

Structure-based drug designing

Shubham Pant¹, Shivani Verma², Rajesh Kumar Pathak³ and Dev Bukhsh Singh⁴

¹*Electrochemical Process Engineering Division, CSIR-Central Electrochemical Research Institute, Karaikudi, India,* ²*Department of Chemistry, College of Basic Sciences & Humanities, G. B. Pant University of Agriculture and Technology, Pantnagar, India,* ³*School of Agricultural Biotechnology, Punjab Agricultural University, Ludhiana, India,* ⁴*Department of Biotechnology, Institute of Biosciences and Biotechnology, Chhatrapati Shahu Ji Maharaj University, Kanpur, India*

14.1 Introduction

In the current scenario, our life is too fast-paced, which comes with major concerns about public health. The public health of every age group has been a never ending story. However, it is evident that utilizing rigorous ongoing developments in the pharmaceutical field, we have successfully found a cure for several diseases. These modern-day medicines can present themselves to be more effective either by direct treating, and preventing the diseases or by generating immunity to prevent the reoccurrence of the disease. For instance, patients with many deadly diseases, such as cancer, which had been a low survival rate in the 1970s, now can able to get treatment more effectively by modern chemotherapy techniques. The death rate for the same has been declined by 20% as compared to the last decade (Siegel, Naishadham, & Jemal, 2013). However, along with this accelerated development of these pharmaceutical medicines, there is a need to be more careful about their long-term desirable/undesirable side effects that may vary according to severity, which may include multidrug resistance, poor absorption and retention, toxicities, and some adverse allergic reactions. The other side of the development of these drugs is manufacturing, which takes a lot of money and time. According to some estimation, it takes up to \$2.6 billion for development and after bringing it to the market raised almost 150% of its cost in the past decades (Tollman, 2001). Furthermore, 90% of drug failure occurs due to lack of efficacy and severe toxicity effects in clinical trials, and clinical trials require a large amount of funding for its completion (Joseph, DiMasi, Henry, & Grabowski, 2012; Khanna, 2012). In recent years due to the emergence of new viruses and other disease causing pathogens, the time needed to develop drugs that come to market is around 10–15 years or as soon as possible.

Advanced techniques, such as combinatorial chemistry or high-throughput screening (HTS), are capable of assay and screen thousands of molecules in a single go as they are having 96–1536-well formats (Li et al., 2017). However, it is seen that these techniques come with limitations, such as time consuming and expensive while taking the biological complexity scenario into mind. Therefore there is an urgent need to view closer into these drugs in terms to get reduced manufacturing time as well as expenditure, for that the researchers and pharmaceutical companies may have attention to computer-aided drug discovery. Generally, these techniques can be comprised different stages of drug discovery and development. These techniques might be helpful to screen hundreds/thousands of compounds quickly and identify putative efficient candidate molecules against the desired molecular targets (i.e., proteins, nucleic acid, lipid, and carbohydrates) and visualize their molecular interactions with the help of molecular modeling software's to determine target–ligand interaction (Sotriffer & Klebe, 2002; Zhao, Xia, Yu, Pan, & Wang, 2015). Therefore, optimization of drug candidates is required for improved potency, lesser side effects, and enhanced selectivity with the target. The structure-based drug design (SBDD) is a very useful and robust technique in the discovery of lead molecules through computational approaches.

Over the decades, technologies, such as X-ray crystallography and nuclear magnetic resonance spectroscopy (NMR), have contributed to identifying the three-dimensional (3D) structures of many biological molecules, which establishes the era of drug discovery by SBDD (Middleton, 2007; Hosfield et al., 2003; Kawato et al., 2015). This approach is followed by designing suitable drug molecules according to the amino acid residues present in binding site

cavities of the target structure for better binding affinity and selectivity. SBDD also predicts the key binding sites of the target and the affinity of probable ligands to the target that is important in terms of respective biological functions (Mandal, Moudgi, & Mandal, 2009). In recent times this technique helped researchers to find many drugs like compounds and some of them are finally brought to the market. Besides this, there are some disadvantages of SBDD, such as the difficulty of considering target flexibility in docking, high false positive rate of virtual screening, and the inaccuracy of algorithms and scoring functions used by tools for predicting target–ligand-binding-free energy (Pei, Yin, Ma, & Lai, 2014; Saleh & Elshemey, 2017; Warren, Do, Kelley, Nicholls, & Warren, 2012).

14.2 Background of structure-based drug design

SBDD has been seen as the most powerful and efficient process in the entire drug discovery program. Databases containing information about the target and small molecules, genetic information with their sequences, 3D structure, interaction, and binding information, cytotoxicity data, absorption, distribution, metabolism, excretion, and toxicity (ADMET) data, and other important biological information serves as the most efficient sources for accelerating the drug discovery process (Batool and Choi, 2017). SBDD works as an iterative process and it proceeds with multiple cycles of screening and simulation that leads to a highly optimized drug candidate for the subsequent developmental clinical trials. In general, any drug discovery process mainly comprises four steps that include the discovery phase, development phase, clinical trial phase, and registry phase. First, in the discovery phase, a metabolically involved therapeutic target and active ligands and their interactions among them are identified.

Many computer algorithm-based tools are available that are utilized for the docking of the ligand database that may also include their modified forms or fragments into the binding pocket of the targeted structure. The operational tool ranks these ligand molecules with respect to the scoring function used that is generally based on steric, binding energies, electrostatic interactions, and hydrogen bond formation of the drug like a molecule with the target-binding cavity/site. Therefore a thorough investigation is done for examining electrostatic behavior of the binding site, and also identifying the cavities, clefts, and allosteric pockets presence that can be visualized by using molecular visualization tools, that is, PyMol. The information of the 3D structure of protein identified by techniques, such as X-ray diffraction and NMR, can be imported from online database, that is, RCSB Protein Data Bank (PDB). With the availability of 3D structure, it is possible to derive detailed information about the intermolecular behavior involved in the process of molecular bond rearrangements and binding of the ligand.

Insights into structure for protein/ligand–protein complex majorly help to examine various binding structural conformations, unknown binding site prediction, important functional groups, and conformational changes resulting from ligand–protein interactions and structural stability studies. A recent trend in the SBDD method focuses on the main examination of the key features of binding pockets/sites of the therapeutic target for the design of efficient ligands (Blaney, 2012; Mamgain, Sharma, Pathak, & Baunthiyal, 2015; Mandal et al., 2009). Furthermore, in the next phase, the top complexes are developed and optimized followed by further checked of stability (Wilson & Lill, 2011). Afterward, the top-shortlisted compounds having high selective target affinity can be examined for *in vitro* biochemical activities. These protein–ligand complex needs to be tested for involvement with some key metabolic/cellular pathways, proceeding to the further development process of drugs. After this step, the drug can be considered as having desired therapeutic functionality (Urwiler, 2001). Qualified drugs analyzed further for crucial biological properties, such as efficacy, affinity, and potency, evaluated by corresponding experimental methods (Ren et al., 2011). Furthermore, multiple cyclic repetitions of the mentioned step increase the specificity and efficacy of the lead compound. The third phase goes the same as a conventional drug trial that includes clinical trials with different phases characterized by trials on the population of various age groups in advanced trials. Candidate compounds that successfully pass the previous clinical trials proceeds for the concluding phase that comprises candidate drug registration by the corresponding competent authorities across the globe, and finally, after the final approval, the drug will be distributed in the market for medical/therapeutic uses.

