1. Background and significance of protein ligand binding site research

Proteins are some of the most important elements for life. They are not only critical cellular components, but they also participate in various critical activities and processes in the life cycle of organisms, which can achieve or help achieve important biological functions. Proteins do not work independently in living organisms. They need to bind to other biomolecules or ions (such as metal ions, nucleic acids, inorganic or organic small molecules) to create specific interactions to achieve corresponding functions [1]. These molecules and ions are called ligands (Fig. 1). Particularly, intermolecular interactions between proteins and ligands, such as small compounds, occur via amino acid residues at specific positions in the protein, usually located in pocket-like regions. These specific key amino acid residues are called ligand binding sites (LBSs). LBSs have attracted much attention in the fields of molecular docking, drug-target interactions, compound design, ligand affinity prediction, and even molecular dynamics [2–6]. Thus, identification of LBSs not only helps to explore the mechanism of intermolecular interactions but also effectively explains the pathogenesis of diseases, which provides insights for drug discovery and design [7].

Compared with highly accurate but time-consuming biological experiments [8], the advantage of computational methods is that LBS predictions can be performed based on sequence and structure information without relying on annotating the biological function of protein binding residues [9]. In addition, combining multiple computational methods, or combining experimental methods with computational methods can improve both accuracy and efficiency of LBS prediction, provide valuable assistance for drug design and drug discovery researches [10–13]. The emergence of Critical Assessment of Protein Structure Prediction (CASP) [14], Continuous Automated Model Evaluation (CAMEO) projects [15], Critical Assessment of Function Annotation (CAFA) [16], PDB database [17,18], and BioLip database [19] etc. have promoted the development of this field and provided some standard evaluation indicators and relatively unified concepts and definitions. According to the definition given in BioLip, if the distance between any one of the atoms in the ligand molecule and at least one of the atoms in the amino acid residue of the protein does not exceed the sum of the radii of these two atoms plus 0.5 Å, the amino acid residue is regarded as a ligand binding residue. Since the prediction of ligand binding residues is a typical dichotomy problem from an...
In the last twenty years, under the promotion of CASP and other research goals, researchers have made great progress in the field of LBS predictions. A series of different prediction methods based on sequence information, structural templates, and three-dimensional structures have been developed. These methods employ various computational methods, including geometry or energy feature searching, sequence or structure similarity comparison, as well as machine learning related algorithms [26–31]. Recently, deep learning-based methods have stood out from machine learning methods and have drawn much attention in computational biology [32–34]. Some state-of-the-art LBS prediction methods that employ machine learning and deep learning algorithms show significant advances over traditional methods [35,36]. In this paper, we systematically introduce the background, principles, algorithms and performance of popular LBS prediction methods by clustering prediction methods into four groups according to their working principles. Particularly, this paper highlights the most recent progress in deep learning-based methods.

2. 3D structure-based LBS prediction methods

Most small ligand binding occurs in hollows or cavities on protein surfaces because high affinity can only be gained by sufficiently large interfaces [37]. This feature has been observed in spatial structures from many detailed studies of protein–ligand complexes in PDB [38]. Therefore, attempting to locate LBSs by searching for special geometry or energy features in protein structures has long been one of the most popular methods in this area. This method generally has two different implementations. One is to perform spatial geometric measurements on the protein structure to find hollows or cavities on the surface of the protein. The second is to place some probes on the surface of the protein and then to find the cavities by estimating the energy potentials between the probe and the cavities. Table 1 lists some published 3D structure-based LBS prediction methods.

The basic idea of LBS prediction methods based on spatial geometry measurements is to locate large or even the largest hollow or cavity on the protein structure by calculating and simulating some certain geometric measures from the protein structure information. Researchers have come up with many different and creative ways to accomplish this over the past few decades.

