REVIEW ARTICLE

TECHNIQUES AND ALGORITHMS FOR STRUCTURE-BASED VIRTUAL SCREENING (SBVS): AN OVERVIEW

Raksha K. Rao[a]*, Somdatta Y. Chaudhari[a], Shailaja B. Jadhav[a], Pravin D. Chaudhari[b] and Ujwala B. Yedake[a]

(Received 21 January 2022) (Accepted 27 October 2022)

ABSTRACT

Today, the world of science is constantly challenged with new genomics, which in turn is responsible for new disease-causing targets. Hence, there is a need for developing drugs acting against such targets. Computational methods are proving to be a mainstay in the drug discovery process, mainly through virtual screening. This review discusses about the recent advancements in structure-based drug design with reference to Virtual Screening along with its procedures from ligand preparation and protein preparation, docking, scoring function, databases, and virtual (VS) algorithms. Application of Structure-based VS in combination with other virtual screening techniques has also been highlighted in this review.

Keywords: Virtual Screening, structure-based virtual screening, docking, library preparation, scoring functions, algorithms, combined application

ABBREVIATIONS

VS= Virtual screening, CADD= Computer aided drug design, SBVS= Structure-based virtual screening, LBVS= Ligand based virtual screening, HTS= High throughput screening, NMR= Nuclear magnetic resonance, PDB= Protein data bank, MD= Molecular Dynamics, ML= Machine learning, ANN= Artificial neural network

INTRODUCTION

Drug discovery aims to find a biologically active compound that has a biological target with greater sensitivity and further generate, the desired biological effect. Due to the high cost and the poor success rate of high-throughput screening (HTS), computer alternatives have been developed, and *in silico* screening has been more widely used[1]. SBVS is a computer-based strategy for evaluating a chemical compound database for new biologically active compounds targeting a specific therapeutic target during the early phase in drug development. Molecular docking is used to study, whether, molecules would attach to a molecular target in its three-dimensional structure[2,3].

In SBVS, the goal is to figure out which molecule binds to the target when the library is screened. Protein-ligand interaction and its bonding label are coupled in a computer model known as a "classical scoring function"[4]. X-ray technology and nuclear magnetic resonance spectroscopy have set the groundwork for SBVS by providing extensive structural information about these receptors as well as knowledge about how they interact with ligands[5].

Introduction to drug design

Despite improvements in biotechnology and our knowledge of living materials, designing a new active synthetic molecule is a lengthy, draining and sophisticated process with a high failure rate of novel therapeutics[6]. Molecular design is a series of stages that begin with a designed molecule[7]. Then, it is charged appropriately, for the target molecule with which it interacts and binds. This process, known as drug design, is the initial step in creating molecules. Drug discovery involves identifying novel medicinal substances using a combination of computer algorithms, translational and clinical models[8]. The entire lead invention or optimization of an existing lead is included

[a] Department of Pharmaceutical Chemistry, Progressive Education Society's Modern College of Pharmacy, Nigdi, Pune – 411 044, Maharashtra, India

[b] Department of Pharmaceutics, Progressive Education Society's Modern College of Pharmacy, Nigdi, Pune – 411 044, Maharashtra, India

* For Correspondence: E-mail: raksharao9598@gmail.com

https://doi.org/10.53879/id.61.01.13346

in drug design. Structure-based drug design relies on the 3D structure of a desired molecule, such as amino acid sequence, etc. to find critical locations and connections that are essential for their biological activities[9]. This knowledge is further useful in discovering medicines that compete with the target's critical interactions, disrupting the biological pathways required for the microbe's existence. Ligand-based drug design (LBDD) approach uses a known set of ligands which have structural similarity[10,11].

Methods in drug designing process

The following are some of the methodologies utilized in drug development.

- *In vitro* approach is used to screen synthetic chemicals and natural materials arbitrarily[12]
- Preparation of new compounds using a known arrangement of physiologically active, organic chemicals of plant and animal origin, such as the lead framework
- Synthesis of a lead substructure of a compound with higher bioactivity[13]

Types of drug designing in Computer-Aided Drug Design (CADD)

Indirect /Ligand-based drug design

A group or library of compounds that adhere to the physiological substrate or ligand of focus is used in ligand-based drug design. These additional compounds are utilized to create a design, which further clearly states the structural properties that a moiety must have to interact with a target[14]. In a different sense, based on knowledge of what binds toward a biological target, a prototype of both the target and the binding site may be produced and this description is further used to invent different structural interaction between the systems with it[15].

Structure or target – based drug design

It relies on the substructure of the physiological target namely protein. A candidate moiety that bonds with greater sensitivity as well as specificity for the target molecule is produced utilizing graphical visuals based on the structure of the configuration[13]. Approaches for structure-based drug design is database searching, and involves the process of finding ligands for a specific receptor. A huge number of possible ligand molecules are screened to see which ones fit the receptor's binding domain. One of the main benefits of data analysis is the effort and money saved[16].

This article reviews the applications of structure-based virtual screening alongside its role as an important virtual screening technique.

Virtual screening

Virtual screening is an application in the field of drug discovery. It is applied as an addition to High-Throughput Screening (HTS), so when paired with biological research, it has the potential to raise the series of tasks at the lead identification phase of the process of discovery while also boosting their efficacy[17]. There are a variety of techniques for undertaking these computational assessments, and they are either ligand-based and receptor-based[18, 19].