Currently SBDD became as a promising computational technique that is being used by R&D groups of institutions and, pharmaceutical companies. This technique also helped scientists to make numerous drug products for the market use including human immunodeficiency virus (HIV)-1-inhibiting FDA-approved drugs (Wlodawer & Vondrasek, 1998). Moreover, some other examples exist that are identified by the SBDD technique: a thymidylate synthase inhibitor, raltitrexed (Anderson, 2003); amprenavir, a potential inhibitor of HIV protease; and the antibiotic norfloxacin (Rutenber & Stroud, 1996; De Paulis, 2007). As SBDD algorithms have certain limitations and require further research work, or one can take a consensus from different tools available for optimization of rankings (Agnihotry, Pathak, Srivastav, Shukla, & Gautam, 2020; Pathak, Singh, Sagar, Baunthiyal, & Kumar, 2020).

14.3 Process of SBDD

The structure-based drug discovery and designing completely depend on bioinformatic methods and tools before going to the wet lab or experimental validations. The computational pipelines involved in SBDD, that is, target identification, structure prediction, visualization, validation, binding site characterization, docking and virtual screening, visualization of docked complex structure and their stability analysis, ADMET screening, and binding-free energy-MM-PBSA, are discussed in the following sections and highlighted in Fig. 14.1.

14.3.1 Target identification

SBDD's fundamental step is the drug target identification and its validation for the drug discovery process. It starts with the identification of the causative agents of disease that has to be established by biochemical and biological basis. This information can be easily retrieved through literature mining. To look carefully at the whole pathophysiological pathway of disease of interest is foremost needed. The ideal target macromolecule for SBDD can be any macromolecule that is completely linked with a selected disease pathway, involved metabolism, and can bind a small candidate drug molecule to facilitate required functions. Following that, the most crucial causative macromolecule can be chosen for the study. Basically, four causative agents or potential drug targets can be seen in organisms, that is, nucleic acids (DNA and RNA), proteins including receptors and enzymes, carbohydrates, and lipid. Presently available drugs in the market are designed focusing majorly on protein as a target. However, from the past year, nucleic acids are much desired drug targets, which may be due to the decoding of several genomes of organisms as well as pathogens (Gashaw, Ellinghaus, Sommer, & Asadullah, 2012). Also, the RNA targets' use with an established secondary structure, such as bacterial ribosome and HIV genome portion, is also effective. Some small molecules that modulate the functions of proteases, kinases, ion channels, and nuclear hormone receptors make up 22% of the total market (Gallego & Varani, 2001).

As the pathophysiological pathway picture is complex so it is a challenging task for the selection of potential drug targets. It should be selected in a way that the forward reaction pathways, after their interaction with the drug, suppress the upcoming sequential reaction events. It should be essential that its deletion should lead to the death of the causative agent. It also needs to be unique; none of the other pathways should be able to supplement the target one. In this time, the use of genomics and proteomics methods came up as the most reliable for drug target identification (Singh, 2014). Furthermore, major advances in technologies, such as high-throughput omics, generated a lot of data for host/organism–pathogen interaction. The scientific community integrated and analyzed these available data through mathematical modeling, network, and systems biology approaches to accelerate the developmental process of target identification in SBDD.

With the progress and development of the conventional structure identification techniques (such as X-ray and NMR), the data for the protein or protein–ligand complex structures in popular databases, such as the RCSB PDB, are increasing (Chaudhari, Heim, & Li, 2015; Franca, 2015). Due to the limitations of these conventional techniques, the determination of many target structures is still difficult by such experimental methods. However, new methods are gradually being discovered to overcome these limitations. Computational methods, such as homology modeling or comparative modeling, are useful to determine the structure of drug target using the amino acid sequence of the target protein and its structural homolog, that is, similar experimentally determined structure present in the PDB database. To determine structure computationally it also includes threading or fold recognition methods when no structural homolog is found in PDB (Nim, Sun, & Wong, 2017).

14.3.2 Target structure determination/prediction

Once a target has been identified it is important to obtain accurate information regarding structure. The structural information of the target is available in structural database, that is, RCSB PDB, but due to unavailability of experimental structure determination, bioinformatics method plays a vital role in the structure determination and accelerates the drug discovery program. The target structure contains binding pockets/sites. Upon binding of the ligand molecule at the binding site of the target structure should reflect the dynamical changes. Several software/tools are available for protein structure modeling using homology modeling, fold recognition, and ab initio method. Homology modeling is recognized as the most accurate computation method, it uses sequence information of target protein for the modeling of 3D structure. This method predicts that two or more proteins can have an evolutionary connection, and these different proteins can have a good extent of structural similarity. So, the target protein structure can be modeled using the known

