Send Orders for Reprints to reprints@benthamscience.ae

Current Topics in Medicinal Chemistry, 2018, *18*, 1146-1167

REVIEW ARTICLE

Overview of Free Software Developed for Designing Drugs Based on Protein-small Molecules Interaction

Piyush Agrawal[1,2], Pawan Kumar Raghav[1], Sherry Bhalla[1], Neelam Sharma[1], and Gajendra P.S. Raghava[1,2,*]

[1]*Center for Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India;* [2]*CSIR-Institute of Microbial Technology, Chandigarh, India*

ARTICLE HISTORY

Received: April 05, 2018
Revised: June 08, 2018
Accepted: August 08, 2018

DOI:
10.2174/1568026618666180816155131

Abstract: One of the fundamental challenges in designing drug molecule against a disease target or protein is to predict binding affinity between target and drug or small molecule. In this review, our focus will be on advancement in the field of protein-small molecule interaction. This review has been divided into four major sections. In the first section, we will cover software developed for protein structure prediction. This will include prediction of binding pockets and post-translation modifications in proteins. In the second section, we will discuss software packages developed for predicting small-molecule interacting residues in a protein. Advances in the field of docking particularly advancement in the knowledge-based force fields will be discussed in the third part of the review. This section will also cover the method developed for predicting affinity between protein and drug molecules. The fourth section of the review will describe miscellaneous techniques used for designing drug molecules, like pharmacophore modelling. Our major emphasis in this review will be on computational tools that are available free for academic use

Keywords: Protein-small molecule interaction, Structure Prediction, Docking, Pharmacophore, Molecular Dynamics, Post Translational Modifications.

1. INTRODUCTION

There are millions of premature deaths per year worldwide, despite tremendous advances in the area of drug discovery. The traditional process of drug development is a costly, time consuming and complicated process that attributed to the high rate of failure of drug discovery process. In the era of Omics, where thousands of disease-associated pathogens have been already sequenced; Computer Aided Drug Designing (CADD) has emerged as the powerful technique for drug discovery [1-3]. One of the major challenges in the area of rational drug design or CADD is to identify drug targets. Fortunately, due to advancement in the field of computational biology, 3000-10000 modified proteins that are responsible for the cause of disease have been already identified [4]. Furthermore, among these proteins, nearly 400 proteins have been targeted by pharmaceutical industries for therapeutic purposes. Broadly, these proteins belong to G-Protein Coupled Receptors (GPCRs), nuclear receptors, ion channels, and enzymes [5]. Majority of the proteins performs their functions when a protein or small molecules (cofactor, substrate, *etc.*) binds to them and the interaction takes place. These interactions are of diverse nature, need careful investigation and cannot be overlooked since they play a significant role in various processes (cellular, biological and physiological) taking place in the cell. The most important role in this interaction is carried out by the size of the protein interface, which determines the stability, and specificity of the interactions. Other factors which governed these interactions are nature of residue and experimental conditions like pI, temperature, ionic strength, *etc.* [6-9]. These interactions are involved in many disease pathogenesis too for *e.g.* In prions diseases [10], Alzheimer's cervical cancer [11], bacterial infection [12], *etc.* and that's why targeting Protein-Protein Interactions (PPIs), or protein-small molecule interaction is the current choice for developing novel therapeutic molecules [13].

In contrast, advances in the field of CADD have accelerated the drug discovery processes. This CADD approach is capable of maintaining extensive data to identify the relation between target and small molecule for designing better therapeutic compounds. Moreover, Quantitative Structure-Activity Relationship (QSAR) based models are developed to understand ligand and small molecules biological activity. Furthermore, ligand and structure-based approaches were performed by virtual screening using energy calculations to calculate binding affinity, conformational state, *etc.* Interestingly, machine-learning techniques are widely used for developing QSAR models, methods for predicting ADME (absorption, distribution, metabolism, and excretion) properties

*Address correspondence to this author at the Center for Computational Biology, Indraprastha Institute of Information Technology, Okhla Industrial Estate, Phase III, New Delhi 110020, India; Tel.: +91 011 26907444; E-mail: raghava@iiitd.ac.in

1873-4294/18 $58.00+.00

of molecules and prediction of binding affinity. Also, these techniques are used to design software for docking, molecular dynamics simulation, and for pocket finding with novel scoring functions [14-18].

In this review, we have discussed freely available computational tools developed for understanding protein-small molecule interaction vital for computer-aided drug design/discovery. As shown in Fig. (**1**), these tools have been categories in following six classes; i) prediction of protein structure, ii) pocket finders, iii) identification of protein interacting residues, iv) pharmacophore based tools, v) molecular docking and vi) prediction of post-translational modification. Each class has the number of tools for nearly same activity. Thus we have mentioned most widely used tools for each class in Fig. (**1**). This will facilitate readers in using suggested tools for their scientific activity if they are not confident which tool they should use.

2. PROTEIN MODELLING

Broadly, computer-aided drug design can be classified into two categories; (i) receptor-based drug design and (ii) ligand-based drug design. In case of receptor-based drug design, one needs to have the tertiary structure of receptor or target protein. Thus receptor-based drug design is also called structure-based drug design. There is a paradigm shift in the field of structure-based drug discovery in past decades; which has reduced the cost and time of drug discovery process [19-22]. Determination of the tertiary structure of protein or drug target is one of the essential requirements for the structure-based drug design. Experimental techniques (*e.g.,* X-ray crystallography, NMR) are commonly used for determining the protein structure of a protein. Protein Data Bank (PDB) that maintains the experimentally determined protein structures has grown over the years. The present release of PDB contains more than one hundred thousand proteins structure solved by X-ray crystallography at high resolution. Despite, tremendous progresses in the field of structure determination, still there are millions of proteins whose sequence is known, but the structure is unknown. This gap is increasing at an increasing rate due to advancement in sequencing techniques. Thus, there is a need to develop methods for predicting tertiary structure of proteins from their amino acid sequence. The structure prediction of the protein is not only crucial for the structure-based drug design, but it is also essential to understand the function of proteins [23]. In this section, we will review advancement in the field computational biology particularly methods developed for predicting tertiary structure of proteins [24, 25]. In Table **1**, we have listed selected standalone software and web server commonly used by the scientific community and is freely available for public use. As shown, in Table **1**, these methods are based on different approaches. Broadly one can classify these methods based on following categories; i) homology or comparative modelling, ii) threading-based approach and iii) Ab initio or de novo modelling [26].

These methods utilize different aspects for protein structure prediction which involves secondary structure, disordered region, structures superimposition, template serach, sequence alignment, solvent accessibility, ancestor relationship, residue-residue contact, loop modelling, three-

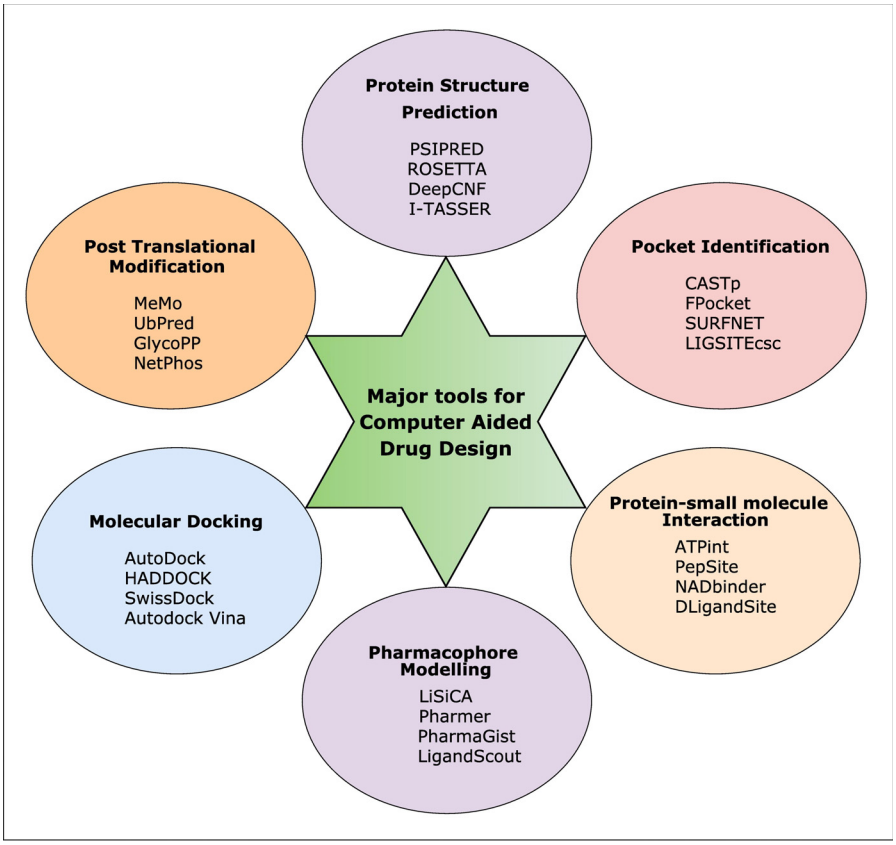

Fig. (1). Shows major categories of freely available computational tools reviewed in this paper and popular tools in each category.

Table 1. List of web servers and standalone software for predicting tertiary structure of proteins and are freely available for public use.

Name of Tool [Reference]	Description of Tool [Web Link]
PHYRE2[#] [30]	It uses HMM technique for tertiary structure prediction. [http://www.sbg.bio.ic.ac.uk/phyre2/]
CPHModels[#] [31]	It is based on secondary structure-guided profile alignment and exposure prediction. [http://www.cbs.dtu.dk/services/CPHmodels/]
SWISS MODEL[#] [32]	A fully automated protein structure homology-modelling server. [https://swissmodel.expasy.org/]
I-TASSER[###] [33]	Prediction of protein tertiary structure using threading approach. [http://zhanglab.ccmb.med.umich.edu/I-TASSER/]
EsyPred3D[#] [34]	It obtains the alignment using neural network technique and final model is built using MODELLER. [http://www.unamur.be/sciences/biologie/urbm/bioinfo/esypred/]
(PS)2[#] [35]	Protein tertiary structure prediction using comparative modelling. [http://ps2v3.life.nctu.edu.tw/]
AS2TS[#] [36]	Homology-based approach for predicting protein structure directly based on amino acid sequence. [http://proteinmodel.org/AS2TS/AS2TS/as2ts.html]
RaptorX[#] [37]	Nonlinear scoring function for predicting the structure of non-homologous proteins. [http://raptorx.uchicago.edu/StructurePrediction/predict/]
IntFOLD-TS[#] [38]	It uses single-template local consensus fold recognition approach for tertiary structure prediction. [http://www.reading.ac.uk/bioinf/IntFOLD/IntFOLD3_form.html]
Robetta[#] [39]	Comparative and de novo methods are used for prediction. [http://robetta.bakerlab.org/]
ModWeb[#] [40]	Structure prediction using the template-based algorithm. [https://modbase.compbio.ucsf.edu/modweb/]
GDFuzz3D[###] [41]	Prediction of the tertiary structure using 2D contacts map. [http://iimcb.genesilico.pl/gdserver/GDFuzz3D/]
CABS-Fold[#] [42]	It uses molecular simulation for structure prediction. [http://biocomp.chem.uw.edu.pl/CABSfold/]
MULTICOM[#] [43]	Utilize information from multiple sources for building models. [http://sysbio.rnet.missouri.edu/multicom_cluster/]
Bhageerath-H[#] [44]	Ab Initio and homology-based approach for structure prediction. [http://www.scfbio-iitd.res.in/bhageerath/bhageerath_h.jsp]
eTASSER[###] [45]	A template-based approach for protein structure prediction. [http://brylinski.cct.lsu.edu/ethread]
MUFOLD[#] [46]	Combination of multiple methods for predicting tertiary structure. [http://mufold.org/prediction.php]
QUARK[#] [47]	Algorithm for ab initio protein folding and structure prediction. [http://zhanglab.ccmb.med.umich.edu/QUARK/]
GalaxyWEB[#] [48]	Structure prediction using template-based modelling. [http://galaxy.seoklab.org/]
GENO3D[#] [49]	Tertiary structure prediction using homology modelling. [https://geno3d-prabi.ibcp.fr/]
MODELLER[###] [50]	Performs comparative modeling of given protein sequence based on its alignment to one or more protein. [https://salilab.org/modeller/]

Web Server; ## Standalone software; ### Web Server and Standalone Software.

dimensional (3D) model evaluation, identification and its refinement [27,28]. In spite of these advancements in the field, there are certain limitations of these methods, which need to be addressed. For example, identification of the suitable template, refinement of the template closer to native structure, and prediction of structures for larger proteins [29]. One of the major challenges for a user is to select best or most appropriate method among the available methods. Fortunately, CASP (Critical Assessment of Structure Prediction) organize a worldwide competition for evaluating the performance of protein structure prediction. Methods such as I-TASSER, ROSETTA, MULTICOM, QUARK, were found to perform better than other methods in recent CASP12 experiments for structure prediction [27]. For the detailed information, we advise users to visit CASP12 site to understand the current state of the art in protein structure prediction.

Although a large number of methods have been developed for predicting the tertiary structure of protein, prediction of the tertiary structure with precision is still a challenge in the absence of template/homology. In order to overcome these limitations, a number of tools have been developed to predict elements of protein structure. These elements are important for designing drug as well as for understanding function of proteins. One such element is protein secondary structure which plays a vital role in predict protein backbone because protein 3D structures are classified into their structural folds based on how their secondary structure element (helix and sheets) are folded [51, 52]. Thus secondary structure can be used for predicting the tertiary structure of the protein as well as the class of proteins [53]. Features like PSSM profile, single residue feature, or feature of neighbouring residues have been used for improving the performance of methods in the past. However, one limit imposed on secondary structure prediction method is the arbitrary definition of three states (helix, sheet, and coil) of secondary structure [54]. We can address this question by incorporating the increasing knowledge of sequence and structural data, sophisticated techniques like deep learning, *etc.* Some of the widely used secondary structure prediction tools used are PSIPRED, and DeepCNF. Similarly, a different method has been developed for predicting protein surface accessibility. In Table **2**, we list major software developed by the scientific community for predicting crucial structural component of proteins. These software packages or servers include methods developed for predicting; i) regular secondary structure (*e.g.,* Helix, Sheet), ii) irregular structures in a protein (*e.g.,* Beta-turn, Alpha-turns), and iii) surface accessibility of proteins.