DOCK, Glide and GOLD are just a few of the protein-ligand docking courses offered[20]. These include molecular docking within each receptor into the target's binding domain, resulting in a projected binding pattern for each database chemical, and therefore a metric of the chemical's fit in the target's binding pocket. This finding aids in the assessment of the chemicals to identify and test a small percentage for biological activity[21, 22].

In many of these implementations, it's algorithm generates ligand conformations, and then uses a scoring function to optimize both alignments and rotations to lower their binding free energy with the biological target[23, 24] as indicated in Fig. 1.

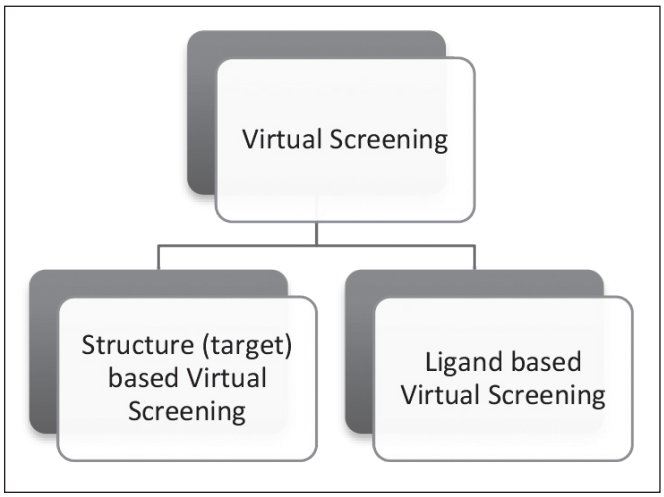

Fig. 1: Types of virtual screening[111]

Structure - based virtual screening

A computerized tool incorporates the ligands to the binding site of the target. Selection of targets and databases, docking and post-docking monitoring, and molecule selection for screening are the processes employed in SBVS[24,25]. This kind of task requires a complex computational infrastructure, which can perform various calculations and process simultaneously using a set of operating systems docking[26]. It is necessary to have a

system in place to process the input from big compound libraries. There is also a professional database of chemicals.

The procedure begins by deriving the target sequence's three-dimensional organization. X-ray, nuclear magnetic resonance (NMR) and other molecular dynamic (MD) stimulation tools work in characterizing the target structure[27]. Another factor to consider for SBVS is the detailed list of the synthetic database to be filtered in the VS operation based on the target of interest, along with library preparation to attribute suitable reaction mechanisms, tautomeric and charge transfer states[28]. Docking procedure attempts to anticipate the ligand-protein conformations by examining the overall structure of such ligands inside the protein's ligand binding. Scoring-system estimates the free binding energy in the protein and the ligand[5, 29, 30].

Advantages of SBVS

- It is cost effective as compared with traditional screening techniques.
- It does not require a synthetic drug molecule, as it can be created virtually.
- Today various commercial tools are available for SBVS.

Disadvantages of SBVS

- There is the possibility of getting misleading results.
- It is problematic in case if knowledge of receptor is unavailable.
- Commercially available tools are usually a bit specific.

Even though there are drawbacks in SBVS tools, advancements in virtual screening are capable of diminishing these setbacks with time[31,32].

Strategies in virtual screening

Combining LB (Ligand- based) and SB (Structural-based) techniques may be done in a variety of ways. In this review, categorization will be used. As a result, the combination of LB and SB methods is discussed in three types: sequential, parallel, and hybrid as mentioned in Fig. 2.

Sequential approach

Sequential methods split the VS pipeline across segments with the goal of performing progressive filtering in the chemical compound library to identify the most suitable candidates, for bioassays at the end. Pre-

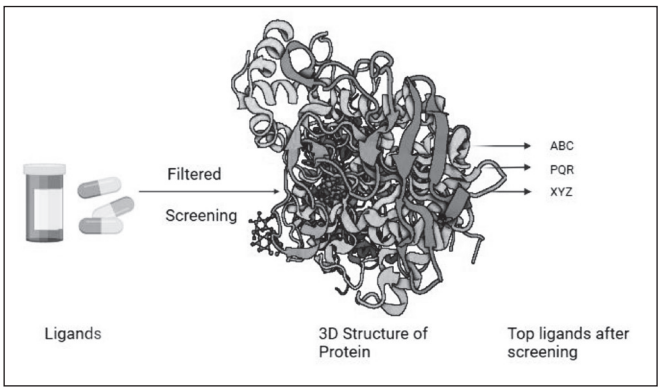

Fig. 2: The screening workflow of ligands against known protein receptor[109, 110]

filtering utilizing LB methods is common since they are less expensive to process. The more computationally intensive SB is used later in the investigation. As a result, such a method aims to maximize the tradeoff between the computational costs of the filtering technique, and thus, the intricacy of the paradigm underpins it along the VS process. It does not employ all the existing evidences instantly, and they keep the limitations of various approaches in mind.

Parallel approach

The LB and SB processes are carried out independently, with leading candidates of each approach being chosen for biological evaluation. The compounds in the final rank order outperform single-modality methods in terms of stability and performance, but still the findings also suggest that performance is susceptible to target structural characteristics, such as the existence of the template ligand in molecule resemblance dimensions and samples were determined in a pocket in docking studies.

Hybrid approach

It consists of methods that are a real mix of LB and SB procedures in a single method. To achieve this aim, two major combinations used are: (i) interaction and (ii) similarity docking techniques. In the first it transforms measurable protein-ligand interaction into pharmacophore parameters and QSAR models[33]. They are applied in VS, ligand profiling, pseudo-receptor analysis, and *de novo* designs. Researchers have recently looked at employing a mix of molecular similarity and docking techniques to test projected ligand poses against relevant templates.