In 1997, Manfred Hendlich et al. published LIGSITE [26], which sets up some regular 3D meshes to cover the target protein. Starting from each grid point, they scan a total of 7 directions, including the x, y, and z axes and the 4 grid diagonals, and then score the grid points. If both ends of the scan line in a certain direction are included in the protein area, the point may be in the pocket, and the grid point is added by one point. After all grid points have been scanned in all directions, the candidate ligand binding residues are determined based on the final score of each point.
are ranked by the interactions between the probe and the protein. The overlapping sites clustered by different probes are used to identify the residues, the potential LBSs are determined according to the clustered residues. A dataset that contains 77 experimentally extracted and further clustered. After mapping the grid points on the entire target protein. A carbon probe and a phosphate probe are released to the grid points and the interaction forces between the molecules of each grid point probe and the protein are calculated. The grid points with higher interaction energies are allowed for years. However, these methods strongly depend on the state of the given protein 3D structure, which means that LBSs may not be discovered if the binding pocket does not exist in the apo state but is induced by protein–ligand interaction in the holo state. In many scenarios which lack the protein structures in holo states, those methods may not be valid.

### 3. Template similarity-based LBS prediction methods

Protein 3D structures provide geometry and energy clues for LBSs that allow us to make predictions using a single structure of a protein. If considering that proteins are not an independent molecule, but are evolved from others, structural or functional information can be transferred between homologous or structurally similar proteins. Hence, an LBS can be predicted using the known proteins as templates to obtain similar characteristics in structurally similar proteins. Hence, an LBS can be predicted using the known proteins as templates to obtain similar characteristics in the query protein. Template similarity-based LBS prediction methods mainly include two types: structure template-based methods and sequence template-based methods. Table 2 lists some template similarity-based LBS prediction methods that have been published in the last twenty years.

The basic idea of the structure template-based LBS prediction method is to search for the most similar proteins in databases that have been labeled with LBSs using a structure alignment algorithm and then to transfer the known LBS from the most similar proteins onto the query protein. This method takes advantage of the increasingly accumulated protein structure databases. It could be highly reliable if proteins are of significant structural similarity.
In 2008, a popular template-based ligand binding site prediction method, FINDSITE, was published [58]. For a given target protein sequence, FINDSITE uses the PROSPECTOR 3 threading algorithm [63,64] to identify a structural template that binds to the ligand from the PDB database and overlays the template with the target protein using TMalign [65]. Then, the LBSs that bound to the structural template are clustered and ranked as predictions. FINDSITE achieved a 67.3% success rate with 75.5% ranking accuracy on protein models that have a less than 35% sequence identity to the closest template structure. Although the prediction accuracy is comparable to some 3D structure-based LBS prediction methods, it can make some very unique LBS discoveries.

Later, in 2010, Mark N. Wass et al. developed the 3DLigandSite prediction method [29]. 3DLigandSite first used MAMMOTH [66] to score the similarity between a target protein and structural templates, and the 25 template proteins with the highest similarity to the target protein structure and their corresponding ligand information were selected as templates. Similar to FINDSITE, these templates are overlaid with the target protein, and these overlaid ligands are clustered using the Single linkage clustering algorithm. The cluster with the most template ligands was chosen as the basis for the prediction of the LBS. The performance of 3DLigandSite has been tested on CASP8 [67] targets with a set of 617 proteins from the FINDSITE test set and achieved an MCC of 0.64, a coverage of 71%, and an accuracy of 60%.

Up to now (December 21, 2019), 158787 protein structures have been published in the PDB [38]. However, for a large number of proteins, it is still impossible to detect their LBS using the above methods. Meanwhile, with the continuous development of sequencing technology, a huge number of protein sequences are published every year. Therefore, sequence template-based LBS prediction methods have received extensive attention. The basic idea of sequence template-based LBS prediction methods is similar to the structure template-based LBS prediction methods, that is, the alignment tool is used to align the sequence of the protein to be tested with the sequence of the known protein, and then, the template is selected according to the similarity. Finally, the ligand-binding residues of the protein to be tested are presumed by referring the known ligand-binding residues on the aligned regions.