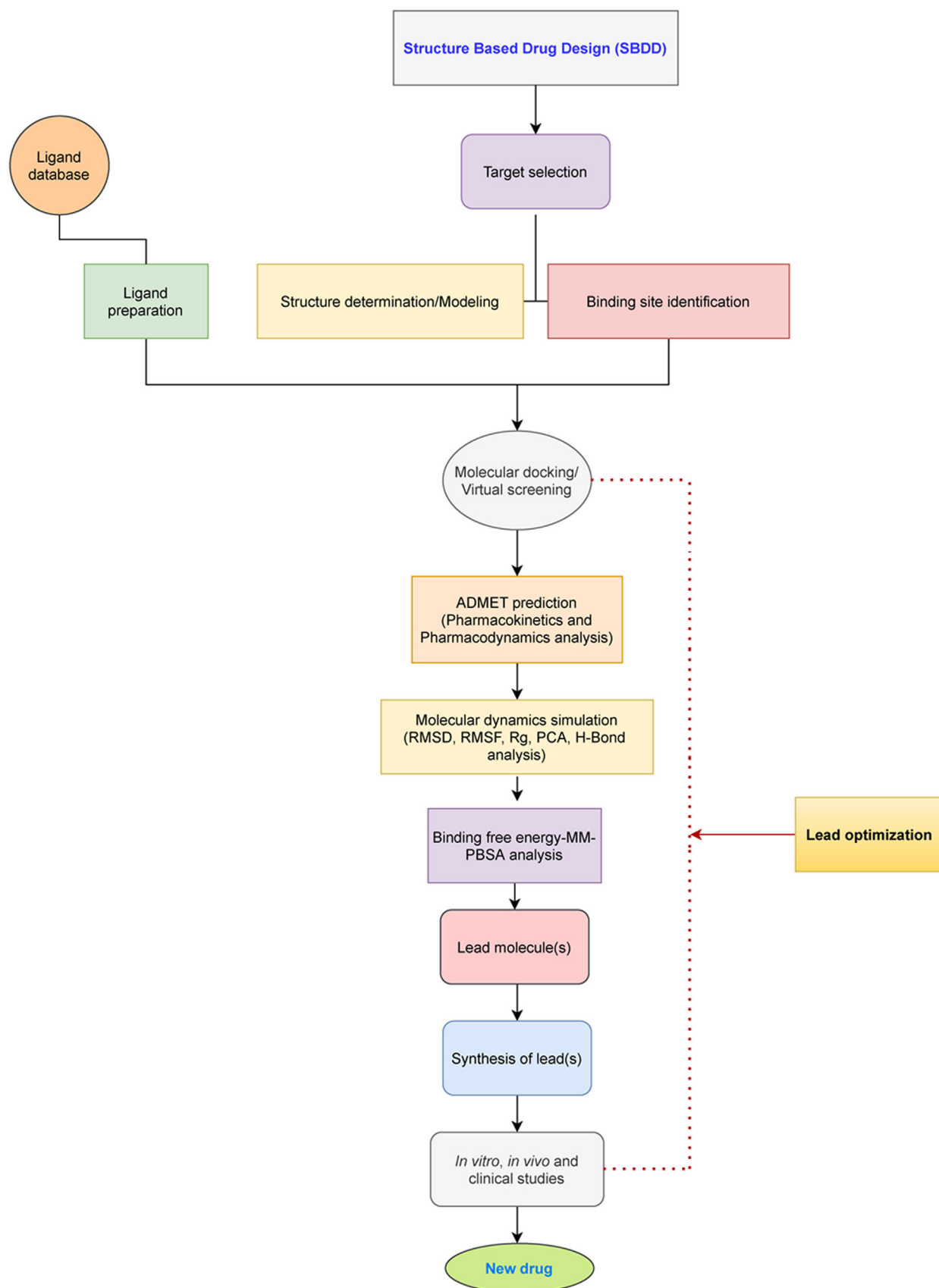


FIGURE 14.1 Computational pipeline used in structure-based drug design for the identification of lead molecule(s).

template with another 3D structure. However, this method has limitations such as it can be used only for more than 35% sequence identity between target and template. This technique may not give reliable results whenever the identity is less than 30% (Franca, 2015).

After modeling the target 3D structure, it is necessary to examine the quality of the predicted structure by checking its stereochemical properties using the Ramachandran plot. After that, we can have an idea about the probable conformation of σ and ψ angles for each amino acid. Second, the fold recognition or threading method is used for constructing a reliable model when no suitable structural homologs are detected. However, this approach is seen to be more precise about fine structural information for predicting the best suitable structure for protein. Protein structural analysis has shown that different amino acid residues in a protein have their unique or very different behavior to remain in an appropriate environmental class. Generally, there are 6 basic environmental classes, and approximately 18 environmental classes or subclass determined for every amino acid residues present in three secondary structures, that is, sheet, helix, and coil exists. The ab initio method can be used when neither structural counterparts nor structural fold is available for the desired target protein. The configuration spaces of atoms in amino acids of protein were determined by using various principles of chemistry, physics, and mathematics. This approach has a search method to explore the energy-based scoring function and entire conformational space to produce the most suitable and accurate structure. Taking different conformational spaces can be used for the different torsion angle's discrete values of amino acids. Conformation energy can be calculated using suitable parameters, such as a number of bonds, bond angles, bond length, interatomic distances, and disulfide linkage. Finally, the structure having global minimum energy conformation can be considered as a stable and reliable model (Banner et al., 1993; Kaur, Utreja, Dhillon, Pathak, & Singh, 2020).

14.3.3 Cavity/binding site prediction

The structure of the target protein determination is followed by the identification of the site of the binding pocket on the 3D protein model. The identification of the functional sites on the target protein model is an essential criterion for the perfect coordination of ligand for inhibiting or activating its function. However, only a few reliable methods exist for spotting the potential binding residues sites. Generally, these methods utilize the knowledge about van der Waals (vdW) forces, hydrogen bonding interaction, and binding energy calculations for mapping a suitable site. Nowadays, many methods are developed in SBDD for determining binding pockets by binding energy calculations, as this identifies the particular sites on the protein model that can compose interaction between the participating functional groups and drug molecules (Laurie & Jackson, 2005). Also, the energetically favorable interactions of specific candidates with the proteins can also be seen by these methods.

Commonly, for the prediction of the binding site, one method that is energy based is Q-SiteFinder (Laurie & Jackson, 2005). vdW binding energies of proteins with a specific methyl probe are calculated by this method. According to those molecules having favorable energy can be clustered and retained. However, these clusters of probes need to be ranked/sorted based on their binding energy calculation. In addition, sometimes the participating residues of proteins are functionally explained for the binding site area determination. The computational approaches used for the target structure determination to the binding site prediction are depicted in Fig. 14.2.

14.3.4 Ligand structure preparation/retrieval

The technique of SBDD emerged as a matter of attention in recent past years. This method implies supplementing the conventional methods as a new era in the identification of probable drug compounds in the developmental process of SBDD. Following the process follows in this approach, the compounds are first identified based on data mining and then can be optimized for various parameters related to ADMET and then linked together to produce high-affinity lead compounds. This optimization can be done based on the structural proximities, which obtained from X-ray crystallography, NMR, and various other analytical techniques. These candidates should be functionally simpler than available drug molecules. They should have efficient binding with the drug target, and should also follow the Lipinski rule of five, which is considered crucial for orally administrated drugs.

Different bioinformatics resources, that is, databases and software, have been developed for the retrieval and preparation of ligand structures. ZINC database is recognized as one of the key resources for the retrieval of a big number of ligand structures in different file formats for virtual screening and molecular docking. Besides, PubChem and ChemSpider are considered as important database resources for the retrieval of ligand structure along with their properties. ChemDraw, MarvinSketch, ChemSketch, etc., are software used for preparation and drawing ligand structure in