Identification of cavity/surface/region on the protein, which interacts with other molecules is essential for understanding the function of a protein. This cavity or pocket on a protein structure can be used to design a drug that can modulate protein function. In case this pocket is involved in protein-protein interaction, the drug can be designed to inhibit protein-protein interaction. Similarly, in case of substrate binding pocket, the drug can be designed to inhibit binding of substrate in binding pockets. Thus, identification of binding sites or pockets is the most crucial step in drug discovery. In the past, large number of software has been developed for predicting pockets on a protein (Table **3**). Broadly, these software can be classified into three categories; i) identification of binding pocket in experimentally determined protein structures, ii) prediction of pockets in predicted protein structures, and iii) prediction of binding pockets from protein amino acid sequence. These pockets are mainly hydrophobic in nature, possess a variety of hydrogen donors and acceptors [79, 80]. Experimental techniques for pocket identification on the surface of the target molecule are complicated, costly and time-consuming. Thus, *in silico* techniques are used as an alternate technique for identification of binding sites or pocket. These methods are based on several features like evolutionary sequence conservation, Voronoi tessellation algorithm, and properties of neighbour residues for identifying the pockets/ligand binding sites on the proteins [17, 81]. However, there are several issues, which are still needs to be answered to identify the correct pocket on the proteins. These issues include ranking the optimal pocket, detection of pocket in the absence of ligands, and quality of the pocket since scoring function depends on the quality of pocket [17]. Some of the widely used pocket detection software includes FPocket, LIGSITEcsc, and CASTp. In Table **3**, we enlist major software, which is available free for academic uses.

Post-Translation Modifications (PTMs) refer to the proteins covalent modification; present in almost all proteins and play an important role in cellular processes. PTMs are the chemical modified protein obtained after translation and performs different functions [106]. It occurs on the amino acid side chains or at the protein's C- or N-termini. PTMs also provide stability to protein by protecting it from enzyme degradation. In the past number of force fields have been developed (*e.g.,* GROMOS 45a3 & 54a7) to model or simulate protein containing PTMs [107]. PTMs have been widely used as a potential drug target since they modulate the protein function. PTMs have been found to have a profound effect on protein stability, activity, and its pharmacokinetics. Thus, characterization of these PTMs and their role in biological function has become a crucial step in drug development. For example, transcription factor RUNX1 is an excellent candidate for targeted therapy in cancer. This protein is associated with PTMs like phosphorylation, acetylation, methylation, and ubiquitination [108, 109]. Thus, one of the significant challenges in the field of protein modelling is to identify different types of PTMs in a protein. Numerous methods have been developed to predict the wide range of PTMs that include phosphorylation, glycosylation, methylation, acetylation (Table **4**). Above methods are developed using features like evolutionary information, substrate primary sequence, and different types of interactions between protein-phospholipid). These methods are used to predict modification sites on proteins based on sophisticated techniques like Artificial Neural Network (ANN), Support Vector Machines (SVM), and Random Forest (RF). Once PTMs are identified on a protein, next challenge is to simulate it. Existing methods have their own limitations like dataset used for training; most of the phosphorylation site prediction methods are based on the same type of kinases datasets. Also, most of these methods take an only sequence-based feature, which lacks incorporation of experimental data [110]. In the past number of web servers have been developed for modeling and simulations of proteins/peptides

Table 2. Standalone software and web servers developed for predicting important component of protein structure like secondary structure, surface accessibility.

Name of Tool [Reference]	Description of Tool [Web Link]
APPSP2[#,###] [55]	A combination of nearest neighbour and neural network for secondary structure prediction. [http://webs.iiitd.edu.in/raghava/apssp2/]
YASPIN[#] [56]	Hidden Neural Network based approach for prediction. [http://www.ibi.vu.nl/programs/yaspinwww/]
CFSSP[#] [57]	Predict secondary structure using Chou Fasman Algorithm. [http://www.biogem.org/tool/chou-fasman/]
PSIPRED[###] [58]	Utilize the evolutionary information in the form of alignment profile for prediction. [http://bioinf.cs.ucl.ac.uk/psipred/]
Jpred4[#] [59]	It predicts secondary structure using Jnet algorithm. [http://www.compbio.dundee.ac.uk/jpred/]
DSC[#] [60]	It uses simple and linear statistical method for prediction. [https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_dsc.html]
PREDATOR[#] [61]	This method is based on recognition of potentially hydrogen-bonded residues in a single amino acid sequence. [https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_predator.html]
GOR V[#] [62]	A statistical method that utilizes evolutionary information. [http://gor.bb.iastate.edu/]
NetSurfP[###] [63]	Predicts secondary structure as well as surface accessibility. [http://www.cbs.dtu.dk/services/NetSurfP/]
Porter[#] [64]	It uses recurrent neural network algorithm for prediction. [http://distill.ucd.ie/porter/]
SCRATCH[#] [65]	Model based on machine learning technique using evolutionary information. [http://scratch.proteomics.ics.uci.edu/index.html]
SOPMA[#] [66]	Utilize alignment profile of protein family for prediction. [https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html]
ALPHAPRED[###] [67]	Prediction of alpha turns using neural network. [http://webs.iiitd.edu.in/raghava/alphapred/]
BetaTPred[###] [68]	Statistical-based method for predicting Beta Turns. [http://webs.iiitd.edu.in/raghava/betatpred/]
BetaTPred2[###] [69]	Uses neural network and multiple sequence alignment for predicting beta turns in a protein. [http://webs.iiitd.edu.in/raghava/betatpred2/]
BetaTPred3[###] [70]	Propensity based method for predicting beta turns. [http://webs.iiitd.edu.in/raghava/betatpred3/]
Bhairpred[###] [71]	Method for prediction of beta hairpin in a protein. [http://webs.iiitd.edu.in/raghava/bhairpred/]
BTEVAL[###] [72, 73]	Evaluation of beta turn predictions methods. [http://webs.iiitd.edu.in/raghava/bteval/]
Gammapred[###] [74]	Method for predicting gamma turns in proteins using neural network algorithm. [http://webs.iiitd.edu.in/raghava/gammapred/]
AcconPRED[##] [75]	Solvent surface accessibility area prediction method. [http://ttic.uchicago.edu/~majianzhu/AcconPred_package_v1.00.tar.gz]
RVPnet[###] [76]	Prediction of real valued accessible surface area. [http://www.netasa.org/rvp-net/]

(Table 2) contd….

Name of Tool [Reference]	Description of Tool [Web Link]
SARpred[#] [77]	ANN-method to predict real value of surface accessibility. [http://www.abren.net/asaview/]
SANN[###] [78]	KNN-based prediction of solvent accessibility. [http://lee.kias.re.kr/~newton/sann/]
DeepCNF[##] [16]	Deep learning-based method for predicting secondary structure element of protein. [http://raptorx.uchicago.edu/download/]

Web Server; ## Standalone software; ### Web Server and Standalone Software.

Table 3. List of software and web servers developed for predicting binding pockets on the surface of a protein.

Name of Tool [Reference]	Description of Tool [Web Link]
PockDrug[#] [82]	It uses physio-chemical and geometry based descriptors to find the pocket. [http://pockdrug.rpbs.univ-paris-diderot.fr/]
SURFNET[##] [83]	It uses PDB 3D coordinates to generate molecular surfaces and gaps. [http://www.ebi.ac.uk/thornton-srv/software/SURFNET/]
Q-siteFinder[#] [84]	Locate energetically favourable binding sites in a protein. [http://www.bioinformatics.leeds.ac.uk/qsitefinder]
CASTp[#] [85]	Identification of surface accessible pockets and interior inaccessible cavities in a protein. [http://sts.bioe.uic.edu/castp/]
PASS[##] [86]	It utilizes geometry to characterize regions of buried volume in proteins. [http://www.ccl.net/cca/software/UNIX/pass/overview.html]
LIGSITEcsc[###] [87]	LIGSITEcsc is based on the notion of surface-solvent-surface events and the degree of conservation of the involved surface residues. [http://projects.biotec.tu-dresden.de/pocket/]
bSiteFinder[#] [88]	Predicts binding sites on a protein using multiple techniques. [http://binfo.shmtu.edu.cn/bsitefinder/]
AutoSite[###] [89]	It identifies binding pockets using energetic aspects. [http://adfr.scripps.edu/AutoDockFR/autosite.html]
KVFinder[##] [90]	A method for cavity prospection and spatial characterization. [http://lnbio.cnpem.br/facilities/bioinformatics/software-2/]
AlloPred[###] [91]	It predicts protein allosteric pockets using normal mode perturbation analysis. [http://www.sbg.bio.ic.ac.uk/allopred/home]
MSPocket[##] [92]	An orientation independent approach to predict surface pockets on proteins. [http://projects.biotec.tu-dresden.de/MSPocket/]
MDPocket[##] [93]	Method for protein cavities prediction and characterization based on molecular dynamics trajectories. [http://fpocket.sourceforge.net/]
GalaxySite[#] [94]	It uses molecular docking for predicting binding sites. [http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=SITE]
POVME[##] [95]	A method for determining the pocket shape and volume characteristics. [http://rocce-vm0.ucsd.edu/data/sw/hosted/POVME/]
MolSite[##] [96]	It uses protein-ligand docking for predicting pockets. [http://presto.protein.osaka-u.ac.jp/myPresto4/index.php?lang=en]
LigandRFs[##] [97]	Prediction of protein pockets using random forest classifier. [https://sfb.kaust.edu.sa/pages/software.aspx]

(Table 3) contd….

Name of Tool [Reference]	Description of Tool [Web Link]
Phosfinder# [98]	Find phosphate binding motifs using structural comparison algorithm [http://phosfinder.bio.uniroma2.it/]
LigPlot# [99]	Generates 2D representations of protein-ligand complexes. [http://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/]
3DLigandSite# [100]	Utilizes protein-structure prediction to find ligand binding sites. [http://www.sbg.bio.ic.ac.uk/~3dligandsite/]
DrosteP## [101]	A method for evaluating the conservation of pockets detected on the protein surface by CastP. [http://www1.na.icb.cnr.it/project/drosteppy/]
PocketAnnotate# [102]	A pipeline for annotation of the binding site on a protein. [http://proline.biochem.iisc.ernet.in/pocketannotate/]
PocketDepth# [103]	Depth based algorithm for identification of ligand binding site. [http://proline.physics.iisc.ernet.in/pocketdepth/]
PocketMatch# [104]	It compares binding sites in two protein structures. [http://proline.physics.iisc.ernet.in/pocketmatch/]
PocketAlign# [105]	It aligns binding sites in protein structures. [http://proline.physics.iisc.ernet.in/pocketalign/]
FPocket## [17]	Pocket detection method based on Voronoi tessellation algorithm. [http://fpocket.sourceforge.net/]

Web Server; ## Standalone software; ### Web Server and Standalone Software.

containing different types of PTMs [111-113]. Some of the recent and widely used methods include NetPhos, PPRED for predicting phosphorylation; GlycoEP, GlycoPP, GlycoMod for predicting glycosylation sites; MeMo for predicting methylation sites; UbPred for ubiquitination.

3. SMALL MOLECULE INTERACTING RESIDUES IN A PROTEIN

Small molecules can be defined as the natural molecules which are present in the biological subject or artificially synthesized chemical (drug) and carry out important biological functions when interacts with their respective target. Small molecules are of different types such as cofactors (ATP, GTP, and NAD), nutrients which trigger operons, plant hormones, vitamins, steroids, cyclic nucleotides, *etc.* These molecules apart from being substrate and product of many biochemical reactions also regulate the function of various proteins. Binding of phosphate ions ($PO4^{3-}$) with protein enzymes results in protein phosphorylation which can either make the enzyme active or non-active [153]. Likewise, haemoglobin function of carrying and transporting oxygen *via* blood is very much dependent on its interaction with metal iron ions (Fe^{3+}) [154]. Thus, identification of native binding sites is required to predict these interactions.

Several computational methods have been designed in the last two decades for identifying ligand binding sites on their target molecule, which can be broadly classified into two categories [155]. First is the sequence-based methods which use the information only from protein sequence. Some of these features include knowledge of residue's conserved during evolution, predicted local structure properties (*e.g.,* secondary structure, relative solvent accessibility, dihedral angles), position and segment specific score of residue conservation [156-160]. Second is the structure based or template based method where structural properties of the protein-ligand complex are exploited as a feature to predict the binding site of the ligand onto given target or protein [161, 162]. Several attempts have been made to improve the accuracy of the correct binding site by combining different structural methods as well as sequence method [163, 164]. These methods have been widely used in drug designing by targeting a number of proteins associated with various diseases. One such example is the identification of inhibitors against the transforming growth factor-β-1 receptor kinase [165,166]. Incorporation of various omics data with sequencing technologies, and a combination of experimental data with computational methods would likely be improved docking. [167, 168]. Table **5** contains the list of tools designed for predicting protein-small molecule interaction.

4. PROTEIN-SMALL MOLECULE DOCKING

Docking is an important tool in the field of drug discovery that predicts the receptor-ligand interactions by estimating its binding affinity [189]. It provides an alternative to experimental assays, which are costly and time-consuming. This is the reason that number of docking publications has increased sharply over the past decade Fig. (**2**). There is tremendous progress in the field of protein-ligand docking specifically in the field of virtual screening and fragment-based drug design [190]. One of the possible reasons to this is the advancement in the field of Molecular Dynamics (MD), which is used to understand the structure-function relationship between macromolecules. Molecular interactions between these macromolecules are the basis of biological functions [191]. MD simulation has a wide range of applications such as modelling a protein structure, understanding allosteric, *etc.* MD simulation minimizes the complex

Table 4. Different software and web servers developed by the scientific community for predicting post-translation modification of amino acids in a protein.

Name of Tool [Reference]	Description of Tool [Web Link]
Tools for predicting phosphorylation	
NetPhos[#] [114]	Predicts phosphorylation sites of serine, threonine or tyrosine (K-specific and K-independent) in eukaryotic proteins. [http://www.cbs.dtu.dk/services/NetPhos/]
Scansite[#] [115]	Scanning of motifs phosphorylated by specific protein kinases. [http://scansite.mit.edu]
PhosphoSitePlus[#] [116]	Predicts the structure and function of experimentally determined PTMs in man and mouse (K-specific) [https://www.phosphosite.org/homeAction.action]
GPS[###] [117]	Predicts phosphorylation sites (K-specific) in protein kinases. [http://gps.biocuckoo.org/online.php]
KinasePhos[#] [118]	Predicts kinase specific phosphorylation sites (K-specific). [http://kinasephos2.mbc.nctu.edu.tw]
PhosphoELM[#] [119]	A database of serine, threonine and tyrosine phosphorylation sites present in the protein. [http://phospho.elm.eu.org]
PPRED[#] [120]	Predicts phosphorylation sites using evolutionary information. [http://biomecis.uta.edu/~ashis/res/ppred/]
Prediction of glycosylation	
bigPI[#] [121]	It predicts GPI-anchored modification sites in a protein. [http://mendel.imp.ac.at/gpi/gpi_server.html]
O-GlycBase[#] [122]	A database of glycoproteins with O-linked glycosylation sites. [http://www.cbs.dtu.dk/databases/OGLYCBASE/]
GlycoMod[#] [123]	Prediction of oligosaccharide structures that occur on proteins. [http://web.expasy.org/glycomod/]
YinOYang[#] [124]	ANN method to predict O-ß-GlcNAc sites in a eukaryotic protein. [http://www.cbs.dtu.dk/services/YinOYang/]
GlyProt[#] [125]	This tool predicts spatially accessible N-glycosylation sites in a protein. [http://www.glycosciences.de/glyprot/]
GPP[#] [126]	A tool for predicting N- and O-linked glycosylation sites in a protein. [http://comp.chem.nottingham.ac.uk/glyco/]
NGlycPred[#] [127]	Prediction of N-linked glycosylation sites at N-X-T/S sequence. [https://exon.niaid.nih.gov/nglycpred/]
GlycoPP[###] [128]	N- and O-linked glycosylation sites in a prokaryotic protein. [http://webs.iiitd.edu.in/raghava/glycopp/]
ProGlycProt [129]	A database of experimentally characterized prokaryotic glycoproteins [http://www.proglycprot.org/]
GlycoEP [130]	Prediction of N-, O- and C-glycosites in a eukaryotic protein. [http://webs.iiitd.edu.in/raghava/glycoep/]
NetOGlyC[#] [131]	ANN-based prediction of mucin type GalNAc O-glycosylation sites. [http://www.cbs.dtu.dk/services/NetOGlyc/]
GlycoMine[#] [132]	Predicts N-, C- and O-linked glycosylation sites in a human protein. [http://www.structbioinfor.org/Lab/GlycoMine/webserver]

(Table 4) contd….