In 2013, Yang Zhang’s team published a ligand binding site prediction method called S-SITE [31], which employs the Needleman–Wunsch algorithm [68] to align the query protein to each of the proteins in the BioLip [19] database and screens similar sequences from the query protein according to the alignment result. The residues of the query protein are aligned with the template protein residues which were annotated as binding residues. Consensus voting is used to score the alignment results of the templates. Residues that received more than 25% of the votes were considered an LBS. S-SITE achieved both an MCC of 0.51 and a Pre of 0.45 on the test datasets.

Hybrid methods have been proposed to further improve LBS predictions. A representative algorithm, TM–SITE [31], mixes the structure template-based and the sequence information-based prediction methods. The TMalign algorithm is first used to align the protein to be tested with the known template proteins. The evolutionary information of the sequence and the spatial distance information of the structure are combined to form a comprehensive scoring function to score the similarity of each template protein, and the qualified template proteins are screened from the BioLip database according to the scoring results. Finally, the ligand-binding residues of the protein being tested are predicted based on these eligible templates. TM–SITE achieved an MCC of 0.51 and a Pre of 0.59 on the test datasets.

### Table 2: Published template similarity-based LBS prediction methods.

| Method       | Type                  | Feature                                                                 | Year   |
|--------------|-----------------------|-------------------------------------------------------------------------|--------|
| FINDSITE     | Structure Template-based | PROSPECTOR 3 threading algorithm and TMalign tool are used                | 2008   |
| S-SITE       | Structure Template-based | Use an automatic approach for cluster identification and residue selection | 2011   |
| 3DLigandSite | Structure Template-based | MAMMOTH is used                                                          | 2010   |
| TM-SITE      | Structure Template-based | Use global-to-local sequence and structural comparison algorithm         | 2012   |
| webPDBinder  | Structure Template-based | Search a protein structure against a library of known binding sites and a collection of control nonbinding pockets. | 2013   |
| S-SITE       | Structure Template-based | Needleman–Wunsch algorithms are used                                     | 2013   |
| TM–SITE      | Structure Template-based | Mix Structure Template-based and Sequence Template-based method          | 2013   |

### 4. Traditional machine learning-based LBS prediction methods

The continuous development of computer technology has promoted the application of artificial intelligence-related theories and algorithms to other fields. In the study of protein LBS predictions, 3D structure-based and template similarity-based prediction methods have shown complementary advantages to LBS predictions. How to integrate that information and further improve the prediction accuracy is one of the urgent questions of this area. Many researchers try to use machine learning algorithms not only for carrying out LBS predictions but also for the binding affinity research, which has caused significant breakthroughs. Table 3 lists some traditional machine learning-based LBS prediction methods and a few related binding affinity research methods published in recent years. However, to focus the topic, we only detail a few representative LBS prediction methods listed above. Binding affinity related methods are elaborated on in the discussion.

As mentioned earlier, predicting protein ligand binding sites is a typical dichotomous problem from a mathematical point of view, and there is a state of sample imbalance. Among the many classic machine learning algorithms that can implement the dichotomy, the naive Bayesian algorithm needs to calculate the prior probability and does not apply to data with a correlation between samples. Although the logistic regression is simple to implement, its accuracy is poor because it tends to under-fit characteristics. Besides, although the KNN algorithm is fast and has low training costs,
Traditional machine learning-based LBS prediction and binding affinity research methods.