Types of structure based virtual screening

Protein- ligand binding

The most well-known structure-based 3D approach is protein-ligand docking. It calculates the potential binding

affinities of a known protein receptor in 3D. A variety of docking tools and techniques, both of which are widely tested are available. They contrast in terms of ligand positioning and scoring systems. Ligand fragmentation and incremental reconstruction, molecular shape-based algorithms, genetic algorithms, systematic search, and surface-based molecular similarity are among some of the strategies used[34]. Ligand placement is not trustworthy for sampling possible ligand binding geometries for all targets. Professionally developed empirical scoring strategies in each application scope are indeed significant, and docking remains one of several ways that provides numerous ideas for logical lead optimization[35].

Pharmacophore based models

A three-dimensional configuration of chemical compositions, such as hydrogen bonds, charges, and lipophilic areas, form a design that depicts the binding of a ligand to a macromolecule[36]. The merits of pharmacophore models include their ease of interpretation, and its flexibility is modified, which allows them to be used in a variety of situations[37]. Researchers can use a transparent, yet nevertheless highly descriptive model to include their knowledge of a certain binding mode[38]. This has led to several success stories, for active ligand affinity elucidation and, as a result, prospective ligand design[34].

Binding site comparisons

The binding site comparison technique is a relatively new concept that comes under SBVS' jurisdiction[35].It is based on the fact that proteins share identical binding affinity due to their amino acid sequence. As a result, binding sites in proteins may be identified, matched to specific chemical structures based on the amino acid sequence. It also considers protein flexibility.

Steps in structure-based virtual screening

Database preparation

The development of a ligand or drug library or database is a crucial step for executing VS using the docking or pharmacophore-based search methods[39]. The database's foundation is often a commercial compilation of physically accessible chemicals[40]. Many physicochemical filters are applied to decrease the size of an initial database of compounds for further screening[99]. Lipinski's rule-of-five is a popular filtering technique for drug-likeness. It is an experimental collection of characteristics based on molecular mass, hydrophobicity and other parameter which are important for orally bioavailable compounds[41]. A restriction on the number of rotating bonds inside the macromolecule or the polar surface area are used as further physiological limits[25]. The "Pfizer's Rule of 3/75," is predicated on calculated partition coefficient along with topographic polar surface area values, and is a more modern rule accounting for the physicochemical characteristics linked to experimental toxicological outcomes (TPSA). Compounds having a ClogP of less than 3 and a TPSA of more than 75 are 50 percent safer, and more reliable *in vivo* studies[42]. At this stage, additional filters are frequently used to exclude compounds that have particular chemical substructures linked to poor chemical stability or toxicity. All these filtering protocols are low-chemically to big databases[36, 43]. Finally, preprocessing compound datasets in realistic 3D representations is recommended, as bond lengths and angles remain similar during docking. The compound set utilized for SBVS should contain true bond lengths and angles[44]. Such compounds will have a suitable conjugated system and filled valences, as well as partial charges, a sufficient proton transfer mode at physiological pH, and precise isomeric configurations[45]. Public databases, like ZINC, are widely utilized specific molecular databases in virtual screening[2].

Target preparation for SBVS

SBVS operation is heavily reliant on acceptable protein and ligand starting structures. Some techniques need the addition of hydrogens, with caution to avoid atomic collisions. A typical PDB format data comprises only heavier atoms, as well as molecules of water, activators, ligands, and ions along with numerous protein units. In addition, there is no information on bond sequencing, topological, or formal atomic charge in general in the structure. Ionized particles along with isomeric states are mostly unallocated, and residues of branched chains are mostly not found due to poor precision of a specific protein region, as well as steric incompatibilities[25]. In order to resolve these issues, the precise protonation states of ionizing residues are necessary. The next stage is to add hydrogen atoms in the protein and optimize them using an optimum hydrogen bond network[46]. Following the insertion of hydrogen atoms into the protein, energy minimization is used to modify the locations of the hydrogens to eliminate any stereo electronic conflicts. A choice is also made, whether water molecules will remain or be eliminated. Several techniques, such as 3D reference interaction site model (3D RISM), WaterMap (WM), and others, have been presented to overcome this complex task. If such protein is co-crystallized with intermediates, co - factors, or other ligands, they should be treated prior to VS. To build 3D geometries, there is need to apply an accurate bond order and establish relevant tautomer and ionization states. These steps could be accomplished

with free tools such as the Protein Preparation Wizard[25] or scripts that generate a protein structure for SBVS using a variety of methods. By comparison of minute structures from the PDB structure, docking efficiency is highly improved.

Identifying the binding site

Identifying the site for binding is often a necessity for SBVS. A target-binding location must ideally be a region with a variety of H-bond acceptors, donors as well as hydrophobic characteristics, and must be usually curved. In the scope of research, there are few ways for finding possible binding sites: Static methods, in which chemical probes are used to discover binding hot spots on a 3D model via computational solvent mapping from X-ray, MD, etc. To use a novel chemical probe every time it determines a new type of binding region[47]. In this perspective, investigations that use microsecond MD simulations or higher sampling approaches to study ligand-binding mechanisms could be used to identify novel binding proteins utilizing minuscule organic compounds as probes, although this strategy is not efficient and robust[48]. To evaluate flexibility, normal mode modeling is utilized to discover elastic residues inside the ligand binding and to investigate alternative conformations of those residues. Water is used as a probe to discover probable protein binding sites in the final process. Approaches founded on the notion of inhomogeneous solvation and evolutionary computation are commercially available[49, 50].