Name of Tool [Reference]	Description of Tool [Web Link]
Prediction of miscellaneous PTM sites	
PS-SNO[#] [133]	Predicting S-nitrosylation sites using a modified GPS algorithm. [http://sno.biocuckoo.org/online.php]
iSNO-PseAAC[#] [134]	Composition-based method for predicting cysteine S-nitrosylation. [http://app.aporc.org/iSNO-PseAAC/]
PMes[#] [135]	Prediction of methylation sites using enhanced feature encoding. [http://bioinfo.ncu.edu.cn/inquiries_PMeS.aspx]
MethK[#] [136]	Identification of lysine (K) methylation sites in histone proteins. [http://csb.cse.yzu.edu.tw/MethK/]
iMethyl-PseAAC[#] [137]	Identifying methylation sites (R or K) on a protein. [http://www.jci-bioinfo.cn/iMethyl-PseAAC]
PSSMe[#] [138]	Predicting species-specific methylation sites in a protein. [http://bioinfo.ncu.edu.cn/PSSMe.aspx]
NetAcet[#] [139]	Predicting N-terminal acetylation in eukaryotes using ANN. [http://www.cbs.dtu.dk/services/NetAcet/]
ASEB[#] [140]	A method for predicting KAT-specific acetylation sites in a protein. [http://bioinfo.bjmu.edu.cn/huac/]
PSKacePred[#] [141]	Position specific analysis and lysine (K) acetylation prediction. [http://bioinfo.ncu.edu.cn/inquiries_PSKAcePred.aspx]
CSS-Palm[#] [142]	Prediction of palmitoylation site using clustering and scoring. [http://csspalm.biocuckoo.org/online.php]
SeqPalm[#] [143]	S-palmitoylation sites prediction in proteins. [http://lishuyan.lzu.edu.cn/seqpalm/]
Myristoylator[#] [144]	Predicting N-terminal myristoylation sites in a protein sequence. [http://web.expasy.org/myristoylator/]
PrePS[#] [145]	Predicting prenylation motifs in the given protein sequence. [http://mendel.imp.ac.at/sat/PrePS/]
SUMOhydro[#] [146]	Sumoylation site prediction tool based on hydrophobic features. [http://protein.cau.edu.cn/others/SUMOhydro/]
GPS-SUMO[#] [147]	Sumoylation site and sumoylation motifs in a given protein sequence. [http://sumosp.biocuckoo.org/online.php]
JASSA[#] [148]	Prediction of sumoylation site and SUMO interacting motifs. [http://www.jassa.fr]
pSumo-CD[#] [149]	Covariant discriminant algorithm for predicting sumoylation site. [http://www.jci-bioinfo.cn/pSumo-CD]
UbPred[###] [150]	In silico method for identification and analysis of ubiquitination sites in proteins. [http://www.ubpred.org]
UbiProber[#] [151]	Species-specific ubiquitination sites prediction in the proteins. [http://bioinfo.ncu.edu.cn/ubiprober.aspx]
iUbiq-Lys[#] [152]	Ubiquitination site prediction using evolutionary information. [http://www.jci-bioinfo.cn/iUbiq-Lys]

Web Server; ## Standalone software; ### Web Server and Standalone Software.

free energy and tries to bring it to stable state [192-194]. All atoms receptor flexibility into docking was introduced using MD simulations which measured its effect on docking accuracy by cross-docking [195]. The obtained stable complex, which contained flexible side chains and multiple flexible backbone segments. In contrast, accuracy was found reduced in complexes, which have flexible loops, and entirely

Table 5. List of computational methods developed for predicting ligand interacting residues in a protein.

Name of Tool [Reference]	Description of Tool [Web Link]
ATPint### [169]	Prediction of ATP interacting residues in a protein. [http://webs.iiitd.edu.in/raghava/atpint/]
TargetATPsite## [170]	A template-free method for predicting residues in ATP binding sites. [http://www.csbio.sjtu.edu.cn:8080/TargetATPsite/]
ATPsite# [171]	Sequence-based method for predicting ATP-binding residues. [http://biomine.ece.ualberta.ca/ATPsite/]
GTPbinder### [172]	Prediction of GTP binding residue in a protein sequence. [http://webs.iiitd.edu.in/raghava/gtpbinder/]
NADbinder### [173]	Prediction of NAD interacting residues in a protein. [http://webs.iiitd.edu.in/raghava/nadbinder/]
FADPred### [174]	A method for predicting FAD interacting residues in a protein. [http://webs.iiitd.edu.in/raghava/fadpred/]
ProtChemSI### [175]	A database of protein-chemical structural interaction derived from PDB. [http://pcidb.russelllab.org/]
STITCH## [176]	A database containing information of known and predicted interactions between proteins and chemicals. [http://stitch.embl.de/]
3did### [177]	A database of 3-D structural templates for domain-domain and domain-peptide interaction. [http://3did.irbbarcelona.org/]
firestar# [178]	Predicting ligand-binding residues in protein structure. [http://firedb.bioinfo.cnio.es/Php/FireStar.php]
metaDBSite# [179]	A meta web server for DNA-binding sites prediction. [http://projects.biotec.tu-dresden.de/metadbsite/]
PepSite# [180]	Predict peptide binding sites on protein surfaces. [http://pepsite2.russelllab.org/]
Pprint### [181]	A method for predicting RNA-binding residues in a protein. [http://webs.iiitd.edu.in/raghava/pprint/]
VitaPred### [182]	Prediction of vitamin-interacting residue in a protein sequence. [http://webs.iiitd.edu.in/raghava/vitapred/]
SiteHound## [183]	Identification of ligand binding site in protein structure. [http://scbx.mssm.edu/sitehound/]
webPDBinder## [184]	Prediction of small ligand binding sites in protein structures. [http://cbm.bio.uniroma2.it/pdbinder/]
3DLigandSite# [185]	Prediction of ligand-binding sites in a protein structure. [http://www.sbg.bio.ic.ac.uk/3dligandsite/]
ConCavity### [186]	Prediction of ligand-binding sites using evolutionary conservation. [http://compbio.cs.princeton.edu/concavity/]
FTSite# [187]	Detection of ligand binding site on unbound protein structures. [http://ftsite.bu.edu]
BioLip# [188]	A semi-manually curated database for ligand-protein interactions. [http://zhanglab.ccmb.med.umich.edu/BioLiP/]

Web Server; ## Standalone software; ### Web Server and Standalone Software.

flexible targets due to the increased noise that affects scoring function.

Broadly protein-ligand has two major challenges, first the generation of potential docking poses of the receptor-ligand complexes, and secondly, the scoring schemes to rank potential docking poses. Docking methods rely on scoring schemes to identify best docking pose that have a maximum binding affinity with the receptor. In the past, numerous

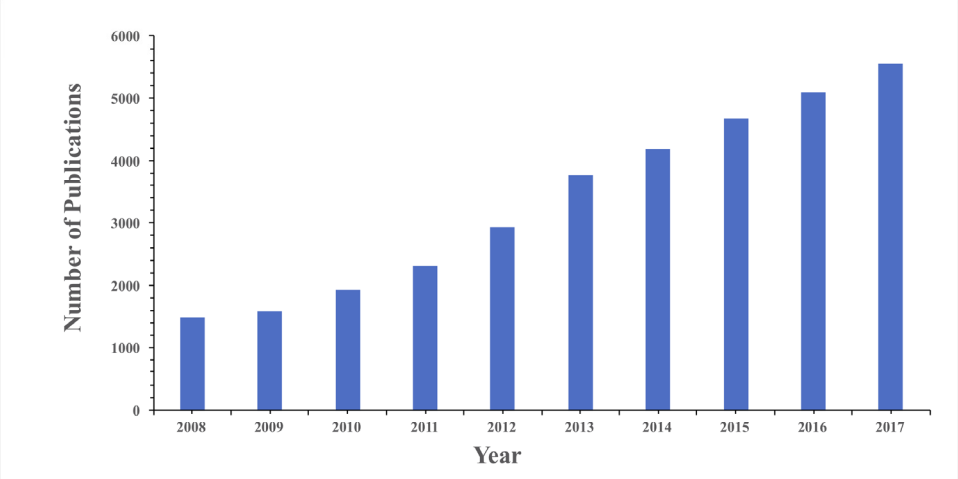

Fig. (2). Bar chart shows cumulative number of docking publications from the year 2008 to 2017; obtained by searching keyword docking in PubMed.

methods for generating docking pose using a wide range of algorithms have been developed [14, 196, 197]. Similarly, a number of strategies have been developed to rank docking pose based affinity or binding between protein and ligand [198, 199]. In Table **6**, we have listed number of commonly used software for docking and virtual screening based on binding affinity between receptor and drug molecules. Some of the widely used protein-small molecule docking software includes AutoDock, AutoDock Vina, HADDOCK, and PatchDock. These docking methods have been found to perform exceptionally well in docking ligands to their active site, like AutoDock Vina can dock small ligand with very high accuracy in comparison to other software [200]. Development of new scoring functions has shown that we can improve the scoring, docking and virtual screening comparison to old state of art methods [15]. However, there are specific challenges still needs to be tackled in this field. One of those challenges associated with docking involves representation of receptor, structural waters, side chain protonation, flexibility, ligand representation [201]. Another problem associated with docking tools is that computational method being used to generate receptor ensembles, create docking ligand databases or even creating individual ligand against large ensembles are expensive [202]. Normal mode analysis is an alternative to MD to generate receptor ensembles [203]. Additionally, an Elastic Network Model (ENM) method that signifies its importance more efficiently than MD in the identification of local conformational changes in side chains and the movement of the protein backbone [204].

5. MISCELLANEOUS TECHNIQUES

In addition to techniques described above, a number of miscellaneous techniques have been developed for designing drugs that include QSAR, pharmacophore, *etc.* [239-245]. QSAR is defined as the study of physio-chemical properties or molecular features/descriptors (two-dimensional (2D) descriptors, 3D descriptors or different fingerprints) of a chemical that correlate these molecular features with its biological activity. QSAR models can be classified into classification models or regression models. Classification model classifies the new chemical based on its molecular descriptors whereas regression model predicts the activity of a given new chemical [246, 247]. In this section, we will only cover methods that are based on pharmacophore modelling. In last few years, the pharmacophore is evolved as one of the most successful approaches in the field of drug discovery. The principle of pharmacophore based docking is defined as the molecular features of a ligand responsible for being recognized by the protein [248, 249]**.** A pharmacophore model can be obtained or generated in two different manners (i) Ligand-based pharmacophore modelling and (ii) Structure-based pharmacophore modelling. In the ligand-based pharmacophore modelling, the pharmacophore generation involves two major steps, (a) Conformational space which generates conformational flexibility in the ligand, and (b) alignment of the multiple ligands that determined essential features responsible for their bioactivity. In case of structure-based pharmacophore modelling, a 3D structure of a macromolecular target and their ligand is required. This involved a spatial relationship in between protein and ligand based on chemical features. In the past, a wide range of web server and software have been developed for designing drug molecules using pharmacophore modelling (Table **7**) [250-253]. However, pharmacophore based methods still require improvement in various fields like correct pharmacophore generation, predicting perfect ligand binding propensities, [254], lack of thorough characterization and in-depth study of ADME properties. ADME property is one of the key property which needs to be studied in depth for any pharmacophore based drug. They are critical in supporting drug discovery and drug development processes for the production of effective and safer biotherapeutics [255]. LigandScout, ZincPharmer, LiSiCA are some of the widely used tools. A complete list of tools for pharmacophore based drug discovery is given in Table **7**.

6. SUGGESTED WORKFLOW FOR CADD

The authors have suggested the following workflow to a user for initiating the drug discovery process and effectively use different tools (Fig. **3**). A user should have protein sequence and small molecule as an input. The CADD approach

Table 6. Docking tools categorized based on features, algorithms, scoring functions, availability with URLs.

Name of Tool [Reference]	Description of Tool [Web Link]
Autodock Vina[##] [205]	It is a Monte Carlo based docking software. [http://vina.scripps.edu]
Autodock[###] [206]	Genetic Algorithm based docking software for docking small molecules. [http://autodock.scripps.edu]
BSP-SLIM[#] [207]	Blind low-resolution ligand-protein docking approach. [https://zhanglab.ccmb.med.umich.edu/BSP-SLIM/]
COPICAT[#] [208]	SVM based method for predicting interactions between proteins & ligands. [http://copicat.dna.bio.keio.ac.jp]
DAIM[##] [209]	Fragment-based docking suite. [http://www.biochem-caflisch.uzh.ch/download/]
DARWIN[###] [210]	Prediction of the interaction between a protein and ligand. [http://darwin.cirad.fr/product.php]
Dock Blaster[#] [211]	A web server for virtual screening using structure-based ligand discovery. [http://blaster.docking.org/]
DockingServer[#] [212]	Integrates a number of computational chemistry software for docking. [https://www.dockingserver.com/]
DockoMatic[##] [213]	Automate the creation and management of AutoDock screening jobs. [https://sourceforge.net/projects/dockomatic/]
DockVision[###] [214]	Docking package including algorithms like Monte Carlo, Genetic algorithm, and database screening docking algorithms. [http://dockvision.sness.net/overview/overview.html]
FINDSITE(X)[#] [215]	A server for modelling GPCR and virtual screening of small-molecules. [http://cssb.biology.gatech.edu/skolnick/webservice/gpcr/index.html]
FITTED 1.0[##] [216]	A suite of programs to perform flexible protein-ligands docking. [http://www.fitted.ca]
Fleksy[##] [217]	Program for flexible and induced fit docking. [http://www.cmbi.ru.nl/software/fleksy/]
FLIPDock[##] [218]	Genetic algorithm based software to performs flexible protein-ligand docking. [http://flipdock.scripps.edu]
GalaxyDock[##] [219]	Protein-ligand docking program that allows flexibility. [http://galaxy.seoklab.org/softwares/galaxydock.html]
GEMDOCK[##] [220]	Generic evolutionary method for performing flexible ligand docking. [http://gemdock.life.nctu.edu.tw/dock/]
GFscore[#] [221]	A general non-linear consensus scoring function for docking. [http://gfscore.cnrs-mrs.fr]
GlamDock[##] [222]	Monte-Carlo based software for protein-ligand docking. [http://www.chil2.de/Glamdock.html]
GPCRautomodel[#] [223]	It automates modelling of GPCR and its interaction with small ligands. [http://genome.jouy.inra.fr/GPCRautomdl/cgi-bin/welcome.pl]
HADDOCK[###] [224]	It is a docking software for flexible as well as rigid protein-ligand docking. [http://haddock.science.uu.nl/services/HADDOCK2.2/]
HYBRID[##] [225]	This docking program is similar to FRED. [https://docs.eyesopen.com/oedocking/hybrid.html]
iScreen[#] [226]	A server for virtual screening and de novo drug design. [http://iscreen.cmu.edu.tw/intro.php]

(Table 6) contd….