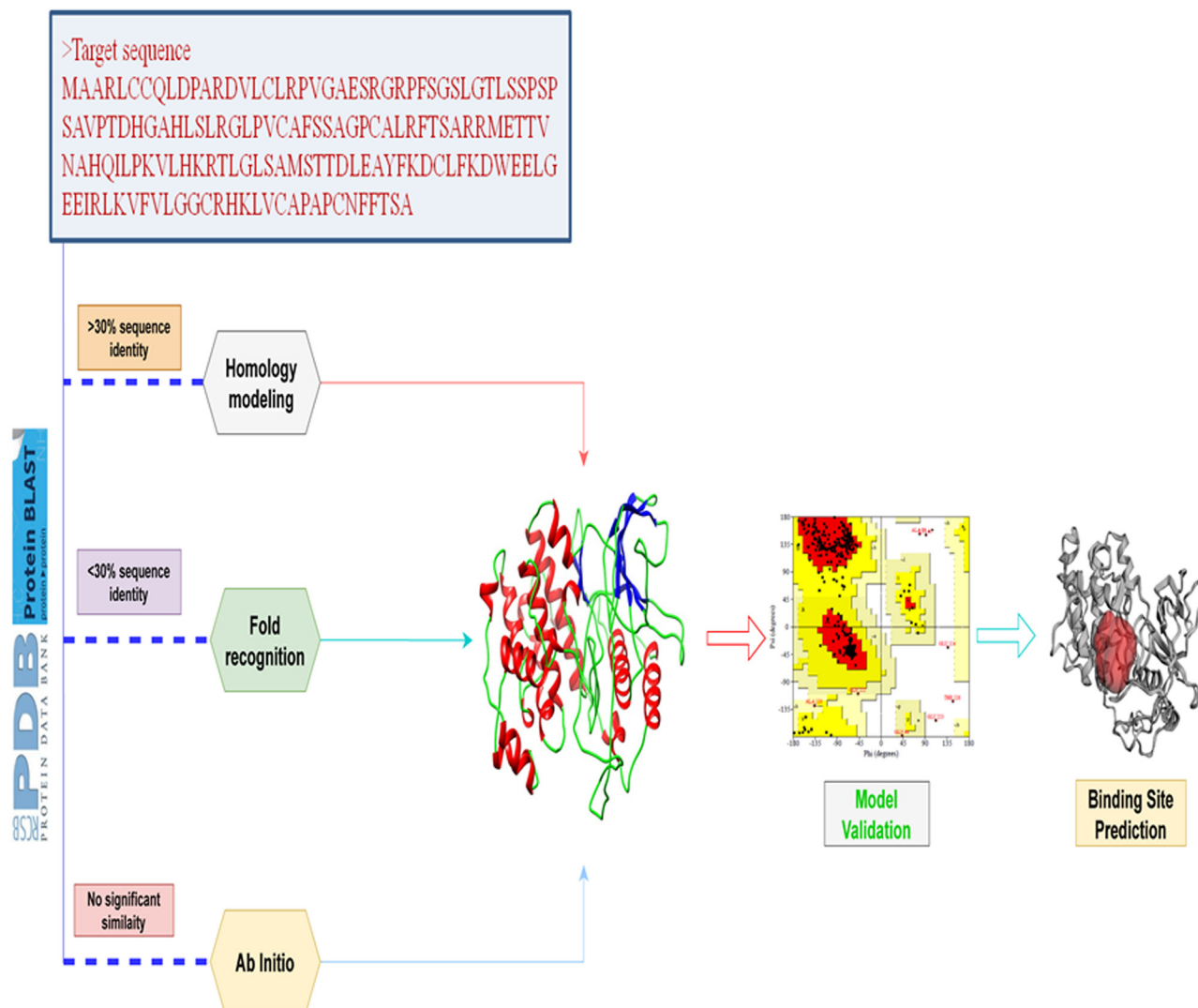


FIGURE 14.2 Computational approaches used in target structure determination, their validation, and binding site prediction using sequence information of selected molecular target for SBDD.

different file formats; besides, it can be utilized for calculation of different physiochemical properties of ligand molecules, that is, molecular weight, h-bond donor and acceptor, and logP.

14.3.5 Molecular docking and virtual screening

Molecular docking is the phenomenon of fitting the ligand into the binding site cavity of the target in 3D space. It is considered a technique of computational simulation for molecular interactions. Basically, the ligand is made to bind on the active site of the target, and then, based on scoring function, the binding energy is analyzed and their interactions are visualized (Agrwal, Pathak, & Kasana, 2021; Sagar et al., 2014). It generally predicts the complex structural conformation as well as the binding affinity of different compounds targeting an active site of the target macromolecular structure accurately. Therefore it has become the most popular and reliable technique in SBDD. Moreover, this technique can also be implemented to examine key molecular phenomena. For instance, intramolecular/intermolecular interactions, a number of hydrogen bond formations during interactions, help researchers for getting an idea about the stability of complex conformation.

Several kinds of docking algorithms can also provide binding energies calculations and ranks the different ligands by the usage of different scoring functions. Therefore the thermodynamic stability of the molecular system can be seen

by theoretically calculating intramolecular/intermolecular forces for stabilizing protein or protein–ligand complex (Huang & Zou, 2010). Generally, most of the suitable ligand-binding conformation mostly depends on two factors: (1) enough structural space defining probable binding conformations and (2) explicit prediction of binding energy with each conformation (Meng, Zhang, Mezei, & Cui, 2011). A significant number of iterations can be performed to attain the minimized energy level followed by assessing ligand-binding by different scoring functions. There are mainly two kinds of docking: (1) flexible docking and (2) rigid docking, which plays a key role in the screening and identification of lead compounds. Various docking simulation methods are based on a kind of key algorithms to handle ligand flexibility. Moreover, the systematic algorithms aimed at analyzing the degree of freedom. This task can be achieved by using the fragmentation method; using this method, a ligand can gradually move with flexibility inward in a suitable binding cavity (Kitchen, Decornez, Furr, & Bajorath, 2004). The second technique combines 360-degree rotatable bonds of the molecule made to rotate systematically and apply a prefixed constant increment rate; either pre-determined libraries of structural conformational ensembles are accessed for the betterment of ligand flexibility using database approach, or some of the random modifications can be implemented for an independent ligand or any single group. However, these modifications might be accepted or rejected, which depends upon probability calculations of genetic algorithm methods (Huang & Zou, 2010; Lopez-Vallejo et al., 2011) as well as the Monte Carlo method.

Virtual screening is a powerful computational method that is based on molecular docking algorithms, in which we screen/dock a large set of molecules on a particular target simultaneously to identify best fitted molecules in terms of binding energy and interacting partners (Pathak, Baunthiyal, Taj, & Kumar, 2014). It will be sorted based on binding-free energy for further research. This technique is very useful in accelerating drug discovery programs through the identification of lead molecules in very less time. A list of some tools and database resources used in SBDD is mentioned in Table 14.1 and Table 14.2, respectively.

TABLE 14.1 List of some tools and their applications used for the SBDD approach.