Docking

The next stage in the screening process is to dock compounds from the database. The procedure involves screening the binding site's coordinate space, and then scoring each potential ligand posture, which is then used to determine the compound's anticipated binding model[18]. The database is added into a target-binding site using a docking tool that simulates the ligand–target interaction computationally to ensure optimum steric and physicochemical compatibility[5]. Given the fact that computers are becoming more powerful, blind docking with all library compounds frequently is a waste of time and resources[51]. As a result, before embarking on the time-consuming docking process, it is typically a good idea to eliminate unwanted compounds and pick just relevant ones from a library. DOCK[52], FlexX[53], GOLD[54], Glide[55], and AutoDock[56] are among the most often utilized docking techniques. When choosing docking software, there are a few things to be kept in mind. These features include the ability to alternatively refine docking parameters/protocols regarding new findings, flexibility to additional classifiers, early and late docking filters, design and validation results, tech support, cost and speed, user interface, input/output, structural formats, code access and upgradeability. In investigating the conformations in ligands, docking algorithms use a variety of conformational search techniques,[25] which are classified as follows: a) Systematic techniques, which examine all degrees of freedom, and are placed in the expected binding site. b) Random torsional searches about rotatable bonds (c) Using molecular dynamics (MD) tools to investigate a molecule's energy landscape[57].

Scoring

Docking algorithms employ scoring methods to assess the free available energy in bonding of a ligand to target. Most often used scoring functions involve estimation of binding free energy and force field algorithm that sums the energy of van der Waals, electrostatic forces and H - binding among the atoms of the two binding proteins in the complex[58]. The next is empirical scoring systems. It calculates the interactions between two complicated binders, such as hydrophobic contacts, hydrogen bonds and stranded rotatable bonds. Using empirical or semi-empirical methodologies, scoring systems measure the stiffness of target and ligand interaction[59].

Post analysis

A docking study scan of a high chemical database vs a receptor generates a massive amount of data, as well as each drug's expected binding posture and a projected binding affinity for the target. Without any previous knowledge of compounds, it is possible to evaluate a list of compounds based on their rank ordering.

Virtual databases

A virtual database is a computer model or a scheme comprising of one or many physical data that provides a database for virtual analysis. The 3D target proteins substructure are present as databases and are commercially available. Below are listed some virtual databases[60].

Binding DB- Binding DB includes 1,419,347 binding data points for thousands of protein molecules and small molecules. It is a accessible library of observed target proteins, with an emphasis on affinities between proteins thought to be drug targets and tiny, drug-like compounds[61].

ChEMBL- It is a library of bioactive compounds with 2-D substructures, computed characteristics, and abstracted bioactivities[62, 63, 102].

Protein data bank (PDB) - A free accessible protein database that contains three-dimensional structures of proteins and complex compounds[64].

PubChem - An open software that collects data from other smaller databases. It has access to over 97 million chemicals[65].

ZINC - A library of commercially available chemicals for VS[66]. ZINC has around 230 million 3D compounds categorized depending on their chemistry, pharmacology, etc.

canSAR - It specifically includes database for cancer drug discovery[67] with a million bioactive, small molecule drugs and compounds equivalent to millions of pharmacological bioactivities and calculated chemical properties.

DrugBank- A mix of specific drug pool and complete drug and target data. The database includes medication entries, authorized biotech (protein/ peptide) pharmaceuticals and nutraceuticals along with investigational drugs[68].

VS Algorithms

Algorithms are performing multiple tasks or problems with a pre-determined group of command or instruction. Virtual screening technique is differentiated based on their algorithms as follows:-

Machine Learning (ML) based algorithms- Deep learning and neural networks

Within the scope of VS, it implies how one builds a model to forecast whether a certain chemical will interact with specific target, once screened on a dataset that includes both known binders, and known non-binders[69]. It is difficult to locate such a model. It is critical to evaluate the level of accuracy of any classifier, regardless of its kind[70]. Standardizing a classifier along the same data it was based on, will always overstate its accuracy; consequently, validation[71]. They are tested on independent training sets. Internal validation occurs once the testing set is drawn from the same cohort. When the training and test sets are from separate cohorts, external validation is used[72]. External validation provides much more accurate evaluation of a model's efficiency, while dealing with unknown data collecting adequate data from several cohorts is not always practicable[101]. As a result, most machine learning models are first assessed using internal validation. When a trained and assessed ML scheme has been developed and approved, it is applied to perform VS on extraordinarily huge chemical libraries. The highest-scoring compounds are known as hits[70]. From here, the most promising compounds (known as leads) can be further developed and evaluated in the hope of becoming commercially available drugs[73].

ML approaches have been used in the industry for 15-20 years. Deep Learning has recently gained popularity in the drug discovery sector, particularly in VS. Recent advances in virtual screening (VS) determine deep learning (DL) applications in improving binding in drug and target, its activity and also its potency[63]. The techniques that employ protein and ligand information are the focus of our review[74, 106].

Neural networks

Numerous classifiers exist, each with its own set of pros and downsides. Under these neural networks, one classifier in particular that is getting a lot of attention and is quite impactful is the Artificial Neural Network (ANN). It is widely studied[75].They are composed up of connected layers of nodes that convert input to output using weights bias. In a typical ANN, input layer, hidden layer and output layer are three types of layers[103].Those layers that manipulate data between the input and output levels in order to create predictions have a less straightforward name but are mathematically straightforward. ANN can have as many hidden layers as it desires[105]. A neural network is a type of Deep Learning technique (DL). It is a subset of ML, where input is converted into other representations to extract patterns more effectively. To convert the data nonlinearly, the resultant numbers are put into an activation function[76].