Name of Tool [Reference]	Description of Tool [Web Link]
MS-DOCK## [227]	Rigid docking protocol, which generates multiple-conformation for virtual ligand screening. [http://dock.compbio.ucsf.edu/Contributed_Code/multiconfdock.htm]
ParaDockS## [228]	It performs docking for small drug-like molecules to a rigid receptor. [https://github.com/cbaldauf/paradocks]
Pardock# [229]	An automatic server for rigid protein-ligand docking. [http://www.scfbio-iitd.res.in/dock/pardock.jsp]
PatchDock### [230]	Docking software designed for protein-small molecule docking. [https://bioinfo3d.cs.tau.ac.il/PatchDock/]
PLANTS## [231]	ACO-based search engine for flexible protein side chains. [http://www.tcd.uni-konstanz.de/research/plants.php]
PLATINUM# [232]	Calculation of hydrophobic properties of molecules receptor-ligand complexes. [http://model.nmr.ru/platinum/]
Pose & Rank# [233]	In silico software for scoring protein-ligand complexes. [https://modbase.compbio.ucsf.edu/poseandrank/]
POSIT## [234]	It is a structure based docking software best suited for pose prediction. [https://docs.eyesopen.com/oedocking/posit_usage.html]
SiMMap# [235]	It recognizes interaction between protein pockets and compound moieties. [http://simfam.life.nctu.edu.tw]
Surflex-Dock## [236]	Docking program based on target active site for generating putative poses of molecules. [http://www.biopharmics.com/downloads.html]
SwissDock# [237]	Docking software for predicting interactions between protein and its ligand. [http://www.swissdock.ch]
VinaMPI## [238]	Massively parallel message-passing interface for virtual docking using Autodock Vina. [https://github.com/mokarrom/mpi-vina]

Web Server; ## Standalone software; ### Web Server and Standalone Software.

Table 7. List of pharmacophore-based tools.

Name of Tool [Reference]	Description of Tool [Web Link]
Pharmer## [256]	A new computational approach to search the pharmacophore. [http://smoothdock.ccbb.pitt.edu/pharmer/]
PharmaGist# [257]	Freely available web server for pharmacophore detection. [http://bioinfo3d.cs.tau.ac.il/PharmaGist/]
LigandScout## [258]	Tool that derives 3D pharmacophores from structural data of macromolecule/ligand complexes. [http://en.bio-soft.net/3d/LigandScout.html]
CoLibri## [259]	Compound collections used as virtual screening library. [https://www.biosolveit.de/CoLibri/]
DecoyFinder## [260]	A graphical tool which helps in finding sets of decoy molecules for a given group of active ligands. [http://urvnutrigenomica-ctns.github.io/DecoyFinder/]
MOLA## [261]	Free software for virtual screening using AutoDock4/Vina in a computer cluster using non-dedicated multi-platform computers. [http://esa.ipb.pt/biochemcore/index.php/ds/m]
NNScore## [262]	Neural network-based scoring function for the characterization of protein-ligand complexes. [http://rocce-vm0.ucsd.edu/data/sw/hosted/nnscore/]

(Table 7) contd….

Name of Tool [Reference]	Description of Tool [Web Link]
DockoMatic[##]* [263]	GUI application that is intended to ease and automate the creation and management of AutoDock jobs for HTS of ligand/receptor interactions. [https://sourceforge.net/projects/dockomatic/]
Shape-it [##] [264]	Free open-source shape-based alignment tool by representing molecules as a set of atomic Gaussians. [https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html]
GMA [##] [265]	Combined 2D/3D approach for the fast superposition of flexible chemical structures. [http://www.chil2.de/Gma.html]
Pharao[##] [266]	Pharmacophore-based tool to align small molecules. [http://silicos-it.be.s3-website-eu-west-1.amazonaws.com/software/align-it/1.0.4/align-it.html#align-it]
AutoclickChem[###] [267]	Performs in silico click-chemistry reactions. [http://nbcr-222.ucsd.edu/autoclickchem/library_1.php]
CATS [268]	It is used to perform chemical similarity searching in a collection of small molecules. [http://www.cadd.ethz.ch/software/catslight2.html]
LiSiCA[##] [269]	Ligand-based virtual screening software. [http://insilab.org/lisica/]
LigPrep 2.5[##] [270]	Performs 2D to 3D structure conversions. [https://www.schrodinger.com/ligprep]
ACPC[##] [271]	The software uses autocorrelation of partial charges for virtual ligand screening. [https://github.com/UnixJunkie/ACPC/blob/master/README.md]
ChemCom[##] [272]	This is a tool searches and compares chemical libraries. [http://bioinformatics.org/chemcom/]
React2D[##] [273]	Creates a library of molecules by combining fragment libraries based on user-defined chemical reaction. [http://molecularforecaster.com/products.html#react]
BALLOON[##] [274]	Free command-line program that creates 3D atomic coordinates from molecular connectivity *via* distance geometry and conformer ensembles using a multi-objective genetic algorithm. [http://web.abo.fi/fak/mnf/bkf/research/johnson/software.php.]
Epik[##] [275]	Enumerates ligand protonation states and tautomers in biological conditions. [https://www.schrodinger.com/Epik]
SwissSimilarity[#] [276]	Complete package for virtual screening of pharmacophores and their prediction. [http://www.swisssimilarity.ch/]
Aggregator Advisor[#] [277]	It identifies the molecule which shows aggregation. [http://advisor.bkslab.org/]
ZincPharmer[#] [278]	It searches the pharmacophore using ZINC database. [http://zincpharmer.csb.pitt.edu/]
pepMmsMIMIC[#] [279]	It performs multi search when a 3D structure of peptide is given among 17 million conformers present in MmsINC database. [http://mms.dsfarm.unipd.it/pepMMsMIMIC/]
ShaEP[##] [280]	It aligns two rigid 3D molecular structures and computes similarity index for the overlay. [http://users.abo.fi/mivainio/shaep/index.php]

Web Server; ## Standalone software; ### Web Server and Standalone Software.

can be divided into two parts, (i) Sequence-based approach or (ii) Structure-based approach. In sequence-based approach, a user can directly predict the interacting residue using protein sequence. Various methods have been developed in the past which implement such kind of methodology (*e.g.,* ATPint, NADbinder). These methods derive features directly from the sequence and apply machine learning techniques for predicting small molecule interacting residues in a given protein sequence. Another approach is the structure-based approach, where a user needs to predict the structure of the protein by using different techniques like hierarchical approach (where firstly secondary structure is predicted, which is followed by super-secondary structure prediction and finally tertiary structure prediction), comparative modeling or ab initio method. Predicted tertiary is then refined using MD techniques. After this, surface accessibility and pocket prediction are done to find the putative binding site for small molecule or a drug. After pocket prediction user

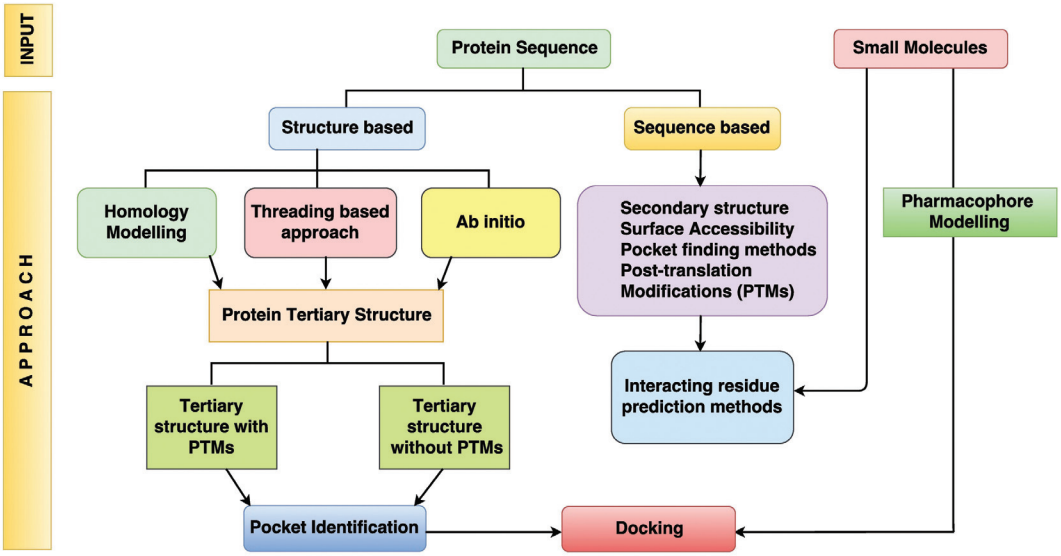

Fig. (3). Schematic representation of CADD.

needs to do virtual screening or pharmacophore based modeling for selecting the putative ligand which can interact with the protein. After ligand identification, docking is performed followed by ranking using various scoring schemes to find the native protein-ligand complex, which can be used for drug designing. A user needs to take care of different parameters before docking experiment is performed such as pre-processing of protein and ligand molecule (removal of water molecules, heteroatoms, steric clashes, the inclusion of ions, charges, *etc.*) since these parameters can have a powerful impact on the docking results.

CONCLUDING REMARKS

Considerable progress and advancement have been made in the field of protein-small molecule interaction based drug designing; still, there are many challenges, which are needs to be addressed. One of the biggest challenges is obtaining a high-quality tertiary structure of protein or macromolecule to which small molecule binds or interacts. Various computational algorithms have been developed to model protein structure since the number of the experimentally obtained structure is very less (< 1%) [281]. Over the years' researchers have developed many algorithms to model protein secondary and tertiary structures which are evaluated and ranked by CASP. Moreover, the software, web servers and algorithms developed for sequence alignment, homology modeling, fold recognition, ab initio, residue-residue contact, and fusion techniques would likely to participate in CASP. Besides, molecular docking is one of the crucial areas in drug designing and has two major challenges first, docking binding affinity calculation and second pose prediction. Predicting the native pose of the complex among the number of poses generated after docking process is one of an important issue and needs attention. There are many scoring functions, which rank these docked poses; however, they are not accurate enough to identify the native pose of the complex, which can be taken further for drug designing. The other issue associated with the molecular docking is predicting the complex stability generated by the docking. Likewise, pharmacophore tools are still unable to address the issue of modeling ligand flexibility, molecular alignment of the ligands, and selection of right training dataset for building pharmacophore model.

In this review, we have collected and compiled the tools, which are related to structure-based drug design. These tools are categorized into four main parts, protein structure prediction and pocket finding, pharmacophore prediction, PTMs finders, and molecular docking.

Although considerable advancement has been observed in the structure-based drug designing with the evident of knowledge-based scoring function, there are certain limitations associated with these tools which revealed another challenge in the field of drug discovery and identification of lead molecule by the virtual screening process.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

AUTHORS CONTRIBUTION

Authors are thankful to funding agencies Department of Science and Technology (DST-INSPIRE), Indian Council of Medical Research (ICMR), J. C. Bose National Fellowship (Department of Science and Technology), and DST-SERB N-PDF for fellowships and financial support.

ACKNOWLEDGEMENTS

P.A., P.K.R., N.S. and S.B. collected the literature. P.A., P.K.R., S.B. and N.S. prepared the tables. P.A. and N.S. draw the figures. P.A., P.K.R., S.B. and G.P.S.R. wrote the manuscript. G.P.S.R. conceived the idea and coordinated the project.

REFERENCES

[1] Śledź, P.; Caflisch, A. Protein structure-based drug design: From docking to molecular dynamics. *Curr. Opin. Struct. Biol.*, **2017**, *48*, 93-102.

[2] Singla, D.; Dhanda, S.K.; Chauhan, J.S.; Bhardwaj, A.; Brahmachari, S.K.; Open source drug discovery consortium; raghava, g.p.s. open source software and web services for designing therapeutic molecules. *Curr. Top. Med. Chem.*, **2013**, *13*, 1172-1191.

[3] Duffy, B.C.; Zhu, L.; Decornez, H.; Kitchen, D.B. Early phase drug discovery: Cheminformatics and computational techniques in identifying lead series. *Bioorg. Med. Chem.*, **2012**, *20*, 5324-5342.

[4] Drews, J.; Ryser, S. The role of innovation in drug development. *Nat. Biotechnol.*, **1997**, *15*, 1318-1319.

[5] Hopkins, A.L.; Groom, C.R. The druggable genome. *Nat. Rev. Drug Discov.*, **2002**, *1*, 727-730.

[6] Markus, M.; Benezra, R. Two Isoforms of Protein Disulfide Isomerase Alter the Dimerization Status of E2A Proteins by a Redox Mechanism. *J. Biol. Chem.*, **1999**, *274*, 1040-1049.

[7] Klemm, J.D.; Schreiber, S.L.; Crabtree, G.R. Dimerization as a regulatory mechanism in signal transduction. *Annu. Rev. Immunol.*, **1998**, *16*, 569-592.

[8] Eyster, K.M. Introduction to signal transduction: A primer for untangling the web of intracellular messengers. *Biochem. Pharmacol.*, **1998**, *55*, 1927-1938.

[9] Furukawa, Y.; Ishimori, K.; Morishima, I. Oxidation-state-dependent protein docking between cytochrome *c* and cytochrome b_5 : High-pressure laser flash photolysis study. *Biochemistry*, **2002**, *41*, 9824-9832.

[10] Cohen, F.E.; Prusiner, S.B. Pathologic conformations of prion proteins. *Annu. Rev. Biochem.*, **1998**, *67*, 793-819.

[11] Selkoe, D.J. The cell biology of beta-amyloid precursor protein and presenilin in alzheimer's disease. *Trends Cell Biol.*, **1998**, *8*, 447-453.

[12] Loregian, A.; Marsden, H.S.; Palù, G. Protein-protein interactions as targets for antiviral chemotherapy. *Rev. Med. Virol.*, **2002**, *12*, 239-262.

[13] Jin, L.; Wang, W.; Fang, G. Targeting protein-protein interaction by small molecules. *Annu. Rev. Pharmacol. Toxicol.*, **2014**, *54*, 435-456.