Stages	Tool name	Application	Reference
Target modeling	SWISS-MODEL	Homology modeling	Waterhouse et al. (2018)
	MODELLER	Homology modeling	Webb and Sali (2016)
	Phyre and Phyre2	Remote template detection, alignment and 3D modeling	Kelley, Mezulis, and Yates (2015)
	HHpred	Template detection, alignment, and 3D modeling	Söding, Biegert, & Lupas, 2005
	I-TASSER	Threading ab initio modeling	Roy et al. (2010)
Binding site prediction	CASTp	Binding site prediction	Tian, Chen, Lei, Zhao, and Liang (2018)
	Active Site Prediction tool	Active site prediction	–
Molecular docking tools	AutoDock Vina	Molecular docking and virtual screening	Trott and Olson (2010)
	Sanjeevini	Molecular docking	Jayaram et al. (2012)
ADMET prediction	admetSAR	ADMET prediction	–
	QikProp	ADME analysis	–
Molecular Dynamic Simulation	AMBER	Suits for biomolecular simulation programs	Case et al. (2020)
	GROMACS	General purpose molecular dynamic simulation package	van Gunsteren and Berendsen (1987)
MM-PBSA	g_mmpbsa	Binding energy calculation using MM-PBSA method	Kumari, Kumar, Open Source Drug Discovery Consortium, and Lynn (2014)

TABLE 14.2 List of some important database resources used in SBDD.

Database name	Information	Link	Reference
ZINC	Database containing commercially available compounds for molecular docking and virtual screening	http://zinc.docking.org/	Irwin and Shoichet (2005)
PubChem	Database containing small molecules and biological activity	https://pubchem.ncbi.nlm.nih.gov/	Kim et al. (2016)
ChemSpider	Chemical structure data for drug discovery	http://www.chemspider.com/	Pence and Williams (2010)
DrugBank	FDA approved drugs for repurposing	https://www.drugbank.ca/	Wishart et al. (2006)
ChemSpider	Ligand structure	http://www.chemspider.com/	Pence and Williams (2010)

14.3.6 ADMET analysis

Absorption, Distribution, Metabolism, Excretion, and Toxicity commonly known as ADMET is the key criteria for sorting ligands in drug discovery programs. It serves as a primary testing requirement for any candidate molecule. A number of tools are available for analyzing and predicting ADMET properties: (1) ADMET parameter predictors, (2) metabolic fate predictors, (3) metabolic stability predictors, (4) cytochrome P450 substrate predictors, and (5) physiology-based pharmacokinetic modeling software. These ADMET parameters majorly reflect the candidate compound's performance as well as pharmacological action as a drug based on the in vivo behavior and the body response for the same; therefore it is useful as it providing a primary key view for the nature of the candidate drug-like compound. Therefore this technique majorly plays an important role in identifying most suited drug-like compounds as its primary screening at preliminary phases of the drug discovery process and it significantly lowers the late-stage failure of candidate drugs during preclinical or clinical advance stage phase trials; thus it can be proving to be an economical and time-effective approach.

Drug designing is still coming out as a challenging task because of high input costs for development and a very low success rate. It is widely suggested that any drug discovery program must pass the computer-based ADMET model prediction prior to in vitro and in vivo model studies. At present some predefined principle descriptors, such as polar surface area (2D), polarizability, vdW surface area, refractivity, and Lipinski's rule, are used for these studies (Pathak, Gupta, Shukla, & Baunthiyal, 2018; Singh & Pathak, 2020), but advancement in the area of two pioneer methods, such as combinatorial chemistry and high-throughput screening, has significantly enhanced the preliminary data in terms of ADMET for small molecules, which is available as a reference for further studies. Furthermore, to make ADMET more accurate adding new principle descriptors and improving algorithms implementation can work; these most relevant pharmacokinetics and pharmacodynamics information for molecules can be modeled to increase the efficiency and fasten the SBDD and development process for several diseases in a cost-effective manner.

14.3.7 Molecular dynamics simulation

The macromolecules or their complex file, that is, docked structures, can be better analyzed for some crucial parameters, affecting its stability, conformation, bioactivity, etc. Molecular dynamics simulation (MDS) is a method comprising mathematical, biophysical, and computational studies of biomolecules for the analysis of their behaviors with respect to time. The analysis of such a complex biological picture appears to be extremely challenging because of the involvement of a large number of molecular components in biological systems. In this study, the participating entities, such as atoms and the molecules, are allowed to interact for a defined specific period of time for the simulation study. This is followed by different parameters estimation and calculation for each entity and designating data to check the overall behavioral activity of the macromolecular molecular structure.

There are many advantages seen over molecular docking because in the case of molecular docking it provides only binding-free energy of ligands and number of participating hydrogen bond with the target receptor, but additionally,

using MDS we can know the specific interaction of the ligand with receptors at the atomic and subatomic levels. There are various parameters in MDS, such as RMSD, which is used to estimate the stability of the target or its ligand complex, and it also shows probable data about the conformational forms and helps to stabilize a molecule with minimized global energy. On the other hand, using root mean square fluctuation is taken to determine the conformational alignments in terms of a time period (Rana, Pathak, Shukla, & Baunthiyal, 2019; Singh & Dwivedi, 2016). Besides, some other analysis, such as Rg, H-bonds, PCA, and Gibbs-free energy, will be conducted to visualize the nature of structures (Rai, Pathak, Singh, Bhatt, & Baunthiyal, 2021; Sidhu, Bhangu, Pathak, Yadav, & Chhuneja, 2020).

Very novel information can be retrieved about the receptor or its ligand–receptor complex that information can be further utilized for the next steps in the SBDD program. Mainly, the software used for MDS is Amber, Gromacs, CHARMM, and Desmond. The working principle of the MDS algorithm is generally comprised majorly three steps; (1) each atom's initial positions' identification and its speed; (2) calculation for the forces given to the identified entity using interatomic potential; and (3) further progression/movement of the atomic positions and its speeds over the short period. The new identified positions and speeds are then introduced as new data inputs in step 2. This will speed up the lead discovery process.

14.3.8 Binding-free-energy calculation-MM-PBSA

In the process of molecular docking, majorly one has to see about the affinity of candidate ligands against any molecular target. In this, we can get through the binding energy, that is, free energy released during the interaction. Generally, target structure can be seen as rigid and free energies for protein–ligand complexes calculated using a different scoring function in build with docking software packages. To validate the binding energy obtained during molecular docking, MM-PBSA plays a key role. As proteins were being considered rigid, therefore, no frequencies were calculated for them. The MM-PBSA method usually combines with a molecular dynamics (MD) simulation for the protein–ligand complex and uses a single trajectory approach.

A real-time snapshot of a structure can be taken at various time durations during the running portion of the MD simulations. These real-time data are then utilized to calculate mean values and uncertainties of various parameters of interest. Generally, most of the time MD simulations are always being carried out using the explicit solvent model to acquire the most accurate real-time data possible before mean values' calculations. It is necessary to obtain many conformations of the solvent model in real time over the desired timeframes of simulation for better statistical inferences (Rai et al., 2021). In MM-PBSA, we can easily calculate real values by using simulated information obtained from MDS, that is, vdW energy, electrostatic energy, polar solvation energy, solvent accessible surface area energy, and binding energy. This method will be very helpful in sorting the best ligand structure for further research.