Support vector machine-SVM is based on knowledge of input-output pairs, learning an operation that transfers an input to an output. It is related to learning algorithms for database classification and regression analysis[77].

Bayesian techniques- Bayesian technique generates a sophisticated approach for analyzing complex data in current computational advancements[78].

Decision tree- Decision tree is a relatively new statistical learning method based on a algorithms to predict QSARs. Training model is determined through decision tree, and is used to understand the targets importance by learning basic past data (training data). Unlike others, the decision tree model may also be applied to regression and classification issues[79].

Ensemble methods- It is a machine learning approach, which integrates many training sets to build a single best

prediction model. It may make extensive use of decision trees to define as well as demonstrate the utility of ensemble methods[80].

Evolutionary algorithms (EA)- It matches with the natural selection process, which involves initialization, selection, genetic operations and termination processes. Several phases all generally correlate to just a different aspect in natural processes, thus giving a simple techniques for planning tasks and implementation of this algorithm category[104]. Briefly said, during an EA, better individuals shall live to multiply, but unsuitable individuals may perish and bring nothing at all to the genetic pool for subsequent generations, like in natural processes[81].

Genetic algorithms- It primarily depends on bioinspired variables such as mutated gene, recombination, and selection. They are employed to develop greater optimal solutions problems[82].

Differential evolution (DE) - DE is a method that is useful for multivariate real-valued functions. They make such little preconceptions about the issue at hand but also can research incredibly huge areas of possible solutions. Gold and Surflex are examples of DE methods[83].

Local search such as AutoDock Vina, Swiss-Dock, Glam-Dock- this are open source softwares, free for academic use[24,84.] Only Swiss dock does not have protein flexibility. It is similar or relates to genetic algorithms. They operate on different platforms like Windows, Linux and Mac.

Exhaustive search like eHiTS- It is an advanced ligand docking technique. It can address that area of the conformational and positional search area, which prevents significant steric conflicts routinely, yielding high precision docking positions at a rate suitable for virtual screening. The scoring functions, involves traditional empirical and statistical view in respect with novel terms, which depends on local surface point contact. It is free for academic use and operates on Unix platform[85].

Simplex method- This technique is a linear programming method that involves an algebraic process that employing a sequence of repeating operations that arrives at a precise solution[107]. The task can have any number of variables and constraints, but it is extremely difficult to solve mechanically, if there are more than four variables. For a high number of variables, a computer is required[86].

Systematic methods like FlexX, Surflex, and Sybyl-X- These are commercial software products available for docking. Here, only FlexX does not involve protein flexibility. They operate on Windows, Mac and Linux. The docking algorithm used here is of incremental construction. Scoring functions are either empirical or force field[87].

Statistical methods- Knowledge based potentials is a type of scoring functions, which is based on observations in pairwise distribution in statistical analysis.

Monte Carlo- It is a dynamic modeling method that mainly interprets about the repetitive recurrent sampling that can provide a statistical solution. The basic approach is to use randomness to create answers and is, in principle, predictable[88]. Algorithms are widely used in theoretical and numerical problems, and they are especially useful when it's difficult for other different techniques. Monte Carlo techniques are commonly applied to overcome problems such as optimization, numerical integration and generating drawings out of a probability distribution.

Simulated annealing (SA)- This method is used for conditions, which are either continuous or differentiable. It is only essential to compute the criteria value for each conceivable solution. As a result, the condition is provided in a sort of function or an algorithm, which provides numeric values based on the values of both the parameters, which specify a result. All variables should be of a subjective character, with the sole requirement being that they can be used to generate a numerical value of the criteria from their values[89, 100].

Conformational space annealing (CSA) - The CSA algorithm is easy to apply, and it decreases the development interval. It's application in optimization makes it useful in protein structure determination and graph analysis[90,91].

Different techniques used in combination with SBVS

There is continuous development and testing of a variety of techniques and algorithms for different purposes[36]. Even though SBVS is a well-established and well-proven technology, it still faces several obstacles and issues that are addressed, including the detrimental impact of protein elasticity or improper estimation of binding affinity[60]. Here, a variety of combinations of tactics or processes is briefly given.

In general, combined techniques begin with cost-cutting measures in the early stages of VS and progresses to more precise but time-consuming procedures in the latter stages. Many notable VS implementations have demonstrated the obvious benefits of integrated

protocol[92]. Novel, non-steroidal- 11 beta-hydroxysteroid dehydrogenase type-1 inhibitors with IC_{50} values in the micro molar range have been discovered using a hybrid ligand and structure-based VS method[93].

The role of DNA G-quadruplex configurations in biological ageing as well as cancer has been established, that is prompting a quest for selective DNA secondary structure binders. By combining ligand and structure-based methods with docking experiments, we were able to perform slightly elevated *in silico* screening of commercially available molecular databases while leveraging existing structural and biological knowledge on such structures. These experiments resulted in the identification of such a potential scaffold for G-quadruplex binders with furan ring fused to a coumarin moiety[94].

Three multiple regression approaches have been applied to multidimensional scoring data from different target proteins[108]. The objective is to build classifiers that could distinguish between bioactive or inactive chemicals using a virtual screen technique based on structure. The score matrices were created using seven distinct scoring functions. Based on these findings, a new strategy for SBVS is proposed when only minimal activity information is available[95].