[14] Trott, O.; Olson, A.J. AutoDock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **2010**, *31*, 455-461.

[15] Quiroga, R.; Villarreal, M.A. Vinardo: A scoring function based on autodock vina improves scoring, docking, and virtual screening. *PLoS One*, **2016**, *11*, e0155183.

[16] Wang, S.; Peng, J.; Ma, J.; Xu, J. Protein secondary structure prediction using deep convolutional neural fields. *Sci. Rep.*, **2016**, *6*, 18962.

[17] Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformat.*, **2009**, *10*, 168.

[18] Case, D.A.; Cheatham, T.E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The amber biomolecular simulation programs. *J. Comput. Chem.*, **2005**, *26*, 1668-1688.

[19] Afshar, M.; Prescott, C.D.; Varani, G. Structure-based and combinatorial search for new rna-binding drugs. *Curr. Opin. Biotechnol.*, **1999**, *10*, 59-63.

[20] Gallego, J.; Varani, G. Targeting RNA with small-molecule drugs: Therapeutic promise and chemical challenges. *Acc. Chem. Res.*, **2001**, *34*, 836-843.

[21] Erickson, J.; Neidhart, D.J.; VanDrie, J.; Kempf, D.J.; Wang, X.C.; Norbeck, D.W.; Plattner, J.J.; Rittenhouse, J.W.; Turon, M.; Wideburg, N. Design, activity, and 2.8 a crystal structure of a c2 symmetric inhibitor complexed to hiv-1 protease. *Science*, **1990**, *249*, 527-533.

[22] Roberts, N.A.; Martin, J.A.; Kinchington, D.; Broadhurst, A. V; Craig, J.C.; Duncan, I.B.; Galpin, S.A.; Handa, B.K.; Kay, J.; Kröhn, A. Rational design of peptide-based hiv proteinase inhibitors. *Science*, **1990**, *248*, 358-361.

[23] Schmidt, T.; Bergner, A.; Schwede, T. Modelling three-dimensional protein structures for applications in drug design. *Drug Discov. Today*, **2014**, *19*, 890-897.

[24] Schwede, T.; Sali, A.; Honig, B.; Levitt, M.; Berman, H.M.; Jones, D.; Brenner, S.E.; Burley, S.K.; Das, R.; Dokholyan, N. V; Dunbrack, R.L.; Fidelis, K.; Fiser, A.; Godzik, A.; Huang, Y.J.; Humblet, C.; Jacobson, M.P.; Joachimiak, A.; Krystek, S.R.; Kortemme, T.; Kryshtafovych, A.; Montelione, G.T.; Moult, J.; Murray, D.; Sanchez, R.; Sosnick, T.R.; Standley, D.M.; Stouch, T.; Vajda, S.; Vasquez, M.; Westbrook, J.D.; Wilson, I.A. Outcome of a workshop on applications of protein models in biomedical research. *Structure*, **2009**, *17*, 151-159.

[25] Schwede, T. Protein modeling: What happened to the "protein structure gapquot. *Structure*, **2013**, *21*, 1531-1540.

[26] Leelananda, S.P.; Lindert, S. Computational methods in drug discovery. *Beilstein J. Org. Chem.*, **2016**, *12*, 2694-2718.

[27] Kryshtafovych, A.; Monastyrskyy, B.; Fidelis, K.; Moult, J.; Schwede, T.; Tramontano, A. Evaluation of the template-based modeling in casp12. *Proteins*, **2018**, *86 Suppl 1*, 321-334.

[28] Li, J.; Deng, X.; Eickholt, J.; Cheng, J. Designing and benchmarking the multicom protein structure prediction system. *BMC Struct. Biol.*, **2013**, *13*, 2.

[29] Zhang, Y. Progress and challenges in protein structure prediction. *Curr. Opin. Struct. Biol.*, **2008**, *18*, 342-348.

[30] Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **2015**, *10*, 845-858.

[31] Nielsen, M.; Lundegaard, C.; Lund, O.; Petersen, T.N. CPHmodels-3.0--remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Res.*, **2010**, *38*, W576-81.

[32] Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Gallo Cassarino, T.; Bertoni, M.; Bordoli, L.; Schwede, T. SWISS-model: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.*, **2014**, *42*, W252-8.

[33] Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The i-tasser suite: Protein structure and function prediction. *Nat. Methods*, **2015**, *12*, 7-8.

[34] Lambert, C.; Léonard, N.; De Bolle, X.; Depiereux, E. ESyPred3D: Prediction of proteins 3d structures. *Bioinformatics*, **2002**, *18*, 1250-1256.

[35] Huang, T.-T.; Hwang, J.-K.; Chen, C.-H.; Chu, C.-S.; Lee, C.-W.; Chen, C.-C. (PS)2: Protein structure prediction server version 3.0. *Nucleic Acids Res.*, **2015**, *43*, W338-42.

[36] Zemla, A.; Zhou, C.E.; Slezak, T.; Kuczmarski, T.; Rama, D.; Torres, C.; Sawicka, D.; Barsky, D. AS2TS system for protein structure modeling and analysis. *Nucleic Acids Res.*, **2005**, *33*, W111-5.

[37] Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-based protein structure modeling using the raptorx web server. *Nat. Protoc.*, **2012**, *7*, 1511-1522.

[38] McGuffin, L.J.; Atkins, J.D.; Salehe, B.R.; Shuid, A.N.; Roche, D.B. IntFOLD: An integrated server for modelling protein structures and functions from amino acid sequences. *Nucleic Acids Res.*, **2015**, *43*, W169-73.

[39] Kim, D.E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the robetta server. *Nucleic Acids Res.*, **2004**, *32*, W526-31.

[40] Sánchez, R.; Sali, A. Large-scale protein structure modeling of the saccharomyces cerevisiae genome. *Proc. Natl. Acad. Sci. USA*, **1998**, *95*, 13597-13602.

[41] Pietal, M.J.; Bujnicki, J.M.; Kozlowski, L.P. GDFuzz3D: A method for protein 3d structure reconstruction from contact maps, based on a non-euclidean distance function. *Bioinformatics*, **2015**, *31*, 3499-3505.

[42] Blaszczyk, M.; Jamroz, M.; Kmiecik, S.; Kolinski, A. CABS-Fold: Server for the de novo and consensus-based prediction of protein structure. *Nucleic Acids Res.*, **2013**, *41*, W406-11.

[43] Li, J.; Bhattacharya, D.; Cao, R.; Adhikari, B.; Deng, X.; Eickholt, J.; Cheng, J. The multicom protein tertiary structure prediction system. *Methods Mol. Biol.*, **2014**, *1137*, 29-41.

[44] Jayaram, B.; Dhingra, P.; Mishra, A.; Kaushik, R.; Mukherjee, G.; Singh, A.; Shekhar, S. Bhageerath-H: A Homology/ab initio hybrid server for predicting tertiary structures of monomeric soluble proteins. *BMC Bioinformat.*, **2014**, *15(Suppl 16)*, S7.

[45] Brylinski, M.; Lingam, D. eThread: A highly optimized machine learning-based approach to meta-threading and the modeling of protein tertiary structures. *PLoS One*, **2012**, *7*, e50200.

[46] Zhang, J.; Wang, Q.; Barz, B.; He, Z.; Kosztin, I.; Shang, Y.; Xu, D. MUFOLD: A new solution for protein 3d structure prediction. *Proteins*, **2010**, *78*, 1137-1152.

[47] Xu, D.; Zhang, Y. Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field. *Proteins*, **2012**, *80*, 1715-1735.

[48] Ko, J.; Park, H.; Seok, C. GalaxyTBM: Template-based modeling by building a reliable core and refining unreliable local regions. *BMC Bioinformat.*, **2012**, *13*, 198.

[49] Combet, C.; Jambon, M.; Deléage, G.; Geourjon, C. Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics*, **2002**, *18*, 213-214.

[50] Webb, B.; Sali, A. Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinforma.*, **2016**, *54*, 5.6.1-5.6.37.

[51] Dai, L.; Zhou, Y. Characterizing the existing and potential structural space of proteins by large-scale multiple loop permutations. *J. Mol. Biol.*, **2011**, *408*, 585-595.

[52] Murzin, A.G.; Brenner, S.E.; Hubbard, T.; Chothia, C. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **1995**, *247*, 536-540.

[53] Koch, O. Use of secondary structure element information in drug design: Polypharmacology and conserved motifs in protein-ligand binding and protein-protein interfaces. *Future Med. Chem.*, **2011**, *3*, 699-708.

[54] Yang, Y.; Gao, J.; Wang, J.; Heffernan, R.; Hanson, J.; Paliwal, K.; Zhou, Y. Sixty-five years of the long march in protein secondary structure prediction: The final stretch? *Brief. Bioinform.*, **2016**, bbw129.

[55] Raghava, G.P.S. APSSP2: Advanced protein secondary structure prediction server. http://webs.iiitd.edu.in/raghava/apssp2/.

[56] Lin, K.; Simossis, V.A.; Taylor, W.R.; Heringa, J. A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics*, **2005**, *21*, 152-159.

[57] Chou, P.Y.; Fasman, G.D. Prediction of protein conformation. *Biochemistry*, **1974**, *13*, 222-245.

[58] McGuffin, L.J.; Bryson, K.; Jones, D.T. The PSIPRED protein structure prediction server. *Bioinformatics*, **2000**, *16*, 404-405.

[59] Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G.J. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.*, **2015**, *43*, W389-94.

[60] King, R.D.; Saqi, M.; Sayle, R.; Sternberg, M.J. DSC: Public domain protein secondary structure predication. *Comput. Appl. Biosci.*, **1997**, *13*, 473-474.

[61] Frishman, D.; Argos, P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.*, **1996**, *9*, 133-142.

[62] Sen, T.Z.; Jernigan, R.L.; Garnier, J.; Kloczkowski, A. GOR V server for protein secondary structure prediction. *Bioinformatics*, **2005**, *21*, 2787-2788.

[63] Petersen, B.; Petersen, T.N.; Andersen, P.; Nielsen, M.; Lundegaard, C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.*, **2009**, *9*, 51.

[64] Pollastri, G.; McLysaght, A. Porter: A New, Accurate server for protein secondary structure prediction. *Bioinformatics*, **2005**, *21*, 1719-1720.

[65] Cheng, J.; Randall, A.Z.; Sweredoski, M.J.; Baldi, P. Scratch: A protein structure and structural feature prediction server. *Nucleic Acids Res.*, **2005**, *33*, W72-6.

[66] Geourjon, C.; Deléage, G. SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.*, **1995**, *11*, 681-684.

[67] Kaur, H.; Raghava, G.P.S. Prediction of alpha-turns in proteins using psi-blast profiles and secondary structure information. *Proteins*, **2004**, *55*, 83-90.

[68] Kaur, H.; Raghava, G.P.S. BetaTPred: Prediction of beta-turns in a protein using statistical algorithms. *Bioinformatics*, **2002**, *18*, 498-499.

[69] Kaur, H.; Raghava, G.P.S. Prediction of beta-turns in proteins from multiple alignment using neural network. *Protein Sci.*, **2003**, *12*, 627-634.

[70] Singh, H.; Singh, S.; Raghava, G.P.S. *In silico* platform for predicting and initiating β-turns in a protein at desired locations. *Proteins*, **2015**, *83*, 910-921.

[71] Kumar, M.; Bhasin, M.; Natt, N.K.; Raghava, G.P.S. Bhairpred: Prediction of beta-hairpins in a protein from multiple alignment information using ann and svm techniques. *Nucleic Acids Res.*, **2005**, *33*, W154-9.

[72] Kaur, H.; Raghava, G.P.S. An evaluation of beta-turn prediction methods. *Bioinformatics*, **2002**, *18*, 1508-1514.

[73] Kaur, H.; Raghava, G.P.S. BTEVAL: A server for evaluation of beta-turn prediction methods. *J. Bioinform. Comput. Biol.*, **2003**, *1*, 495-504.

[74] Kaur, H.; Raghava, G.P.S. A Neural-network based method for prediction of gamma-turns in proteins from multiple sequence alignment. *Protein Sci.*, **2003**, *12*, 923-929.

[75] Ma, J.; Wang, S. AcconPred: Predicting solvent accessibility and contact number simultaneously by a multitask learning framework under the conditional neural fields model. *Biomed Res. Int.*, **2015**, *2015*, 678764.

[76] Ahmad, S.; Gromiha, M.M.; Sarai, A. RVP-Net: Online prediction of real valued accessible surface area of proteins from single sequences. *Bioinformatics*, **2003**, *19*, 1849-1851.

[77] Garg, A.; Kaur, H.; Raghava, G.P.S. Real value prediction of solvent accessibility in proteins using multiple sequence alignment and secondary structure. *Proteins*, **2005**, *61*, 318-324.

[78] Joo, K.; Lee, S.J.; Lee, J. Sann: Solvent accessibility prediction of proteins by nearest neighbor method. *Proteins*, **2012**, *80*, 1791-1797.

[79] Filikov, A. V; Mohan, V.; Vickers, T.A.; Griffey, R.H.; Cook, P.D.; Abagyan, R.A.; James, T.L. Identification of ligands for rna targets *via* structure-based virtual screening: HIV-1 TAR. *J. Comput. Aided. Mol. Des.*, **2000**, *14*, 593-610.

[80] Lind, K.E.; Du, Z.; Fujinaga, K.; Peterlin, B.M.; James, T.L. Structure-based computational database screening, *in vitro* assay, and nmr assessment of compounds that target tar rna. *Chem. Biol.*, **2002**, *9*, 185-193.

[81] Capra, J.A.; Laskowski, R.A.; Thornton, J.M.; Singh, M.; Funkhouser, T.A. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3d structure. *PLoS Comput. Biol.*, **2009**, *5*, e1000585.

[82] Borrel, A.; Regad, L.; Xhaard, H.; Petitjean, M.; Camproux, A.-C. PockDrug: A model for predicting pocket druggability that overcomes pocket estimation uncertainties. *J. Chem. Inf. Model.*, **2015**, *55*, 882-895.

[83] Laskowski, R.A. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graph.*, **1995**, *13*, 323-330, 307-308.

[84] Laurie, A.T.R.; Jackson, R.M. Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics*, **2005**, *21*, 1908-1916.

[85] Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J. CASTp: Computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res.*, **2006**, *34*, W116-8.

[86] Brady, G.P.; Stouten, P.F. Fast prediction and visualization of protein binding pockets with PASS. *J. Comput. Aided. Mol. Des.*, **2000**, *14*, 383-401.

[87] Huang, B.; Schroeder, M. LIGSITEcsc: Predicting ligand binding sites using the connolly surface and degree of conservation. *BMC Struct. Biol.*, **2006**, *6*, 19.

[88] Gao, J.; Zhang, Q.; Liu, M.; Zhu, L.; Wu, D.; Cao, Z.; Zhu, R. bSiteFinder, an improved protein-binding sites prediction server based on structural alignment: More accurate and less time-consuming. *J. Cheminform.*, **2016**, *8*, 38.