| Method                        | Machine Learning Algorithm                                    | Year  |
|-------------------------------|-----------------------------------------------------------------|-------|
| Knowledge-based QSAR approach | Kernel-Partial Least Squares (K-PLS) [70]                      | 2004  |
| Multi-RELIEF [71]             | RELIEF algorithm [72]                                           | 2007  |
| SFCScore [73]                 | multiple linear regression                                      | 2008  |
| ATPrint [74]                  | partial least squares analysis                                   | 2009  |
| ConCavity [75]                | K-Means algorithm                                               | 2009  |
| MetaPocket [76]               | hierarchical clustering algorithm                               | 2009  |
| RF-Score [4]                  | The Random Forest algorithm                                     | 2010  |
| MetaDBSite [78]               | Support Vector Machine                                          | 2011  |
| NsitePred [79]                | Support Vector Machine                                          | 2011  |
| NNSCORE [80,81]               | Artificial Neural Network                                       | 2011  |
| L1pred [30]                   | L1-Logreg Regression classifier                                 | 2012  |
| TargetS [83]                  | Support Vector Machine                                          | 2013  |
| eFndSite [84]                 | Support Vector Machine                                          | 2013  |
| VitaPred [85]                 | Support Vector Machine                                          | 2013  |
| COACH [31]                    | Support Vector Machine                                          | 2013  |
| LigandRFs [86]                | The Random Forest algorithm                                     | 2014  |
| GOML [87]                     | Support Vector Machine                                          | 2015  |
| LigandDSES [88]               | The Random Forest algorithm                                     | 2015  |
| PRANK [89]                    | The Random Forest algorithm                                     | 2015  |
| A method for protein-ligand   | Gradient Boosting Regressor                                     | 2018  |
| binding affinity prediction   |                                                                 |       |
| [90]                          |                                                                 |       |
| SANDEs [92]                   | Regression Analysis                                             | 2016  |
| P2Rank [93]                   | The Random Forest algorithm                                     | 2018  |
| COACH-D [94]                  | Support Vector Machine                                          | 2018  |
| Tab9 [95]                     | Regression Analysis                                             | 2019  |

In 2011, Jingna Si et al. developed the MetaDBSite server [78], relying on sequence information to predict protein-DNA binding residues. MetaDBSite uses SVM to integrate the results of the six predictive tools: DISIS [96], DNABindR [97], BindN [98], BindN-rf [99], DP-Bind [100] and DBS-PRED [101]. The final output is superior to any single prediction method. The prediction results returned by DISIS, DNABindR, BindN, and BindN-rf are the main input parameters of SVM, while DP-Bind and DBS-PRED provide smaller score effects as auxiliary parameters. MetaDBSite achieved ACC, Spe, Sen of 0.77 and MCC of 0.32 on a test set, which is better than any of the single methods it combined.

In 2011, Ke Chen et al. published the NsitePred algorithm [79], which predicted the five most common nucleotide residues in the PDB database. The main steps of the NsitePred algorithm are to first extract the secondary structure, relative solvent accessibility and dihedral angles, determine the PSSM profile and other information from a given protein sequence to be tested, and use sliding window technology to process the information to generate an eigenvector describing the residue. These eigenvectors are used as inputs to the SVM to obtain a classification model. The model is used to predict the protein, and the SVM-based prediction results are combined with the BLAST [102] results as the final output. In the benchmarks, NsitePred showed better performance over ATPrint [74] and GTP binder [103].

In 2013, Yang Zhang’s team published the SVM-based prediction method COACH [31]. It combines the structure template-based and sequence information-based prediction methods S-SITE and TM_SITE with the prediction results of the three methods of the new COFACTOR [104], FINDSITE [58], and ConCavity [75] as eigenvectors to the SVM for training and to form a classification model, and finally uses this classification model to output the prediction results. The benchmark results show that COACH outperforms other classical prediction algorithms (MCC = 0.54 and Pre = 0.59), making it the most popular protein LBS prediction method over the past few years.

5. Deep learning-based LBS prediction methods

In 2006, deep learning led the third wave of artificial intelligence [105], which far surpassed traditional machine learning in text classification, speech recognition, semantic modeling, image recognition, image segmentation and computer vision [106–109]. In some areas, it has even surpassed the human brain [110] and has become the most popular research branch in the field of machine learning. Therefore, an increasing number of researchers have seen the possibility of using deep learning techniques to solve complex problems in the fields of bioinformatics and medical research, such as small-compound-drug discovery, activity prediction, chemical structure design, bioimaging, and medical imaging-based diagnosis [35,90,111–114].