Applications of structure based virtual screening in COVID-19

SARS-major CoV-2's proteolytic enzymes main protease M^{PRO}, has a proteolytic function in the production of viral polyproteins that are required for virus replication. Because of its crucial involvement in viral replication, main protease is a prospective target for antagonist effects and possible therapeutic therapy for new coronavirus infections[96]. Alphaketo-amide inhibitor exhibited inhibitory activity against SARS-CoV2 major protease purified recombinant protein using SBVS[97]. This investigation has two objectives, one to perform comparison in protein sequence, and second, to perform three - dimensional structural assessment to analyse mutation on an active amino acid sequence. Mutations of Ser46 and Phe134, produces a substantial alteration in SARS-active CoV-2's sites[98].

CONCLUSION

Previously, the drug discovery process was lengthy and was based on trial and error, making it difficult to screen such large chemical libraries against biological targets. Today, the advancements in virtual screening have been proven to ease the traditional drug discovery process. They show the possibility of compounding new molecules outside of the large pools of bioactive compounds. Today, many other strategies are applied along with SBVS methods, like deep learning techniques or its combination techniques with other virtual screening techniques. They overcome some of the traditional difficulties of using only SBVS. Docking software is advancing day by day, coming up with new ways to cope with previous hurdles. Virtual screening is an ever-evolving field of study and research.

REFERENCES

1. Blundell T. L., Jhoti H. and Abell C.: High-throughput crystallography for lead discovery in drug design, **Nat. Rev. Drug Discov.,** 2002, 1(1), 45-54.
2. Ghosh S., Nie A., An Jing and Huang Z.: Structure-based virtual screening of chemical libraries for drug discovery, **Curr. Opin. Chem. Biol.,** 2006, 10(3), 194–202.
3. Jain A. N.: Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine, **J. Med. Chem.,** 2003, 46(4), 499–511.
4. Hongjian Li, Kam-Heung Sze, Gang Lu, Pedro J. and Ballester.: Machine-learning scoring functions for structure-based virtual screening, **Wil. Into. Rev. Comput. Mol. Sci.,** 2021, 11(1), 1–21.
5. Cheng T., Li Q., Zhou Z., Wang Y. and Bryant SH.: Structure-based virtual screening for drug discovery: a problem-centric review, **AAPS J.,** 2012, 14(1), 133-141.
6. Myers S. and Baker A.: Drug discovery – an operating model for a new era, **Nat. Biotechnol.,** 2001, 19(8), 727–730.
7. Pinzi Land Rastelli G.: Molecular docking: Shifting paradigms in drug discovery, **Int. J. Mol. Sci.,** 2019, 20(18), 4331.
8. Zhou S-F and Zhong W-Z.: Drug design and discovery: Principles and Applications, **Molecules,** 2017, 22(2), 279.
9. Mestres, J. and Knegtel.: Similarity versus docking in 3D virtual screening, Perspect. **Drug Discov. Des.,** 2000, 20, 191-207.
10. Yu W., and MacKerell AD Jr.: Computer-aided drug design methods. **Antibiotics: Methods and Protocols,** 2017,1520, 85–106.
11. Golender V. E. and Vorpagel E. R.: In Drug Design: Theory, methods, and applications, H. Kubinyi, Ed., ESCOM, Leiden, 1993, 137–149.
12. Nicholls and Anthony.: What do we know and when do we know it?, **J. Comput. Aided Mol. Des.,** 2008, 22(3–4), 239–255.
13. Bandyopadhyay S., Bagchi A. and Maulik U.: Active site driven ligand design: An evolutionary approach, **J. Bioinform. Comput. Biol.,** 2005, 3(5), 1053–1070.
14. Jorgensen W. L.: Efficient drug lead discovery and optimization, **Acc. Chem. Res.,** 2009, 42(6), 724-733.
15. Patel Y., Gillet, V. J., Bravi G. and Leach A. R.: A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP, **J. Comput.-Aided Mol. Des.,** 2002, 16(8-9), 653-681.
16. Zhong W. Z. and Zhou S.F.: Molecular science for drug development and biomedicine, **Int. J. Mol. Sci.,** 2014, 15(11), 20072–20078.
17. Otto S., Furlan R. L. and Sanders J. K.: Dynamic combinatorial chemistry, **Drug Discov. Today,** 2002, 7(2), 117-125.
18. Lyne P. D.: Structure-based virtual screening: An overview, **Drug Discov. Today,** 2002, 7(20), 1047–1055.
19. Dolata P. D., Leach A. R. and Prout K. J.: WIZARD: AI in conformational analysis, **Comput.-Aided Mol. Des.,** 1987, 1, 73-86.