[89] Ravindranath, P.A.; Sanner, M.F. AutoSite: An automated approach for pseudo-ligands prediction-from ligand-binding sites identification to predicting key ligand atoms. *Bioinformatics*, **2016**, *32*, 3142-3149.

[90] Oliveira, S.H.P.; Ferraz, F.A.N.; Honorato, R. V; Xavier-Neto, J.; Sobreira, T.J.P.; de Oliveira, P.S.L. KVFinder: Steered identification of protein cavities as a pymol plugin. *BMC Bioinformat.*, **2014**, *15*, 197.

[91] Greener, J.G.; Sternberg, M.J.E. AlloPred: Prediction of allosteric pockets on proteins using normal mode perturbation analysis. *BMC Bioinformat.*, **2015**, *16*, 335.

[92] Zhu, H.; Pisabarro, M.T. MSPocket: An orientation-independent algorithm for the detection of ligand binding pockets. *Bioinformatics*, **2011**, *27*, 351-358.

[93] Schmidtke, P.; Bidon-Chanal, A.; Luque, F.J.; Barril, X. MDpocket: Open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics*, **2011**, *27*, 3276-3285.

[94] Heo, L.; Shin, W.-H.; Lee, M.S.; Seok, C. GalaxySite: Ligand-binding-site prediction by using molecular docking. *Nucleic Acids Res.*, **2014**, *42*, W210-4.

[95] Durrant, J.D.; Votapka, L.; Sørensen, J.; Amaro, R.E. POVME 2.0: An enhanced tool for determining pocket shape and volume characteristics. *J. Chem. Theory Comput.*, **2014**, *10*, 5047-5056.

[96] Fukunishi, Y.; Nakamura, H. Prediction of ligand-binding sites of proteins by molecular docking calculation for a random ligand library. *Protein Sci.*, **2011**, *20*, 95-106.

[97] Chen, P.; Huang, J.Z.; Gao, X. LigandRFs: Random forest ensemble to identify ligand-binding residues from sequence information alone. *BMC Bioinformat.*, **2014**, *15 Suppl 15*, S4.

[98] Parca, L.; Mangone, I.; Gherardini, P.F.; Ausiello, G.; Helmer-Citterich, M. Phosfinder: A web server for the identification of phosphate-binding sites on protein structures. *Nucleic Acids Res.*, **2011**, *39*, W278-82.

[99] Wallace, A.C.; Laskowski, R.A.; Thornton, J.M. LIGPLOT: A program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.*, **1995**, *8*, 127-134.

[100] Wass, M.N.; Kelley, L.A.; Sternberg, M.J.E. 3DLigandsite: Predicting ligand-binding sites using similar structures. *Nucleic Acids Res.*, **2010**, *38*, W469-73.

[101] Cammisa, M.; Correra, A.; Andreotti, G.; Cubellis, M.V. Identification and analysis of conserved pockets on protein surfaces. *BMC Bioinformat.*, **2013**, *14 Suppl 7*, S9.

[102] Anand, P.; Yeturu, K.; Chandra, N. PocketAnnotate: Towards site-based function annotation. *Nucleic Acids Res.*, **2012**, *40*, W400-8.

[103] Kalidas, Y.; Chandra, N. PocketDepth: A new depth based algorithm for identification of ligand binding sites in proteins. *J. Struct. Biol.*, **2008**, *161*, 31-42.

[104] Yeturu, K.; Chandra, N. PocketMatch: A new algorithm to compare binding sites in protein structures. *BMC Bioinformat.*, **2008**, *9*, 543.

[105] Yeturu, K.; Chandra, N. PocketAlign a novel algorithm for aligning binding sites in protein structures. *J. Chem. Inf. Model.*, **2011**, *51*, 1725-1736.

[106] Wang, Y.-C.; Peterson, S.E.; Loring, J.F. Protein post-translational modifications and regulation of pluripotency in human stem cells. *Cell Res.*, **2014**, *24*, 143-160.

[107] Petrov, D.; Margreitter, C.; Grandits, M.; Oostenbrink, C.; Zagrovic, B. A systematic framework for molecular dynamics simulations of protein post-translational modifications. *PLoS Comput. Biol.*, **2013**, *9*, e1003154.

[108] Goyama, S.; Schibler, J.; Cunningham, L.; Zhang, Y.; Rao, Y.; Nishimoto, N.; Nakagawa, M.; Olsson, A.; Wunderlich, M.; Link, K.A.; Mizukawa, B.; Grimes, H.L.; Kurokawa, M.; Liu, P.P.; Huang, G.; Mulloy, J.C. Transcription Factor RUNX1 promotes survival of acute myeloid leukemia cells. *J. Clin. Invest.*, **2013**, *123*, 3876-3888.

[109] Goyama, S.; Huang, G.; Kurokawa, M.; Mulloy, J.C. Posttranslational modifications of RUNX1 as potential anticancer targets. *Oncogene*, **2015**, *34*, 3483-3492.

[110] Audagnotto, M.; Dal Peraro, M. Protein post-translational modifications:in silicoprediction tools and molecular modeling. *Comput. Struct. Biotechnol. J.*, **2017**, *15*, 307-319.

[111] Singh, S.; Singh, H.; Tuknait, A.; Chaudhary, K.; Singh, B.; Kumaran, S.; Raghava, G.P.S. PEPstrMOD: Structure prediction of peptides containing natural, non-natural and modified residues. *Biol. Direct*, **2015**, *10*, 73.

[112] Khoury, G.A.; Thompson, J.P.; Smadbeck, J.; Kieslich, C.A.; Floudas, C.A. Forcefield_PTM: Ab initio charge and amber forcefield parameters for frequently occurring post-translational modifications. *J. Chem. Theory Comput.*, **2013**, *9*, 5653-5674.

[113] Margreitter, C.; Petrov, D.; Zagrovic, B. Vienna-PTM Web Server: A toolkit for md simulations of protein post-translational modifications. *Nucleic Acids Res.*, **2013**, *41*, W422-6.

[114] Blom, N.; Sicheritz-Pontén, T.; Gupta, R.; Gammeltoft, S.; Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics*, **2004**, *4*, 1633-1649.

[115] Obenauer, J.C.; Cantley, L.C.; Yaffe, M.B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. *Nucleic Acids Res.*, **2003**, *31*, 3635-3641.

[116] Hornbeck, P.V; Kornhauser, J.M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.*, **2012**, *40*, D261-70.

[117] Xue, Y.; Zhou, F.; Zhu, M.; Ahmed, K.; Chen, G.; Yao, X. GPS: A comprehensive www server for phosphorylation sites prediction. *Nucleic Acids Res.*, **2005**, *33*, W184-7.

[118] Wong, Y.-H.; Lee, T.-Y.; Liang, H.-K.; Huang, C.-M.; Wang, T.-Y.; Yang, Y.-H.; Chu, C.-H.; Huang, H.-D.; Ko, M.-T.; Hwang, J.-K. KinasePhos 2.0: A web server for identifying protein kinase-specific phosphorylation sites based on sequences and coupling patterns. *Nucleic Acids Res.*, **2007**, *35*, W588-94.

[119] Dinkel, H.; Chica, C.; *Via*, A.; Gould, C.M.; Jensen, L.J.; Gibson, T.J.; Diella, F. Phospho.ELM: A database of phosphorylation sites--update 2011. *Nucleic Acids Res.*, **2011**, *39*, D261-7.

[120] Biswas, A.K.; Noman, N.; Sikder, A.R. Machine learning approach to predict protein phosphorylation sites by incorporating evolutionary information. *BMC Bioinformat.*, **2010**, *11*, 273.

[121] Eisenhaber, B.; Bork, P.; Yuan, Y.; Löffler, G.; Eisenhaber, F. Automated annotation of gpi anchor sites: Case study *C. elegans*. *Trends Biochem. Sci.*, **2000**, *25*, 340-341.

[122] Gupta, R.; Birch, H.; Rapacki, K.; Brunak, S.; Hansen, J.E. O-GLYCBASE Version 4.0: A revised database of o-glycosylated proteins. *Nucleic Acids Res.*, **1999**, *27*, 370-372.

[123] Cooper, C.A.; Gasteiger, E.; Packer, N.H. GlycoMod--a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*, **2001**, *1*, 340-349.

[124] Gupta, R.; Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.*, **2002**, 310-322.

[125] Bohne-Lang, A.; von der Lieth, C.-W. GlyProt: In silico glycosylation of proteins. *Nucleic Acids Res.*, **2005**, *33*, W214-9.

[126] Hamby, S.E.; Hirst, J.D. Prediction of glycosylation sites using random forests. *BMC Bioinformat.*, **2008**, *9*, 500.

[127] Chuang, G.-Y.; Boyington, J.C.; Joyce, M.G.; Zhu, J.; Nabel, G.J.; Kwong, P.D.; Georgiev, I. Computational prediction of n-linked glycosylation incorporating structural properties and patterns. *Bioinformatics*, **2012**, *28*, 2249-2255.

[128] Chauhan, J.S.; Bhat, A.H.; Raghava, G.P.S.; Rao, A. GlycoPP: A webserver for prediction of n- and o-glycosites in prokaryotic protein sequences. *PLoS One*, **2012**, *7*, e40155.

[129] Bhat, A.H.; Mondal, H.; Chauhan, J.S.; Raghava, G.P.S.; Methi, A.; Rao, A. ProGlycProt: A repository of experimentally characterized prokaryotic glycoproteins. *Nucleic Acids Res.*, **2012**, *40*, D388-93.

[130] Chauhan, J.S.; Rao, A.; Raghava, G.P.S. *In silico* platform for prediction of n-, o- and c-glycosites in eukaryotic protein sequences. *PLoS One*, **2013**, *8*, e67008.

[131] Steentoft, C.; Vakhrushev, S.Y.; Joshi, H.J.; Kong, Y.; Vester-Christensen, M.B.; Schjoldager, K.T.-B.G.; Lavrsen, K.; Dabelsteen, S.; Pedersen, N.B.; Marcos-Silva, L.; Gupta, R.; Bennett, E.P.; Mandel, U.; Brunak, S.; Wandall, H.H.; Levery, S.B.; Clausen, H. Precision mapping of the human O-GalNAc glycoproteome through simplecell technology. *EMBO J.*, **2013**, *32*, 1478-1488.

[132] Li, F.; Li, C.; Wang, M.; Webb, G.I.; Zhang, Y.; Whisstock, J.C.; Song, J. GlycoMine: A machine learning-based approach for predicting n-, c- and o-linked glycosylation in the human proteome. *Bioinformatics*, **2015**, *31*, 1411-1419.

[133] Xue, Y.; Liu, Z.; Gao, X.; Jin, C.; Wen, L.; Yao, X.; Ren, J. GPS-SNO: Computational prediction of protein s-nitrosylation sites with a modified gps algorithm. *PLoS One*, **2010**, *5*, e11290.

[134] Xu, Y.; Ding, J.; Wu, L.-Y.; Chou, K.-C. iSNO-PseAAC: Predict cysteine s-nitrosylation sites in proteins by incorporating position specific amino acid propensity into pseudo amino acid composition. *PLoS One*, **2013**, *8*, e55844.

[135] Shi, S.-P.; Qiu, J.-D.; Sun, X.-Y.; Suo, S.-B.; Huang, S.-Y.; Liang, R.-P. PMeS: Prediction of methylation sites based on enhanced feature encoding scheme. *PLoS One*, **2012**, *7*, e38772.

[136] Lee, T.-Y.; Chang, C.-W.; Lu, C.-T.; Cheng, T.-H.; Chang, T.-H. Identification and characterization of lysine-methylated sites on