Deep learning is a complex machine learning technique that simulates the learning mechanism of the human brain by building and simulating the neural networks in the human brain and uses this mechanism to interpret data. Deep learning is mainly implemented in three ways: convolutional neural networks (CNNs), deep belief networks (DBNs) and self-encoding neural networks. Among them, CNN is the most popular approach used in fields other than computer science since it is relatively simple to use and generalize. CNN is a kind of feedforward neural network. Similar to traditional artificial neural networks (ANNs) [115], CNN is also composed of multiple neurons and each of them does a part of the calculation base on a part of the input and give a part of the output, as below:
where $x$ is the input, $w$ is a set of weights, and $b$ is the bias. $f(x)$ is the activation function, which makes the neural network approximate the nonlinear function so that the network can be used in a nonlinear model. As described in Fig. 3, CNNs are mainly composed of three layers: the convolution layer, the pooling layer and the fully connected layer. The convolutional layer is used to extract different local features of the input; it consists of several convolutional units, and the parameters of each convolutional unit are optimized by backpropagation [116]. The pooling layer cuts the high dimensional local features obtained by convolutional layers into several regions and calculates the maximum value or the average value of them so that new low dimensional features can be generated. Finally, the fully connected layer combines all the local features into global features and calculates the score for each final class.

DBN is a highly scalable deep neural network; it consists of multiple layers of Restricted Boltzmann Machine (RBM) [117], which is used to learn a probability distribution of the inputs. The DBN training process can be divided into two main steps: First, unsupervised training is performed for each layer of RBM independently. Then, a supervised classifier is set after the last layer of RBM to receive the output features of RBMs and generate classification results. The structure of DBNs is shown in Fig. 4.

In the past two years, some protein LBS prediction methods using deep learning techniques have been reported. Developing new deep learning-based prediction methods has become a new hotspot in LBS prediction. Table 4 lists some deep learning-based LBS prediction methods and related studies. Some representative LBS prediction methods or LBS highly related methods are introduced below.

In 2017, J. Jiménez et al. developed the DEEPSite algorithm [36] for predicting binding sites for protein ligands. The basic idea of the algorithm is to treat the protein structure as a three-dimensional image and discretize it into a mesh with certain size voxels. A series of atomic attributes, such as hydrophobicity and hydrogen bond acceptors or donors, are used as features to calculate the occupancy of each attribute on each voxel. Finally, subgrids of a certain size are sampled, and the features of the subgrid are used as inputs to the convolutional neural network. The probability of the site being labeled a binding site is output. DEEPSite was compared with Fpocket and Concavity on the same test dataset, and the result indicated that DEEPSites outperforms other methods.

In 2019, Yifeng Cui et al. developed the DeepCSeqSite algorithm [121], which used the seven characteristics of the position-specific score matrix, relative solvent accessibility, secondary structure, the dihedral angle, conservation scores, residue type and position embeddings to construct the eigenspace. Each residue in the amino
acid sequence is embedded in the eigenspace such that the amino acid sequence is converted to a feature map, and then the map is used as an input to the convolutional neural network. The output of the network is the predicted result of protein ligand binding residues. Instead of using any template, including the three-dimensional structure, DeepCSeqSite directly predicts the binding sites of protein ligands. Its performance on test datasets is significantly better than COACH, the most accurate SVM-based prediction method mentioned above.