20. Yasuo N. and Sekijima M.: Improved method of structure-based virtual screening via interaction-energy-based learning, **J. Chem. Inf. Model.,** 2019, 59(3), 1050–1061.
21. Clark D. E.: What has virtual screening ever done for drug discovery? **Expert Opin. Drug Discov.,** 2008, 3, 841–851.
22. Walters W. P., Stahl M. T. and Murcko M. A.: Virtual screening - an overview, **Drug Discov. Today,** 1998, 3, 160–178.
23. Leach A. R., Gillet V. J., Lewis R. A. and Taylor R.: Three-dimensional pharmacophore methods in drug discovery, **J. Med. Chem.,** 2010, 53(2), 539–558.
24. Trott O. and 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(2), 455–461.
25. Lionta E., Spyrou G., Vassilatis D. and Cournia Z.: Structure-based virtual screening for drug discovery: Principles, applications and recent advances, **Curr. Top. Med. Chem.,** 2014, 14(16), 1923–1938.
26. Villoutreix B. O., Eudes R. and Miteva M. A.: Structure-based virtual ligand screening: recent success stories, **Comb. Chem. High Throughput Screen.,** 2009, 12(10), 1000–1016.
27. Pitt W.R., Calmiano M.D., Kroeplien B., Taylor R.D., Turner J. P. and King M. A.: Structure-based virtual screening for novel ligands, **Met. Mol. Bio.,** 2013, 1008, 501-519.
28. Lavecchia A. and Di Giovanni C.: Virtual screening strategies in drug discovery: a critical review, **Curr. Med. Chem.,** 2013, 20(23), 2839-2860.
29. Reddy A. S., Patil S. P., Kumar P. P., Pradeep H. N. and Sastry G.N.: Virtual screening in drug discovery – a computational perspective, **Curr. Progr. Pept. Sci.,** 2007, 8(4), 329-351.
30. Köppen H.:Virtual screening - what does it give us?, **Curr. Opin. Drug Dis. Dev.,** 2009, 12(3), 397-407.
31. Maia E. H. B., Assis L. C., Oliveira T. A., Silva A. M., and Taranto A. G.: Structure-based virtual screening: from classical to artificial intelligence, **Front. Chem.,** 2020, 8, 343.
32. Kar S. and Roy K.: How far can virtual screening take us in drug discovery?, **Expert Opin. Drug. Discov.,** 2013, 8(3), 245–261.
33. Fossa P., Boggia R. and Mosti L.: An updated topographical model for phosphodiesterase 4 (PDE4) Catalytic Site, **Quant. Struct.-Act. Relat.,** 2001, 20(1), 17-22.
34. Murgueitio M. S., Bermudez M., Mortier J. and Wolber G.: *In silico* virtual screening approaches for anti-viral drug discovery, **Drug Discov. Today Technol.,** 2012, 9(3), 219-215.
35. Wu K. J., Lei P. M., Liu H., Wu C., Leung C. H. and Ma D. L.: Mimicking strategy for protein-protein interaction inhibitor discovery by virtual screening, **Molecules,** 2019, 24(24), 4428.
36. Yuriev E., Agostino M. and Ramsland Pa.: Challenges and advances in computational docking, **J. Mol. Recognit.,** 2010, 24 (2), 149–164.
37. Clark R. D. and Roe D. C.: Ligand- and structure-based virtual screening, Handbook of Chemoinformatics Algorithms, 2010, 145–171.
38. Güner and Osman F.: In Pharmacophore perception, development and use in drug design, **UIL Biotechnology Series,** 2000, 3-12.
39. Khedkar S., Malde A., Coutinho E., and Srivastava S.: Pharmacophore modeling in drug discovery and development: An Overview, **Curr. Med. Chem.,** 2007, 3(2), 187–197.
40. Lipinski C. A., Lombardo F., Dominy B. W. and Feeney P. J.: Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, **Adv. Drug Deliv. Rev.,** 2001, 46(1-3), 3-26.
41. Walters W. P. and Murcko M.A.: Prediction of 'drug-likeness', **Adv. Drug Deliv. Rev.,** 2002, 54(3), 255–271.
42. Hughes J. D., Blagg J., Price D. A., Bailey S., Decrescenzo G. A., Devraj R. V., Ellsworth E., Fobian Y. M., Gibbs M.E., Gilles R. W., Greene N., Huang E., Krieger-Burke T., Loesel J., Wager T., Whiteley L. and Zhang Y.: Physiochemical drug properties associated with *in vivo* toxicological outcomes, **Bioorg. Med. Chem. Lett.,** 2008, 18(17), 4872-4875.
43. Palm K., Luthman K., Ungell A. L., Strandlund G. and Artursson P.: Correlation of drug absorption with molecular surface properties, **J. Pharm. Sci.,** 1996, 85(1), 32–39.
44. Kalliokoski T., Salo H. S., Lahtela-Kakkonen M. and Poso A.: The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening, **J. Chem. Inf. Model.,** 2009, 49(12), 2742-2748.
45. Sadowski J., Rudolph C. and Gasteiger J.: The generation of 3D-models of host-guest, **Anal. Chim. Acta,** 1992, 265, 233-241.
46. Sastry G. M., Adzhigirey M., Day T., Annabhimoju R. and Sherman W.: Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments **J. Comput. Aided. Mol. Des.,** 2013, 27(3), 221-234.
47. Buch I., Giorgino T. and Fabritiis G.: Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations, **Proc. Natl. Acad. Sci. U S A,** 2011, 108(25), 10184-10189.
48. Shan Y., Kim E. T., Eastwood M. P., Dror R. O., D., Seeliger M.A. and Shaw D. E.: How does a drug molecule find its target binding site?, **J. Am. Chem.,** 2011, 133(24), 9181-9183.
49. Lazaridis T.: Inhomogeneous fluid approach to solvation thermodynamics. 2. Applications to simple fluids, **J. Phys. Chem. B.,** 1998, 102(18), 3542-3550.
50. M.I. Ecemis, J. H. Wikel, C. Bingham and Eric Bonabeau.: A drug candidate design environment using evolutionary computation, Presented at **IEEE Trans Evolutionary Computation,** 2008, 12, 591-603.
51. Soufan O., Ba-Alawi W., Magana-Mora A., Essack M. and Bajic VB.: DPubChem: A web tool for QSAR modeling and high-throughput virtual screening, **Sci. Rep.,** 2018, 8(1), 9110.
52. Ewing T. J., Makino S., Skillman A. G. and Kuntz I. D.: DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases, **J. Comput. Aided Mol. Des.,** 2001, 15, 411-428.
53. Rarey M., Kramer B., Lengauer T. and Klebe G.: A fast flexible docking method using an incremental construction algorithm, **J. Mol. Biol.,** 1996, 261, 470-489.
54. Jones G., Willett P., Glen R. C., Leach A. R. and Taylor R.D.: Development and validation of a genetic algorithm for flexible docking, **J. Mol. Biol.,** 1997, 267, 727-748.
55. Halgren T. A., Murphy R. B., Friesner R. A., Beard H. S., Frye L. L., Pollard W. T. and Banks J. L.: Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, **J. Med. Chem.,** 2004, 47, 1750-1759.
56. Goodsell D. S., Lauble H., Stout C. D. and Olson A. J.: Automated docking in crystallography: analysis of the substrates of aconitase, **Proteins,** 1993, 17(1), 1-10.
57. Guido R. V., Oliva G. and Andricopulo A.D.: Virtual screening and its integration with modern drug design technologies, **Curr. Med. Chem.,** 2008, 15(1), 37-46.
58. Halperin I., Ma B., Wolfosn H. and Nussinov R.: Principle of docking: an overview of search algorithms and guide to scoring functions, **Proteins,** 2002, 47, 409-443.