14.4 Recent development in SBDD

Any small molecular interference by any entity that can interact with the corresponding protein is responsible for disease mechanism or any undesired inhibition reactions it forms as the primary step to implement a very novel concept for the discovery process, that is, SBDD. To enhance the effectiveness of such a design, atomic state computational modeling can be seen as significant progress. However, it has been realized over the years that the accurate calculation of binding-free energy for target protein and ligand emerged as one of the most significant problems of such modeling as it involves the use of many scoring functions. However, in past years numerous important, reliable, and well-established docking tools are brought into consideration. This shortlisting is done to overcome the problems related to global optimization. Other than this efforts are made to use different algorithms, and the well-known reliable genetic algorithm is well established and received attention because of its finely define design and popularity with respect to some other algorithms. Besides this, methods utilizing solvent implicit models are also being considered for more accurate free-energy calculations.

Some newly developed docking programs are also available, as they are capable to detect the minimized lower energy spectra for ligand–protein complex. Those are being considered more accurate programs because they lack a predefined pool of potential from ligand–protein interaction. These new docking algorithms analyze docking ligand into a protein and utilize some dozen of mobile entities on the other side to look for the search of surrounding energy minimization (Sulimov, Kutov, & Ilin, 2019). These algorithms improve the accuracy positioning of the ligand in the docking. Quantum chemistry involvement and advancement have improved the accuracy of the docking process. In addition, other side developments have been made in calculations of molecular energy, utilizing solvent models, quantum-chemical methods, and the flexibility and atoms of the ligand mobility to get minimized energy spectra. The artificial intelligence (AI) technique uses human behavior by implementing human intelligence as the means of

computer techniques (Mak & Pichika, 2019). Another machine learning (ML), a subset of AI, uses different statistical methods with or without being programmed (Bishop, 2013). In SBDD, AI has come up with shifting the mood from hype to hope (Mak & Pichika, 2019).

In past years AI and ML algorithms have revolutionized the process of drug discovery in various concerned industries by providing a stage to efficiently identified more new compounds and more accurately. In this integrated approach of ML and AI, big data emerges as an extra extension to improve these methods. Big Pharma has been seen to increase investment in AI; this is a clear benefit insight of success of these methods. For example, SYNSIGHT has launched an AI-based VS and molecular modeling-integrated platform in combination to create large biological models for drug development. In addition, Pfizer is also been collaborating with IBM since December 2016 utilizing their multi cloud platform Watson drug discovery process in terms of immuno-oncology. ML is characterized into three types; these are supervised learning, unsupervised learning, and reinforced Learning. The variable ML subcategory methods analyze the model based on the availability of input and output data sources. Supervised ML integrated with ADMET analysis is majorly applicable in disease diagnostic methods; ADMET data can be passed as an output, and it improves drug efficacy.

Algorithms of supervised ML use binary coded prediction to differentiate between a potential drug compound and an incapable candidate. The methods using the unsupervised learning category use a disease similar type as an output and a feature-finding method can be used to identify a target for any disease. ML helps in stabilizing the relationship between a biological action and chemical structure during drug design that is a primary development. Predicting the structure of biological targets and QSAR (quantitative structure-activity relationship) models, molecular docking analysis, and many scoring functions that are being used for similarity searches can be implemented as well as cross-validated by the use of ML techniques. Pharmacokinetic and toxicological (ADMET)-related data classification, discovery, biologically active enhancement versatile compounds, and the modified constructed biological activity of a new ligand can help in the SBDD process at several steps using ML techniques. ML-processed data can be developed to libraries so that can lower the generation of false positives and can have reproducibility along with a good number of active compounds as an output. Using versatile training datasets and predicted inactive compounds can achieve this target.

14.5 Challenges and limitations

Although the SBDD technique has proved as a pioneer in the drug developmental process but there is a lot of complexities or challenges where the community has to look upon, generally these involve enhancement of the efficacy of computational screening methods, chemogenomic computational data improvement, quantity, and quality of improvement of computational tools and databases, modifying in multitarget drug structures, toxicity predicting algorithm improvement, and collaboration or integrating the approach with other related areas to get better efficacy via lead identification as well as optimization. Combinatorial chemistry, HTS, and VS approaches are used to limit the time and cost of the drug discovery process.

As to overcome some significant limitations of these techniques, toxicological informatics system development is needed. Utilizing de novo lead identification process, keeping aside its efficiency and acceptability, there are many limitations associated with SBDD. These automated methods do not provide flexibility to accommodate the lead inside the binding cavity and associated manual work one has to do. Furthermore, due to instability, lead compounds developed using this approach are not always flexible to modify or to easily synthesize in laboratory conditions. Thus some novel reliable sources are needed that overcome the difficulty for synthesis factors. Coming to molecular docking, several tools utilizing different evaluating algorithms and scoring functions are available in the scientific community.

However, it is not so certain to rely on a particular tool; it is important to choose a suitable scoring/ranking function. Among different docking programs, these scoring functions have a major drawback because this evaluates only ligand-binding energy but not accuracy. Other important surrounding parameters, such as electrostatic interactions and entropy calculations, are completely ignored. There is not a single software package that can work for all types of molecular targets and ligands. In addition, after some incidence, it has been seen that accurate binding energy calculation algorithms are not trustworthy. Despite SBDD get a lot of advancement, a consistent package is yet to be developed. On the other hand, issues, such as considering the involvement of water molecules and probable conformation for the target, and other significant innovations, are still needed.

14.6 Future prospective

SBDD emerged as a promising tool when used as an armamentarium for the identification of new potential compounds. New compounds having drug-like activity can be designed against a target followed by further optimization and

synthesis for practical use. As areas, such as computational biology, mathematical modeling, and system biology, are in their advancement stage, structural genomics techniques are also being benefitted and seem to be getting more in the future. The integration of AI and ML and their integration with these structure-based techniques emerge as a revolutionary step in this regard. As discussed above, it has immense potential in the discovery of therapeutics. So, a scientific community just needs to overcome the problems associated with the algorithms, scoring functions, and availability of present data. This is going to accelerate the clinical trials and limit the failure of promising drug candidates. Afterward, SBDD can emerge as a novel, reliable, and efficient technique for drug designing. That can help industries to cut their cost and time to launch a drug into the market.