Recently, Ingoo Lee et al. reported the DeepConv-DTI prediction model [122] to identify interactions between drugs and targets. The idea of the model is to input the entire protein sequence into a convolutional neural network, convolve the various amino acid subsequences of the protein to capture how the protein matches the local residue pattern participating in the DTI, and use that as the input to the higher layer network to build the model and extract features. The new features will connect the model to the drug signature and predict the likelihood of DTI through a higher fully connected layer in the network. By further optimizing the model, it achieves better predictions of performance. Through the model, new features will be linked to drug characteristics and predict the likelihood of DTI through a higher fully connected layer in the network. Finally, the model is further optimized to achieve better predictive performance. As a result, the local features detected by DeepConv-DTI show better performance than other protein descriptors, such as CTD and SW scores according to the authors.

In 2019, Limeng Pu et al. presented DeepDrug3D [35], a new deep learning-based binding pockets characterization and classification algorithm, which can classify nucleotide- and heme-binding sites by learning the patterns of specific molecular interactions between ligands and their protein targets. First, the ligand–protein complexes are converted into 3D pocket grids, and the physicochemical properties of binding pockets are considered and characterized. These 3D pocket grids are then voxelized into a 3D image with 14 channels. These voxels are used as inputs for a designed convolutional neural network to get the classification result. DeepDrug3D was tested on the PDB dataset of nucleotide- and heme-binding sites and achieved an accuracy of 95%, which is much better than volume- and shape-based approaches.

6. Discussion

From the long history of LBS prediction methods, we have seen that the research focus of LBS predictions has shifted from analyzing simple 3D structure features and sequence/structure similarities to the integration of multiple features. Machine learning algorithms [21,22,24,124–130] have played a critical role in this process. Particularly, the application of deep learning algorithms has begun to show great value in LBS predictions. Furthermore, information about binding affinity and crystal structures can be used as inputs to machine learning or deep learning algorithms to help complete the LBS prediction, which makes LBS predictions more closely integrated with areas such as affinity prediction and molecular docking [23,131].

With the continuous publication of more excellent machine learning and deep learning-based LBS prediction methods, other biological studies using these methods, such as protein structure and function prediction, protein–protein interaction site prediction, and drug design, have also made new breakthroughs [132–137]. For instance, in 2015, COACH was used in drug design studies targeting MARK4 regulatory enzymes related to cancer, type 2 diabetes and many other diseases [138]. In 2019, DeepDTA was used to research protein kinases to help develop a predictive model which can estimate kinase-ligand pKi values [139].

New solutions often bring new challenges and problems while solving problems. Although deep learning-based LBS prediction methods have been used and applied in the past 2 years, there are still some problems and deficiencies to this type of solution. A key problem is that deep learning algorithms often require extremely high training costs (expensive computing resources, huge training sets, etc.) compared with traditional machine learning algorithms [140,141]. Studies have also been inconclusive about whether deep learning approaches are always superior to traditional machine learning algorithms in all cases. In fact, traditional machine learning algorithms and even some 3D structure-based binding affinity prediction methods are constantly being optimized. For instance, some methods can predict binding affinity based on the known crystal structure of a specific ligand or a protein can accurately identify the key LBS [131,142–144]. Additionally, the performance of deep learning algorithms is similar to traditional machine learning algorithms in some cases with low dimensional or small amounts of data. Thus, how to take advantage of deep learning to obtain the best solution for LBS predictions in the near future is still an open question.

In addition, researchers also think that the series of LBS prediction methods mentioned in the article cannot completely solve the problem of LBS detection since there exist some cryptic sites that are not evident in the unbound protein but form upon ligand binding [145]. Conformational change is critical to reveal these cryptic sites. Thus, detecting cryptic binding sites has received lots of attention in the past few years, and molecular dynamics simulations have become one of the most popular methods for conformational sampling in this field [2,5,146–148]. For instance, Bowman and Geissler built Markov state models from molecular dynamics (MD) simulations that can identify prospective cryptic sites [149], and a series of studies have been carried out by Gorfe's team to find hidden binding sites in Ras proteins using probe-based molecular dynamics simulations [150–153]. We believe that in the future, the advanced machine learning or deep learning approaches together with protein conformational sampling technique is also likely to become a new development direction in the field of LBS prediction.
Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 2015;43:W174–81.