59. Wang R., Lu Y. and Wang S.: Comparative evaluation of scoring functions for molecular docking, **J. Med. Chem.,** 2003, 46, 2287-2303.
60. Neves B. J., Braga R. C., Melo-Filho C. C., Moreira-Filho J. T., Muratov E. N. and Andrade C. H.: QSAR-based virtual screening: Advances and applications in drug discovery, **Front. Pharmacol.,** 2018, 9, 1275.
61. Gilson M. K., Liu T., Baitaluk M., Nicola G., Hwang L. and Chong J.: Binding DB: A public database for medicinal chemistry, computational chemistry and systems pharmacology, **Nucleic Acids Res.,** 2015, 44, D1045-D1063.
62. Gaulton A., Hersey A., Nowotka M., Bento A. P., Chambers J., Mendez D., Mutowo P., Atkinson F., Bellis L. J., Cibrián-Uhalte E., Davies M., Dedman N., Karlsson A., Magariños M. P., Overington J. P., Papadatos G., Smit I. and Leach A.R.: The ChEMBL database in 2017, **Nucleic Acids Res.,** 2017, 45(D1), D945-D954.
63. Mayr A., Klambauer G. and Unterthiner T.: Large-scale comparison of machine learning methods for drug target prediction on ChEMBL, **Chem Sci.,** 2018, 9(24), 5441–5451.
64. H. M. Berman, K. Henrick and H. Nakamura.: Announcing the worldwide protein data bank Nature, **J. Struct. Biol.,** 2003, 10(12), 980.
65. Kim S., Chen J. and Cheng T.: PubChem: new data content and improved web interface, **Nucleic Acids Res.,** 2021, 49(D1), D1388-D1395.
66. Sterling T. and Irwin J. J.: ZINC 15--Ligand Discovery for Everyone, **J. Chem. Inf. Model,** 2015, 55(11), 2324-2337.
67. Mitsopoulos C., Micco P., Fernandez E. V., Dolciami D., Holt E., Mica I. L., Coker E. A., Tym J. E., Campbell J., Che K. H., Ozer B., Kannas C., Antolin A. A., Workman P. and Al-Lazikani B.: canSAR: update to the cancer translational research and drug discovery knowledgebase, **Nucleic Acids Res.,** 2021, 49(D1), D1074-D1082.
68. Wishart D. S., Knox C., Guo A. C., Shrivastava S., Hassanali M., Stothard P., Chang Z. and Woolsey J.: Drugbank: a comprehensive resource for in silico drug discovery and exploration, **Nucleic Acids Res.,** 2006,1(34), D668-72.
69. Chen D., Liu S. and Kingsbury.: Deep learning and alternative learning strategies for retrospective real-world clinical data, **NPJ Digit Med.,** 2019, 2(1), 43.
70. Carpenter K. A., Cohen D. S., Jarrell J. T. and Huang X:. Deep learning and virtual drug screening, **Future Med. Chem.,** 2018, 10(21), 2557–2567.
71. Dias R. and Azevedo Jr. WF.: Molecular docking algorithms, **Curr. Drug Targets,** 2008, 9, 1040–1047.
72. Vapnik V. N.: 'The Nature of Statistical Learning Theory', 2; 978-1-4757-3264-1 Springer New York, NY, USA 1995, 314.
73. Waljee A. K., Higgins P. D. R. and Singal A.G.: A primer on predictive models, **Clin. Transl. Gastroenterol.,** 2014, 5(1), 44.
74. Kimber T. B., Chen Y. and Volkamer A.: Deep learning in virtual screening: Recent applications and developments, **Int. J. Mol. Sci.,** 2021, 22(9), 4435.
75. Ramsundar B., Kearnes S., Riley P., Webster D., Konerding D. and Pande V.: Massively multitask networks for drug discovery, **arXiv Prepr,** 2015, arXiv:1502.02072v1.
76. Nair V. and Hinton G. E.: Rectified linear units improve restricted boltzmann machines. Presented at: **27th International Conference on Machine Learning,** Haifa, 2010, 21–24.
77. