 19. https://doi.org/10.1186/1472-6807-6-19.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637. https://doi.org/10.1002/bip.360221211.

Krivák, R., Hoksza, D., 2015. Improving protein-ligand binding site prediction accuracy by classification of inner pocket points using local features. J Cheminform 7, 12. https://doi.org/10.1186/s13321-015-0059-5.

Krivák, R., Hoksza, D., 2018. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. J Cheminform 10, 39. https://doi.org/10.1186/s13321-018-0285-8.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157 (1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0.

Laskowsk, R.A., Luscombe, N.M., Swindells, M.B., Thornton, J.M., 1996. Protein clefts in molecular recognition and function. Protein Sci. 5 (12), 2438–2452. https://doi.org/10.1002/pro.5560051206.

Laskowski, R., 1995. SURFNET: a program for visualizing molecular surfaces, cavities and intermolecular interactions. J. Mol. Graph. 13, 323–330. https://doi.org/10.1016/0263-7855(95)00073-9.

Laurie, A., Jackson, R., 2005. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 21, 1908–1916. https://doi.org/10.1093/bioinformatics/bti315.

Le Guilloux, V., Schmidtke, P., Tuffery, P., 2009. Fpocket: an open source platform for ligand pocket detection. BMC Bioinf. 10, 168. https://doi.org/10.1186/1471-2105-10-168.

Levitt, D., Banaszak, L., 1992. POCKET: a computer graphics method for identifying and displaying protein cavities and their surrounding amino acids. J. Mol. Graph. 10, 229–234. https://doi.org/10.1016/0263-7855(92)80074-n.

Lewis, R.A., 1991. Clefts and binding sites in protein receptors. Meth Enzymol 202, 126–156. https://doi.org/10.1016/0076-6879(91)02010-7.

Liang, S., Zhang, C., Liu, S., Zhou, Y., 2006. Protein binding site prediction using an empirical scoring function. Nucl Acids Res 34 (13), 3698–3707. https://doi.org/10.1093/nar/gkl454.

Nelson, D.L., Cox, M.M., 2008. Principles of Biochemistry, fifth ed. W. H. Freeman and Company, New York.

Nisius, B., Sha, F., Gohlke, H., 2012. Structure-based computational analysis of protein binding sites for function and druggability prediction. J. Biotechnol. 159 (3), 123–134. https://doi.org/10.1016/j.jbiotec.2011.12.005.

Panchenko, A.R., Kondrashov, F., Bryant, S., 2004. Prediction of functional sites by analysis of sequence and structure conservation. Protein Sci. 13 (4), 884–892. https://doi.org/10.1110/ps.03465504.

Pettit, F.K., Bare, E., Tsai, A., Bowie, J.U., 2007. HotPatch: a Statistical approach to finding biologically relevant features on protein surfaces. J. Mol. Biol. 369, 863–879. https://doi.org/10.1016/j.jmb.2007.03.036.

Puvanendrampillai, D., Mitchell, J.B., 2003. L/D Protein Ligand Database (PLD): additional understanding of the nature and specificity of protein-ligand complexes. Bioinformatics 19 (14), 1856–1857. https://doi.org/10.1093/bioinformatics/btg243.

Regad, L., Martin, J., Camproux, A.C., 2011. Dissecting protein loops with a statistical scalpel suggests a functional implication of some structural motifs. BMC Bioinf. 12, 247. https://doi.org/10.1186/1471-2105-12-247.

Reynolds, C.H., Merz, K.M., Ringe, D., 2010. Drug Design: Structure- and Ligand-Based Approaches. Cambridge University Press, Cambridge, UK. ISBN 978-0521887236.

Sankararaman, S., Sha, F., Kirsch, J.F., Jordan, M.I., Sjölander, K., 2010. Active site prediction using evolutionary and structural information. Bioinformatics 26 (5), 617–624. https://doi.org/10.1093/bioinformatics/btq008.

Singh, K., Lahiri, T., 2017a. A new search subspace to compensate failure of cavity-based localization of ligand-binding sites. Comput. Biol. Chem. 68, 6–11. https://doi.org/10.1016/j.compbiolchem.2017.01.013.

Singh, K., Lahiri, T., 2017b. An improved protein surface extraction method using rotating cylinder probe. Inter Sci. Comput Life Sci 9, 65–71. https://doi.org/10.1007/s12539-016-0201-8.

Smith, H.J., Williams, H.J., 2002. In: Liljefors, T., Krogsgaard-Larsen, P., Madsen, U. (Eds.), Textbook of Drug Design and Discovery, third ed. CRC Press. https://doi.org/10.1201/b12381.

Sotriffer, C., Klebe, G., 2002. Identification and mapping of small-molecule binding sites in proteins: computational tools for structure-based drug design. Farmaco 57 (3), 243–251. https://doi.org/10.1016/S0014-827X(02)01211-9.

Tan, K.P., Nguyen, T.B., Patel, S., Varadarajan, R., Madhusudhan, M.S., 2013. Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Res. 41, W314–W321. https://doi.org/10.1093/nar/gkr356.

Via, A., Ferr, F., Brannetti, B., Helmer-Citterich, M., 2000. Protein surface similarities: a survey of methods to describe and compare protein surfaces. Cell. Mol. Life Sci. 57 (13–14), 1970–1977. https://doi.org/10.1007/PL00000677.

Villar, O., Kauvar, L.M., 1994. Amino acid preferences at protein binding sites. FEBS Lett. 349 (1), 125–130. https://doi.org/10.1016/0014-5793(94)00648-2.

Weisel, M., Proschak, E., Schneider, G., 2007. Pocketpicker: analysis of ligand binding-sites with shape descriptors. Chem. Cent. J. 1 (1), 7. https://doi.org/10.1186/1752-153X-1-7.

Yu, J., Zhou, Y., Tanaka, I., Yao, M., 2010. Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. Bioinformatics 26, 46–52. https://doi.org/10.1093/bioinformatics/btp599.

Zhang, Z., Li, Y., Lin, B., Schroeder, M., Huang, B., 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics 27 (15), 2083–2088. https://doi.org/10.1093/bioinformatics/btr331.