14.7 Conclusion

The introduction of molecular modeling tools, growth in structures determination through experimental techniques, such as NMR and X-ray crystallography, for the macromolecular target structure, and availability of the database resources for ligand structures provide a lot of advantages over other conventional studies. SBDD techniques are a significant approach and can be applied in terms of target-based therapies. AI and ML are the reliable approaches that are being prioritized in terms of computational methods and identifying versatile molecules or their fragments having drug-like capabilities. Drug developmental studies should be more accurate and should cut down the failure rate of promising compounds. There is a need to integrate ML, AI, and available genomic information to improve the available resources used for SBDD. Consistent packages and algorithms need to be developed so that compounds having drug-like activity can be screened and shorted using molecular docking, ADMET, MDS, and MM-PBSA as well as can be synthesized for practical applications through synthetic chemistry approaches so that we can have a very reliable and cost-effective method for drug designing in the future.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- Agnihotry, S., Pathak, R. K., Srivastav, A., Shukla, P. K., & Gautam, B. (2020). Molecular docking and structure-based drug design. In D. B. Singh (Ed.), *Computer-aided drug design* (pp. 115–131). Springer. (ISBN: 978-981-15-6815-2).
- Agrwal, A., Pathak, R. K., & Kasana, V. (2021). Molecular docking and antibacterial studies of pyranopyrazole derivatives synthesized using [Pap@Glu-Chi] biocatalyst through a greener approach. *Arabian Journal for Science and Engineering*. Available from <https://doi.org/10.1007/s13369-021-05377-1>.
- Anderson, A. C. (2003). The process of structure-based drug design. *Chemistry & Biology*, *10*(9), 787–797.
- Banner, D., & Wirtz, B. (1993). Serine proteases: 3D structures, mechanisms of action and inhibitors. In B. Testa, E. Kyburz, W. Fuhrer, & R. Giger (Eds.), *Perspectives in medicinal chemistry. Proceedings of the XIIIth international symposium on medicinal chemistry* (pp. 27–43). Basel: Verlag Chimica Acta.
- Batool, M., & Choi, S. (2017). Identification of druggable genome in *Staphylococcus aureus* multidrug resistant strain. In *Proceedings of the 2017 IEEE life sciences conference (LSC)*, Sydney, NSW, Australia (pp. 270–273), December 13–15, 2017.
- Bishop, C. M. (2013). Model-based machine learning. *Philosophical Transactions of The Royal Society A: Mathematical Physical and Engineering Sciences*, *2013*(371), 20120222.
- Blaney, J. (2012). A very short history of structure-based design: How did we get here and where do we need to go? *Journal of Computer-Aided Molecular Design*, *2012*(26), 13–14.
- Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, ... Kollman, P. A. (2020). *AMBER 2020*. San Francisco: University of California.
- De Paulis, T. (2007). Drug evaluation (2007) Prx-00023, a selective 5-HT_{1A} receptor agonist for depression. *Current Opinion in Investigational Drugs (London, England: 2000)*, *8*, 78–86.
- Franca, T. C. (2015). Homology modelling: An important tool for the drug discovery. *Journal of Biomolecular Structure & Dynamics*, *33*(8), 1780–1793.
- Gallego, J., & Varani, G. (2001). Targeting RNA with small molecule drugs: therapeutic promise and chemical challenges. *Accounts of Chemical Research*, *34*, 836–843.
- Gashaw, I., Ellinghaus, P., Sommer, A., & Asadullah, K. (2012). What makes a good drug target? *Drug Discovery Today*, *17*(Suppl), S24–S30.
- Hosfield, D., Palan, J., Hilgers, M., Scheibe, D., McRee, D. E., & Stevens, R. C. (2003). A fully integrated protein crystallization platform for small-molecule drug discovery. *Journal of Structural Biology*, *142*(1), 207–217.

- Huang, S. Y., & Zou, X. (2010). Advances and challenges in protein-ligand docking. *International Journal of Molecular Sciences*, 2010(11), 3016–3034.
- Irwin, J. J., & Shoichet, B. K. (2005). ZINC_ a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, 45(1), 177–182.
- Jayaram, B., Singh, T., Mukherjee, G., Mathur, A., Shekhar, S., & Shekhar, V. (2012). Sanjeevini: A freely accessible web-server for target directed lead molecule discovery. *BMC Bioinformatics*, 13(17), 1–13.
- Joseph, A., DiMasi, Henry, G., & Grabowski. (2012). R&D costs and returns to new drug development: A review of the evidence. *Journal of Health Economics*, 47, 20–33.
- Kaur, K., Utreja, D., Dhillon, N. K., Pathak, R. K., & Singh, K. (2020). N-alkyl isatin derivatives: Synthesis, nematocidal evaluation and protein target identifications for their mode of action. *Pesticide Biochemistry and Physiology*, 104736.
- Kawato, T., Mizohata, E., Shimizu, Y., Meshizuka, T., Yamamoto, T., Takasu, N., Matsuoka, M., Matsumura, H., Kodama, T., Kanai, M., Doi, H., Inoue, T., & Sugiyama, A. (2015). Structure-based design of a streptavidin mutant specific for an artificial biotin analogue. *Journal of Biological Chemistry*, 157(6), 467–475.
- Kelley, L., Mezulis, S., & Yates, C. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10, 845–858.
- Khanna, I. (2012). Drug discovery in pharmaceutical industry: Productivity challenges and trends. *Drug Discovery Today*, 17(19–20), 1088–1102.
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., & Wang, J. (2016). PubChem substance and compound databases. *Nucleic Acids Research*, 44(D1), D1202–D1213.
- Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nature Reviews. Drug Discovery*, 3, 935–949.
- Kumari, R., Kumar, R., Open Source Drug Discovery Consortium., & Lynn, A. (2014). g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. *Journal of Chemical Information and Modeling*, 54(7), 1951–1962.
- Laurie, A. T., & Jackson, R. M. (2005). Q-SiteFinder: An energy-based method for the prediction of protein-ligand binding sites. *Bioinformatics (Oxford, England)*, 21, 1908–1916.
- Li, S., Hsu, C. W., Sakamuru, S., Zou, C., Huang, R., & Xia, M. (2017). Identification of angiogenesis inhibitors using a co-culture cell model in a high-content and high-throughput screening platform. *SLAS Technology*, 1, 2472630317729792.
- Lopez-Vallejo, F., Caulfield, T., Martinez-Mayorga, K., Giulianotti, M. A., Nefzi, A., Houghten, R. A., & Medina-Franco, J. L. (2011). Integrating virtual screening and combinatorial chemistry for accelerated drug discovery. *Combinatorial Chemistry & High Throughput Screening*, 2011(14), 475–487.
- Mak, K. K., & Pichika, M. R. (2019). Artificial intelligence in drug development: Present status and future prospects. *Drug Discovery Today*, 2019(24), 773–780.
- Mamgain, S., Sharma, P., Pathak, R. K., & Baunthiyal, M. (2015). Computer aided screening of natural compounds targeting the E6 protein of HPV using molecular docking. *Bioinformation*, 11(5), 236.
- Mandal, S., Moudgi, M., & Mandal, S. K. (2009). Rational drug design. *European Journal of Pharmacology*, 625(1–3), 90–100.
- Meng, X. Y., Zhang, H. X., Mezei, M., & Cui, M. (2011). Molecular docking: A powerful approach for structure-based drug discovery. *Current Computer-Aided Drug Design*, 2011(7), 146–157.
- Middleton, D. A. (2007). Solid-state NMR spectroscopy as a tool for drug design: From membrane-embedded targets to amyloid fibrils. *Biochemical Society Transactions*, 35(Pt5), 985–990.
- Nim, Y. S., Sun, S., & Wong, K. B. (2017). Using homology modeling to understand the structural basis of specific interaction of a plants specific AtSarla-AtSec23a pair involved in protein ER Export. *Methods in Molecular Biology*, 2017(1662), 59–73.
- Pathak, R. K., Baunthiyal, M., Taj, G., & Kumar, A. (2014). Virtual screening of natural inhibitors to the predicted HBx protein structure of Hepatitis B Virus using molecular docking for identification of potential lead molecules for liver cancer. *Bioinformation*, 10(7), 428.
- Pathak, R. K., Gupta, A., Shukla, R., & Baunthiyal, M. (2018). Identification of new drug-like compounds from millets as Xanthine oxidoreductase inhibitors for treatment of Hyperuricemia: A molecular docking and simulation study. *Computational Biology and Chemistry*, 76, 32–41.
- Pathak, R. K., Singh, D. B., Sagar, M., Baunthiyal, M., & Kumar, A. (2020). Computational approaches in drug discovery and design. In D. B. Singh (Ed.), *Computer-Aided Drug Design* (pp. 1–21). Springer. (ISBN: 978-981-15-6815-2).
- Pei, J., Yin, N., Ma, X., & Lai, L. (2014). Systems biology brings new dimensions for structure-based drug design. *Journal of the American Chemical Society*, 136(33), 11556–11565.
- Pence, H. E., & Williams, A. (2010). ChemSpider: An online chemical information resource. *Journal of Chemical Education*, 87(11), 1123–1124.
- Rai, S., Pathak, R., Singh, D. B., Bhatt, A., & Baunthiyal, M. (2021). Chemo-informatics guided study of natural inhibitors targeting rho GTPase: A lead for treatment of glaucoma. *In Silico Pharmacology*, 9. Available from <https://doi.org/10.1007/s40203-020-00061-y>.
- Rana, G., Pathak, R. K., Shukla, R., & Baunthiyal, M. (2019). *In silico* identification of mimicking molecule(s) triggering Von Willebrand factor (VWF) in human: A molecular drug target for regulating coagulation pathway. *Journal of Biomolecular Structure & Dynamics*, 38(1), 124–136.
- Ren, J. X., Li, L. L., Zheng, R. L., Xie, H. Z., Cao, Z. X., Feng, S., Pan, Y. L., Chen, X., Wei, Y. Q., & Yang, S. Y. (2011). Discovery of novel pim-1 kinase inhibitors by a hierarchical multistage virtual screening approach based on svm model, pharmacophore, and molecular docking. *Journal of Chemical Information and Modeling*, 51, 1364–1375.
- Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738.
- Rutenber, E. E., & Stroud, R. M. (1996). Binding of the anticancer drug zd1694 to E. coli thymidylate synthase: Assessing specificity and a_nity. *Structure (London, England: 1993)*, 4, 1317–1324.