histones and non-histone proteins. *Comput. Biol. Chem.*, **2014**, *50*, 11-18.
[137] Qiu, W.-R.; Xiao, X.; Lin, W.-Z.; Chou, K.-C. iMethyl-PseAAC: Identification of protein methylation sites *via* a pseudo amino acid composition approach. *Biomed. Res. Int.*, **2014**, *2014*, 947416.
[138] Wen, P.-P.; Shi, S.-P.; Xu, H.-D.; Wang, L.-N.; Qiu, J.-D. Accurate in silico prediction of species-specific methylation sites based on information gain feature optimization. *Bioinformatics*, **2016**, *32*, 3107-3115.
[139] Kiemer, L.; Bendtsen, J.D.; Blom, N. NetAcet: Prediction of N-Terminal Acetylation Sites. *Bioinformatics*, **2005**, *21*, 1269-1270.
[140] Wang, L.; Du, Y.; Lu, M.; Li, T. ASEB: A web server for kat-specific acetylation site prediction. *Nucleic Acids Res.*, **2012**, *40*, W376-9.
[141] Suo, S.-B.; Qiu, J.-D.; Shi, S.-P.; Sun, X.-Y.; Huang, S.-Y.; Chen, X.; Liang, R.-P. Position-specific analysis and prediction for protein lysine acetylation based on multiple features. *PLoS One*, **2012**, *7*, e49108.
[142] Zhou, F.; Xue, Y.; Yao, X.; Xu, Y. CSS-Palm: Palmitoylation site prediction with a clustering and scoring strategy (CSS). *Bioinformatics*, **2006**, *22*, 894-896.
[143] Li, S.; Li, J.; Ning, L.; Wang, S.; Niu, Y.; Jin, N.; Yao, X.; Liu, H.; Xi, L. *In Silico* identification of protein s-palmitoylation sites and their involvement in human inherited disease. *J. Chem. Inf. Model.*, **2015**, *55*, 2015-2025.
[144] Bologna, G.; Yvon, C.; Duvaud, S.; Veuthey, A.-L. N-terminal myristoylation predictions by ensembles of neural networks. *Proteomics*, **2004**, *4*, 1626-1632.
[145] Maurer-Stroh, S.; Eisenhaber, F. Refinement and prediction of protein prenylation motifs. *Genome Biol.*, **2005**, *6*, R55.
[146] Chen, Y.-Z.; Chen, Z.; Gong, Y.-A.; Ying, G. SUMohydro: A novel method for the prediction of sumoylation sites based on hydrophobic properties. *PLoS One*, **2012**, *7*, e39195.
[147] Zhao, Q.; Xie, Y.; Zheng, Y.; Jiang, S.; Liu, W.; Mu, W.; Liu, Z.; Zhao, Y.; Xue, Y.; Ren, J. GPS-SUMO: A tool for the prediction of sumoylation sites and sumo-interaction motifs. *Nucleic Acids Res.*, **2014**, *42*, W325-30.
[148] Beauclair, G.; Bridier-Nahmias, A.; Zagury, J.-F.; Saïb, A.; Zamborlini, A. JASSA: A comprehensive tool for prediction of sumoylation sites and sims. *Bioinformatics*, **2015**, *31*, 3483-3491.
[149] Jia, J.; Zhang, L.; Liu, Z.; Xiao, X.; Chou, K.-C. pSumo-CD: Predicting sumoylation sites in proteins with covariance discriminant algorithm by incorporating sequence-coupled effects into general pseaac. *Bioinformatics*, **2016**, *32*, 3133-3141.
[150] Radivojac, P.; Vacic, V.; Haynes, C.; Cocklin, R.R.; Mohan, A.; Heyen, J.W.; Goebl, M.G.; Iakoucheva, L.M. Identification, analysis, and prediction of protein ubiquitination sites. *Proteins*, **2010**, *78*, 365-380.
[151] Chen, X.; Qiu, J.-D.; Shi, S.-P.; Suo, S.-B.; Huang, S.-Y.; Liang, R.-P. Incorporating key position and amino acid residue features to identify general and species-specific ubiquitin conjugation sites. *Bioinformatics*, **2013**, *29*, 1614-1622.
[152] Wang, P.; Xiao, X.; Chou, K.-C. NR-2L: A two-level predictor for identifying nuclear receptor subfamilies based on sequence-derived features. *PLoS One*, **2011**, *6*, e23505.
[153] Burnett, G.; Kennedy, E.P. The enzymatic phosphorylation of proteins. *J. Biol. Chem.*, **1954**, *211*, 969-980.
[154] Hsia, C.C. Respiratory function of hemoglobin. *N. Engl. J. Med.*, **1998**, *338*, 239-247.
[155] Hu, X.; Dong, Q.; Yang, J.; Zhang, Y. Recognizing metal and acid radical ion-binding sites by integrating ab initio modeling with template-based transferals. *Bioinformatics*, **2016**, *32*, 3260-3269.
[156] Rausell, A.; Juan, D.; Pazos, F.; Valencia, A. Protein interactions and ligand binding: From protein subfamilies to functional specificity. *Proc. Natl. Acad. Sci. USA*, **2010**, *107*, 1995-2000.
[157] Magliery, T.J.; Regan, L. Sequence variation in ligand binding sites in proteins. *BMC Bioinformat.*, **2005**, *6*, 240.
[158] Chen, P.; Hu, S.; Zhang, J.; Gao, X.; Li, J.; Xia, J.; Wang, B. A sequence-based dynamic ensemble learning system for protein ligand-binding site prediction. *IEEE/ACM Trans. Comput. Biol. Bioinfor.*, **2016**, *13*, 901-912.
[159] Capra, J.A.; Singh, M. Predicting functionally important residues from sequence conservation. *Bioinformatics*, **2007**, *23*, 1875-1882.
[160] Chauhan, J.S.; Mishra, N.K.; Raghava, G.P. Identification of ATP binding residues of a protein from its primary sequence. *BMC Bioinformatics*, **2009**, *10*, 434.
[161] Brylinski, M.; Skolnick, J. A Threading-based method (findsite) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 129-134.
[162] Capra, J.A.; Laskowski, R.A.; Thornton, J.M.; Singh, M.; Funkhouser, T.A. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3d structure. *PLoS Comput. Biol.*, **2009**, *5*, e1000585.
[163] Hu, X.; Wang, K.; Dong, Q. Protein ligand-specific binding residue predictions by an ensemble classifier. *BMC Bioinformat.*, **2016**, *17*, 470.
[164] Yang, J.; Roy, A.; Zhang, Y. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, **2013**, *29*, 2588-2595.
[165] Singh, J.; Chuaqui, C.E.; Boriack-Sjodin, P.A.; Lee, W.C.; Pontz, T.; Corbley, M.J.; Cheung, H.-K.; Arduini, R.M.; Mead, J.N.; Newman, M.N.; Papadatos, J.L.; Bowes, S.; Josiah, S.; Ling, L.E. Successful shape-based virtual screening: The discovery of a potent inhibitor of the type I TGFbeta receptor kinase (TbetaRI). *Bioorg. Med. Chem. Lett.*, **2003**, *13*, 4355-4359.
[166] Sawyer, J.S.; Anderson, B.D.; Beight, D.W.; Campbell, R.M.; Jones, M.L.; Herron, D.K.; Lampe, J.W.; McCowan, J.R.; McMillen, W.T.; Mort, N.; Parsons, S.; Smith, E.C.R.; Vieth, M.; Weir, L.C.; Yan, L.; Zhang, F.; Yingling, J.M. Synthesis and Activity of New Aryl- and heteroaryl-substituted pyrazole inhibitors of the transforming growth factor-beta type i receptor kinase domain. *J. Med. Chem.*, **2003**, *46*, 3953-3956.
[167] Katsila, T.; Spyroulias, G.A.; Patrinos, G.P.; Matsoukas, M.-T. Computational approaches in target identification and drug discovery. *Comput. Struct. Biotechnol. J.*, **2016**, *14*, 177-184.
[168] Ekins, S.; Mestres, J.; Testa, B. *In Silico* pharmacology for drug discovery: Methods for virtual ligand screening and profiling. *Br. J. Pharmacol.*, **2007**, *152*, 9-20.
[169] Chauhan, J.S.; Mishra, N.K.; Raghava, G.P.S. Identification of atp binding residues of a protein from its primary sequence. *BMC Bioinformat.*, **2009**, *10*, 434.
[170] Yu, D.-J.; Hu, J.; Huang, Y.; Shen, H.-B.; Qi, Y.; Tang, Z.-M.; Yang, J.-Y. TargetATPsite: A template-free method for atp-binding sites prediction with residue evolution image sparse representation and classifier ensemble. *J. Comput. Chem.*, **2013**, *34*, 974-985.
[171] Chen, K.; Mizianty, M.J.; Kurgan, L. ATPsite: Sequence-based prediction of atp-binding residues. *Proteome Sci.*, **2011**, *9 Suppl 1*, S4.
[172] Chauhan, J.S.; Mishra, N.K.; Raghava, G.P.S. Prediction of GTP interacting residues, dipeptides and tripeptides in a protein from its evolutionary information. *BMC Bioinformat.*, **2010**, *11*, 301.
[173] Ansari, H.R.; Raghava, G.P.S. Identification of NAD interacting residues in proteins. *BMC Bioinformat.*, **2010**, *11*, 160.
[174] Mishra, N.K.; Raghava, G.P.S. Prediction of FAD interacting residues in a protein from its primary sequence using evolutionary information. *BMC Bioinformat.*, **2010**, *11(Suppl 1)*, S48.
[175] Kalinina, O. V; Wichmann, O.; Apic, G.; Russell, R.B. combinations of protein-chemical complex structures reveal new targets for established drugs. *PLoS Comput. Biol.*, **2011**, *7*, e1002043.
[176] Kuhn, M.; Szklarczyk, D.; Franceschini, A.; Campillos, M.; von Mering, C.; Jensen, L.J.; Beyer, A.; Bork, P. STITCH 2: An interaction network database for small molecules and proteins. *Nucleic Acids Res.*, **2010**, *38*, D552-6.
[177] Mosca, R.; Céol, A.; Stein, A.; Olivella, R.; Aloy, P. 3did: A catalog of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.*, **2014**, *42*, D374-9.
[178] López, G.; Valencia, A.; Tress, M.L. Firestar--prediction of functionally important residues using structural templates and alignment reliability. *Nucleic Acids Res.*, **2007**, *35*, W573-7.
[179] Si, J.; Zhang, Z.; Lin, B.; Schroeder, M.; Huang, B. MetaDBSite: A meta approach to improve protein dna-binding sites prediction. *BMC Syst. Biol.*, **2011**, *5 Suppl 1*, S7.
[180] Trabuco, L.G.; Lise, S.; Petsalaki, E.; Russell, R.B. Pepsite: Prediction of peptide-binding sites from protein surfaces. *Nucleic Acids Res.*, **2012**, *40*, W423-7.
[181] Kumar, M.; Gromiha, M.M.; Raghava, G.P.S. Prediction of RNA binding sites in a protein using svm and pssm profile. *Proteins*, **2008**, *71*, 189-194.

[182] Panwar, B.; Gupta, S.; Raghava, G.P.S. Prediction of vitamin interacting residues in a vitamin binding protein using evolutionary information. *BMC Bioinformat.*, **2013**, *14*, 44.

[183] Ghersi, D.; Sanchez, R. EasyMIFS and SiteHound: A toolkit for the identification of ligand-binding sites in protein structures. *Bioinformatics*, **2009**, *25*, 3185-3186.

[184] Bianchi, V.; Mangone, I.; Ferrè, F.; Helmer-Citterich, M.; Ausiello, G. webPDBinder: A server for the identification of ligand binding sites on protein structures. *Nucleic Acids Res.*, **2013**, *41*, W308-13.

[185] Wass, M.N.; Kelley, L.A.; Sternberg, M.J.E. 3DLigandsite: Predicting ligand-binding sites using similar structures. *Nucleic Acids Res.*, **2010**, *38*, W469-73.

[186] Capra, J.A.; Laskowski, R.A.; Thornton, J.M.; Singh, M.; Funkhouser, T.A. Predicting protein ligand binding sites by combining evolutionary sequence conservation and 3d structure. *PLoS Comput. Biol.*, **2009**, *5*, e1000585.

[187] Ngan, C.-H.; Hall, D.R.; Zerbe, B.; Grove, L.E.; Kozakov, D.; Vajda, S. FTSite: High accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics*, **2012**, *28*, 286-287.

[188] Yang, J.; Roy, A.; Zhang, Y. BioLiP: A Semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Res.*, **2013**, *41*, D1096-103.

[189] Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular docking: A powerful approach for structure-based drug discovery. *Curr. Comput. Aided. Drug Des.*, **2011**, *7*, 146-157.

[190] Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R.P.; Foloppe, N. Current progress in structure-based rational drug design marks a new mindset in drug discovery. *Comput. Struct. Biotechnol. J.*, **2013**, *5*, e201302011.

[191] Hospital, A.; Goñi, J.R.; Orozco, M.; Gelpí, J.L. Molecular dynamics simulations: Advances and applications. *Adv. Appl. Bioinform. Chem.*, **2015**, *8*, 37-47.

[192] Dolatkhah, Z.; Javanshir, S.; Sadr, A.S.; Hosseini, J.; Sardari, S. Synthesis, molecular docking, molecular dynamics studies, and biological evaluation of 4h-chromone-1,2,3,4-tetrahydropyrimidine-5-carboxylate derivatives as potential antileukemic agents. *J. Chem. Inf. Model.*, **2017**, *57*, 1246-1257.

[193] Noha, S.M.; Schmidhammer, H.; Spetea, M. Molecular docking, molecular dynamics, and structure-activity relationship explorations of 14-oxygenated n-methylmorphinan-6-ones as potent μ-opioid receptor agonists. *ACS Chem. Neurosci.*, **2017**, *8*, 1327-1337.

[194] Priya, R.; Sumitha, R.; Doss, C.G.P.; Rajasekaran, C.; Babu, S.; Seenivasan, R.; Siva, R. Molecular docking and molecular dynamics to identify a novel human immunodeficiency virus inhibitor from alkaloids of toddalia asiatica. *Pharmacogn. Mag.*, **2015**, *11*, S414-22.

[195] Armen, R.S.; Chen, J.; Brooks, C.L. An evaluation of explicit receptor flexibility in molecular docking using molecular dynamics and torsion angle molecular dynamics. *J. Chem. Theory Comput.*, **2009**, *5*, 2909-2923.

[196] Shin, W.-H.; Seok, C. Galaxydock: Protein-ligand docking with flexible protein side-chains. *J. Chem. Inf. Model.*, **2012**, *52*, 3225-3232.

[197] Sauton, N.; Lagorce, D.; Villoutreix, B.O.; Miteva, M.A. MS-DOCK: Accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinformat.*, **2008**, *9*, 184.

[198] Fan, H.; Schneidman-Duhovny, D.; Irwin, J.J.; Dong, G.; Shoichet, B.K.; Sali, A. Statistical potential for modeling and ranking of protein-ligand interactions. *J. Chem. Inf. Model.*, **2011**, *51*, 3078-3092.

[199] Betzi, S.; Suhre, K.; Chétrit, B.; Guerlesquin, F.; Morelli, X. GFscore: A general nonlinear consensus scoring function for high-throughput docking. *J. Chem. Inf. Model.*, **2006**, *46*, 1704-1712.

[200] Rentzsch, R.; Renard, B.Y. Docking small peptides remains a great challenge: An assessment using autodock vina. *Brief. Bioinform.*, **2015**, *16*, 1045-1056.

[201] Corbeil, C.; Therrien, E.; Moitessier, N. Modeling reality for optimal docking of small molecules to biological targets. *Curr. Comput. Aided-Drug Des.*, **2009**, *5*, 241-263.

[202] Huang, S.-Y.; Zou, X. Efficient molecular docking of NMR structures: Application to HIV-1 protease. *Protein Sci.*, **2007**, *16*, 43-51.

[203] Moroy, G.; Sperandio, O.; Rielland, S.; Khemka, S.; Druart, K.; Goyal, D.; Perahia, D.; Miteva, M.A. Sampling of conformational ensemble for virtual screening using molecular dynamics simulations and normal mode analysis. *Future Med. Chem.*, **2015**, *7*, 2317-2331.

[204] May, A.; Zacharias, M. Protein-ligand docking accounting for receptor side chain and global flexibility in normal modes: Evaluation on kinase inhibitor cross docking. *J. Med. Chem.*, **2008**, *51*, 3499-3506.

[205] Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **2010**, *31*, 455-461.

[206] Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, **2009**, *30*, 2785-2791.

[207] Lee, H.S.; Zhang, Y. BSP-SLIM: A blind low-resolution ligand-protein docking approach using predicted protein structures. *Proteins*, **2012**, *80*, 93-110.

[208] Sakakibara, Y.; Hachiya, T.; Uchida, M.; Nagamine, N.; Sugawara, Y.; Yokota, M.; Nakamura, M.; Popendorf, K.; Komori, T.; Sato, K. COPICAT: A software system for predicting interactions between proteins and chemical compounds. *Bioinformatics*, **2012**, *28*, 745-746.

[209] Kolb, P.; Caflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.*, **2006**, *49*, 7384-7392.

[210] Taylor, J.S.; Burnett, R.M. DARWIN: A program for docking flexible molecules. *Proteins*, **2000**, *41*, 173-191.

[211] Irwin, J.J.; Shoichet, B.K.; Mysinger, M.M.; Huang, N.; Colizzi, F.; Wassam, P.; Cao, Y. Automated docking screens: A feasibility study. *J. Med. Chem.*, **2009**, *52*, 5712-5720.

[212] Bikadi, Z.; Hazai, E. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of autodock. *J. Cheminform.*, **2009**, *1*, 15.

[213] Bullock, C.W.; Jacob, R.B.; McDougal, O.M.; Hampikian, G.; Andersen, T. Dockomatic - automated ligand creation and docking. *BMC Res. Notes*, **2010**, *3*, 289.

[214] Hart, T.N.; Read, R.J. A multiple-start monte carlo docking method. *Proteins*, **1992**, *13*, 206-222.

[215] Zhou, H.; Skolnick, J. FINDSITE(X): A structure-based, small molecule virtual screening approach with application to all identified human GPCRs. *Mol. Pharm.*, **2012**, *9*, 1775-1784.

[216] Corbeil, C.R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. development and validation of fitted 1.0. *J. Chem. Inf. Model.*, **2007**, *47*, 435-449.