Yang J, Yan R, Roy A, Xu D, Poon J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods 2015;12:7.

Li G-Q, Liu Z, Shen H-B, Yu D-J. Target Mi6A: identifying N 6-methyladenosine sites from RNA sequences via position-specific nucleotide propensities and a support vector machine. IEEE Trans Nanobiosci 2016;15:674–82.

Wei Z-S, Yang J-Y, Shen H-B, Yu D-J. A cascade random forests algorithm for predicting protein-protein interaction sites. IEEE Trans Nanobiosci 2015;14:746–60.

Wei Z-S, Han K, Yang J-Y, Shen H-B, Yu D-J. Protein–protein interaction sites prediction by ensembling SVM and sample-weighted random forests. Neurocomputing 2016;193:201–12.

Wass MN, Barton G, Sternberg MJ. CombFunc: predicting protein function using heterogeneous data sources. Nucleic Acids Res 2012;40:W466–70.

Naz F, Shahbaaz M, Bisetty K, Islam A, Ahmad F, Hassan MI. Designing new kinase inhibitor derivatives as therapeutics against common complex diseases: structural basis of microtubule affinity-regulating kinase 4 (MARK4) inhibition. OMICS 2015;19:700–11.

Govinda K, Hassan MM, Sirimulla S. KinaseKiPred: a predictive model for estimating ligand-kinase inhibitor constant (pKi). BioRxiv 2019:798561.

Goodfellow I, Bengio Y, Courville A. Deep learning. MIT press; 2016.

de Ávila MB, Bitencourt-Ferreira G, de Azevedo Jr WF. Structural basis for inhibition of enoyl-[acyl carrier protein] reductase (InhA) from Mycobacterium tuberculosis. Curr Med Chem 2019.

Volkart PA, Bitencourt-Ferreira G, Souto AA, de Azevedo WF. Cyclin-dependent kinase 2 in cellular senescence and cancer. A structural and functional review. Curr Drug Targets 2019;20:716–26.

de Ávila MB, Xavier MM, P intro VO, de Azevedo Jr WF. Supervised machine learning techniques to predict binding affinity. A study for cyclin-dependent kinase 2. Biochem Biophys Res Commun 2017;484:305–10.

Cimermancic P et al. CryptoSite: expanding the druggable proteome by characterization and prediction of cryptic binding sites. J Mol Biol 2016;428:709–19.

Guterres H, Lee HS, Im W. Ligand-binding-site structure refinement using molecular dynamics with restraints derived from predicted binding site templates. J Chem Theory Comput 2019;15:6524–35.

Bowman GR, Bolin ER, Hart KM, Maguire BC, Marqusee S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. Proc Natl Acad Sci 2015;112:2734–9.

Udi Y et al. Unraveling hidden regulatory sites in structurally homologous metalloproteases. J Mol Biol 2013;425:2330–46.

Bowman GR, Geissler PL. Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. Proc Natl Acad Sci 2012;109:11681–6.

Prakash P, Hancock JF, Gorfe AA. Computational allosteric ligand binding site identification on Ras proteins. Acta Biochim Biophy Sin 2015;48:3–10.

Prakash P, Hancock JF, Gorfe AA. Binding hotspots on K-ras: Consensus ligand binding sites and other reactive regions from probe-based molecular dynamics analysis. Proteins 2015;83:898–909.

Prakash P, Sayyed-Ahmad A, Gorfe AA. pMD-membrane: a method for ligand binding site identification in membrane-bound proteins. PLoS Comput Biol 2015;11:e1004469.

Prakash P, Zhou Y, Liang H, Hancock JF, Gorfe AA. Oncogenic K-Ras binds to an anionic membrane in two distinct orientations: a molecular dynamics analysis. Biophys J 2016;110:1125–38.