- Sagar, M., Pathak, R. K., Pandey, R. K., Singh, D. B., Pandey, N., & Gupta, M. K. (2014). Binding affinity analysis and ADMET prediction of epigallocatechin gallate (EGCG) derivatives for AP-1 protein: A drug target for liver cancer. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 3(1), 66.
- Saleh, N. A., & Elshemy, W. M. (2017). Structure-based drug design of novel peptidomimetic cellulose derivatives as HCV-NS3 protease inhibitors. *Life Sciences*, 2017(187), 58–63.
- Sidhu, K. S., Bhangu, S. K., Pathak, R. K., Yadav, I. S., & Chhuneja, P. (2020). Identification of natural lead compounds for leaf rust of Wheat: A molecular docking and simulation study. *Journal of Proteins and Proteomics*. Available from <https://doi.org/10.1007/s42485-020-00048-5>.
- Siegel, R., Naishadham, D., & Jemal, A. (2013). Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians*, 63, 11–30.
- Singh, D. B. (2014). Success, limitation and future of computer aided drug designing. *Translational Medicine*, 4, 4.
- Singh, D. B., & Dwivedi, S. (2016). Docking and molecular dynamics simulation study of inhibitor 2-fluoroaristeromycin with anti-malarial drug target PfSAHH. *Network Modelling and Analysis in Health Informatics and Bioinformatics*, 5, 16.
- Singh, D. B., & Pathak, R. K. (2020). *Computational approaches in drug designing and their applications. Experimental protocols in biotechnology* (pp. 95–117). New York, NY: Humana.
- Sotriffer, C., & Klebe, G. (2002). Identification and mapping of small molecule binding sites in proteins: Computational tools for structure-based drug design. *Pharmacology*, 57(3), 243–251.
- Sulimov, A. V., Kutov, D. K., & Ilin, I. S. (2019). Docking of oligopeptides. *Russian Chemical Bulletin*, 68, 1780–1786.
- Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research*, 33, W244–W248. (Web Server issue).
- Tian, W., Chen, C., Lei, X., Zhao, J., & Liang, J. (2018). CASTp 3.0: Computed atlas of surface topography of proteins. *Nucleic Acids Research*, 46(W1), W363–W367.
- Tollman, P. (2001). A revolution in R&D: How genomics and genetics are transforming the biopharmaceutical industry. *Acta Metallurgica Sinica*, 41(6), 611–616.
- Trott, O., & Olson, A. J. (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Journal of Computational Chemistry*, 31, 455–461.
- Urwiler, S. (2001). Allosteric modulation of family c g-protein-coupled receptors: From molecular insights to therapeutic perspectives. *Pharmaceutical Reviews*, 63, 59–126.
- van Gunsteren, W. F., & Berendsen, H. J. C. (1987). *Groningen Molecular Simulation (GROMOS) Library Manual* (pp. 1–221). Groningen, The Netherlands: Biomos.
- Warren, G. L., Do, T. D., Kelley, B. P., Nicholls, A., & Warren, S. D. (2012). Essential considerations for using protein-ligand structures in drug discovery. *Drug Discovery Today*, 17(23–24), 1270–1281.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303.
- Webb, B., & Sali, A. (2016). Comparative protein structure modelling using modeller. *Current Protocols in Bioinformatics*, 54, 5.6.1–5.6.37.
- Wilson, G. L., & Lill, M. A. (2011). Integrating structure-based and ligand-based approaches for computational drug design. *Future Medicinal Chemistry*, 3, 735–750.
- Wishart, D. S., Knox, C., Guo, A. C., Shrivastava, S., Hassanali, M., Stothard, P., Chang, Z., & Woolsey, J. (2006). DrugBank: A comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Research*, 34(Suppl. 1), D668–D672.
- Wlodawer, A., & Vondrasek, J. (1998). Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Annual Review of Biophysics and Biomolecular Structure*, 1998(27), 249–284.
- Zhao, C., Xia, C. G., Yu, M., Pan, Y., & Wang, L. (2015). Application of molecule docking software in drug design. *Chinese Journal of Antibiotics*, 4(3), 234–240.