[217] Nabuurs, S.B.; Wagener, M.; de Vlieg, J. A flexible approach to induced fit docking. *J. Med. Chem.*, **2007**, *50*, 6507-6518.

[218] Zhao, Y.; Sanner, M.F. FLIPDock: Docking flexible ligands into flexible receptors. *Proteins*, **2007**, *68*, 726-737.

[219] Shin, W.-H.; Seok, C. GalaxyDock: Protein-ligand docking with flexible protein side-chains. *J. Chem. Inf. Model.*, **2012**, *52*, 3225-3232.

[220] Yang, J.-M.; Chen, C.-C. GEMDOCK: A generic evolutionary method for molecular docking. *Proteins*, **2004**, *55*, 288-304.

[221] Betzi, S.; Suhre, K.; Chétrit, B.; Guerlesquin, F.; Morelli, X. GFscore: A general nonlinear consensus scoring function for High-Throughput Docking. *J. Chem. Inf. Model.*, **2006**, *46*, 1704-1712.

[222] Tietze, S.; Apostolakis, J. Glamdock: Development and validation of a new docking tool on several thousand protein-ligand complexes. *J. Chem. Inf. Model.*, **2007**, *47*, 1657-1672.

[223] Launay, G.; Téletchéa, S.; Wade, F.; Pajot-Augy, E.; Gibrat, J.-F.; Sanz, G. Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Eng. Des. Sel.*, **2012**, *25*, 377-386.

[224] Dominguez, C.; Boelens, R.; Bonvin, A.M.J.J. HADDOCK: A Protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **2003**, *125*, 1731-1737.

[225] McGaughey, G.B.; Sheridan, R.P.; Bayly, C.I.; Culberson, J.C.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.-F.; Cornell, W.D. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.*, **2007**, *47*, 1504-1519.

[226] Tsai, T.-Y.; Chang, K.-W.; Chen, C.Y.-C. iScreen: World's first cloud-computing web server for virtual screening and de novo drug design based on tcm database@taiwan. *J. Comput. Aided. Mol. Des.*, **2011**, *25*, 525-531.

[227] Sauton, N.; Lagorce, D.; Villoutreix, B.O.; Miteva, M.A. MS-DOCK: Accurate multiple conformation generator and rigid

docking protocol for multi-step virtual ligand screening. *BMC Bioinformatics*, **2008**, *9*, 184.

[228] Meier, R.; Pippel, M.; Brandt, F.; Sippl, W.; Baldauf, C. ParaDockS: A framework for molecular docking with population-based metaheuristics. *J. Chem. Inf. Model.*, **2010**, *50*, 879-889.

[229] Gupta, A.; Gandhimathi, A.; Sharma, P.; Jayaram, B. ParDOCK: An all atom energy based monte carlo docking protocol for protein-ligand complexes. *Protein Pept. Lett.*, **2007**, *14*, 632-646.

[230] Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. Patchdock and symmdock: Servers for rigid and symmetric docking. *Nucleic Acids Res.*, **2005**, *33*, W363-7.

[231] Korb, O.; Stützle, T.; Exner, T.E. Empirical scoring functions for advanced protein-ligand docking with plants. *J. Chem. Inf. Model.*, **2009**, *49*, 84-96.

[232] Pyrkov, T. V; Chugunov, A.O.; Krylov, N.A.; Nolde, D.E.; Efremov, R.G. PLATINUM: A Web tool for analysis of hydrophobic/hydrophilic organization of biomolecular complexes. *Bioinformatics*, **2009**, *25*, 1201-1202.

[233] Fan, H.; Schneidman-Duhovny, D.; Irwin, J.J.; Dong, G.; Shoichet, B.K.; Sali, A. statistical potential for modeling and ranking of protein-ligand interactions. *J. Chem. Inf. Model.*, **2011**, *51*, 3078-3092.

[234] Kelley, B.P.; Brown, S.P.; Warren, G.L.; Muchmore, S.W. POSIT: Flexible shape-guided docking for pose prediction. *J. Chem. Inf. Model.*, **2015**, *55*, 1771-1780.

[235] Chen, Y.-F.; Hsu, K.-C.; Lin, S.-R.; Wang, W.-C.; Huang, Y.-C.; Yang, J.-M. SiMMap: A web server for inferring site-moiety map to recognize interaction preferences between protein pockets and compound Moieties. *Nucleic Acids Res.*, **2010**, *38*, W424-30.

[236] Spitzer, R.; Jain, A.N. Surflex-Dock: Docking benchmarks and real-world application. *J. Comput. Aided. Mol. Des.*, **2012**, *26*, 687-699.

[237] Grosdidier, A.; Zoete, V.; Michielin, O. Swissdock, a protein-small molecule docking web service based on eadock DSS. *Nucleic Acids Res.*, **2011**, *39*, W270-7.

[238] Ellingson, S.R.; Smith, J.C.; Baudry, J. VinaMPI: Facilitating multiple receptor high-throughput virtual docking on high-performance computers. *J. Comput. Chem.*, **2013**, *34*, 2212-2221.

[239] Singh, H.; Kumar, R.; Singh, S.; Chaudhary, K.; Gautam, A.; Raghava, G.P.S. Prediction of anticancer molecules using hybrid model developed on molecules screened against nci-60 cancer cell lines. *BMC Cancer*, **2016**, *16*, 77.

[240] Kumar, R.; Chauhan, J.S.; Raghava, G.P.S. In silico designing and screening of antagonists against cancer drug target XIAP. *Curr. Cancer Drug Targets*, **2015**, *15*, 836-846.

[241] Yadav, I.S.; Singh, H.; Khan, M.I.; Chaudhury, A.; Raghava, G.P.S.; Agarwal, S.M. EGFRIndb: Epidermal growth factor receptor inhibitor database. *Anticancer. Agents Med. Chem.*, **2014**, *14*, 928-935.

[242] Mishra, N.K.; Agarwal, S.; Raghava, G.P. Prediction of Cytochrome P450 Isoform Responsible for Metabolizing a Drug Molecule. *BMC Pharmacol.*, **2010**, *10*, 8.

[243] Ortuso, F.; Langer, T.; Alcaro, S. GBPM: GRID-based pharmacophore model: Concept and application studies to protein-protein recognition. *Bioinformatics*, **2006**, *22*, 1449-1455.

[244] Chen, J.; Lai, L. Pocket v.2: Further developments on receptor-based pharmacophore modeling. *J. Chem. Inf. Model.*, **2006**, *46*, 2684-2691.

[245] Garg, A.; Tewari, R.; Raghava, G.P.S. KiDoQ: Using docking based energy scores to develop ligand based model for predicting antibacterials. *BMC Bioinformat.*, **2010**, *11*, 125.

[246] Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R.; Consonni, V.; Kuz'min, V.E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR modeling: Where have you been? where are you going to? *J. Med. Chem.*, **2014**, *57*, 4977-5010.

[247] Winkler, D.A. The role of quantitative structure--activity relationships (qsar) in biomolecular discovery. *Brief. Bioinform.*, **2002**, *3*, 73-86.

[248] Gao, Q.; Yang, L.; Zhu, Y. Pharmacophore based drug design approach as a practical process in drug discovery. *Curr. Comput. Aided. Drug Des.*, **2010**, *6*, 37-49.

[249] Yang, S.-Y. Pharmacophore modeling and applications in drug discovery: Challenges and recent advances. *Drug Discov. Today*, **2010**, *15*, 444-450.

[250] Tschinke, V.; Cohen, N.C. The NEWLEAD program: A new method for the design of candidate structures from pharmacophoric hypotheses. *J. Med. Chem.*, **1993**, *36*, 3863-3870.

[251] Zhao, H. Scaffold selection and scaffold hopping in lead generation: A medicinal chemistry perspective. *Drug Discov. Today*, **2007**, *12*, 149-155.

[252] Barillari, C.; Marcou, G.; Rognan, D. Hot-Spots-guided receptor-based pharmacophores (hs-pharm): A knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J. Chem. Inf. Model.*, **2008**, *48*, 1396-1410.

[253] Wolber, G.; Langer, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J. Chem. Inf. Model.*, **2005**, *45*, 160-169.

[254] Horvath, D. Pharmacophore-based virtual screening. *Methods Mol. Biol.*, **2011**, *672*, 261-298.

[255] Tibbitts, J.; Canter, D.; Graff, R.; Smith, A.; Khawli, L.A. Key Factors Influencing ADME properties of therapeutic proteins: A need for adme characterization in drug discovery and development. *MAbs*, **2016**, *8*, 229-245.

[256] Koes, D.R.; Camacho, C.J. Pharmer: Efficient and exact pharmacophore search. *J. Chem. Inf. Model.*, **2011**, *51*, 1307-1314.

[257] Dror, O.; Schneidman-Duhovny, D.; Inbar, Y.; Nussinov, R.; Wolfson, H.J. Novel approach for efficient pharmacophore-based virtual screening: Method and applications. *J. Chem. Inf. Model.*, **2009**, *49*, 2333-2343.

[258] Wolber, G.; Dornhofer, A.A.; Langer, T. Efficient overlay of small organic molecules using 3d pharmacophores. *J. Comput. Aided. Mol. Des.*, **2006**, *20*, 773-788.

[259] Oloff, S.; Zhang, S.; Sukumar, N.; Breneman, C.; Tropsha, A. Chemometric analysis of ligand receptor complementarity: Identifying complementary ligands based on receptor information (CoLiBRI). *J. Chem. Inf. Model.*, **2006**, *46*, 844-851.

[260] Cereto-Massagué, A.; Guasch, L.; Valls, C.; Mulero, M.; Pujadas, G.; Garcia-Vallvé, S. DecoyFinder: An easy-to-use python gui application for building target-specific decoy sets. *Bioinformatics*, **2012**, *28*, 1661-1662.

[261] Abreu, R.M.; Froufe, H.J.; Queiroz, M.J.R.; Ferreira, I.C. MOLA: A bootable, self-configuring system for virtual screening using autodock4/vina on computer clusters. *J. Cheminform.*, **2010**, *2*, 10.

[262] Durrant, J.D.; McCammon, J.A. NNScore: A neural-network-based scoring function for the characterization of protein-ligand complexes. *J. Chem. Inf. Model.*, **2010**, *50*, 1865-1871.

[263] Jacob, R.B.; Bullock, C.W.; Andersen, T.; McDougal, O.M. DockoMatic: Automated peptide analog creation for high throughput virtual screening. *J. Comput. Chem.*, **2011**, *32*, 2936-2941.

[264] O'Connell, J.; Gurdasani, D.; Delaneau, O.; Pirastu, N.; Ulivi, S.; Cocca, M.; Traglia, M.; Huang, J.; Huffman, J.E.; Rudan, I.; McQuillan, R.; Fraser, R.M.; Campbell, H.; Polasek, O.; Asiki, G.; Ekoru, K.; Hayward, C.; Wright, A.F.; Vitart, V.; Navarro, P.; Zagury, J.-F.; Wilson, J.F.; Toniolo, D.; Gasparini, P.; Soranzo, N.; Sandhu, M.S.; Marchini, J. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.*, **2014**, *10*, e1004234.

[265] Marialke, J.; Körner, R.; Tietze, S.; Apostolakis, J. Graph-based molecular alignment (GMA). *J. Chem. Inf. Model.*, **2007**, *47*, 591-601.

[266] Taminau, J.; Thijs, G.; De Winter, H. Pharao: Pharmacophore alignment and optimization. *J. Mol. Graph. Model.*, **2008**, *27*, 161-169.

[267] Durrant, J.D.; McCammon, J.A. Autoclickchem: Click chemistry in silico. *PLoS Comput. Biol.*, **2012**, *8*, e1002397.

[268] Reutlinger, M.; Koch, C.P.; Reker, D.; Todoroff, N.; Schneider, P.; Rodrigues, T.; Schneider, G. Chemically advanced template search (cats) for scaffold-hopping and prospective target prediction for "orphan" molecules. *Mol. Inform.*, **2013**, *32*, 133-138.

[269] Lešnik, S.; Štular, T.; Brus, B.; Knez, D.; Gobec, S.; Janežič, D.; Konc, J. LiSiCA: A software for ligand-based virtual screening and its application for the discovery of butyrylcholinesterase inhibitors. *J. Chem. Inf. Model.*, **2015**, *55*, 1521-1528.

[270] Meraj, K.; Mahto, M.K.; Christina, N.B.; Desai, N.; Shahbazi, S.; Bhaskar, M. Molecular modeling, docking and admet studies towards development of novel disopyramide analogs for potential inhibition of human voltage gated sodium channel proteins. *Bioinformation*, **2012**, *8*, 1139-1146.

[271] Berenger, F.; Voet, A.; Lee, X.Y.; Zhang, K.Y. A rotation-translation invariant molecular descriptor of partial charges and its use in ligand-based virtual screening. *J. Cheminform.*, **2014**, *6*, 23.

[272] Saeedipour, S.; Tai, D.; Fang, J. ChemCom: A software program for searching and comparing chemical libraries. *J. Chem. Inf. Model.*, **2015**, *55*, 1292-1296.

[273] Pottel, J.; Moitessier, N. Customizable generation of synthetically accessible, local chemical subspaces. *J. Chem. Inf. Model.*, **2017**, *57*, 454-467.

[274] Puranen, J.S.; Vainio, M.J.; Johnson, M.S. Accurate conformation-dependent molecular electrostatic potentials for high-throughput in silico drug discovery. *J. Comput. Chem.*, **2010**, *31*, 1722-1732.

[275] Shelley, J.C.; Cholleti, A.; Frye, L.L.; Greenwood, J.R.; Timlin, M.R.; Uchimaya, M. Epik: A software program for pk(a) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided. Mol. Des.*, **2007**, *21*, 681-691.

[276] Zoete, V.; Daina, A.; Bovigny, C.; Michielin, O. SwissSimilarity: A web tool for low to ultra high throughput ligand-based virtual screening. *J. Chem. Inf. Model.*, **2016**, *56*, 1399-1404.

[277] Irwin, J.J.; Duan, D.; Torosyan, H.; Doak, A.K.; Ziebart, K.T.; Sterling, T.; Tumanian, G.; Shoichet, B.K. An aggregation advisor for ligand discovery. *J. Med. Chem.*, **2015**, *58*, 7076-7087.

[278] Koes, D.R.; Camacho, C.J. ZINCPharmer: Pharmacophore search of the zinc database. *Nucleic Acids Res.*, **2012**, *40*, W409-14.

[279] Floris, M.; Masciocchi, J.; Fanton, M.; Moro, S. Swimming into peptidomimetic chemical space using pepMMsMIMIC. *Nucleic Acids Res.*, **2011**, *39*, W261-9.

[280] Vainio, M.J.; Puranen, J.S.; Johnson, M.S. Shaep: Molecular overlay based on shape and electrostatic potential. *J. Chem. Inf. Model.*, **2009**, *49*, 492-502.

[281] Fiser, A. Template-based protein structure modeling. *Methods Mol. Biol.*, **2010**, *673*, 73-94.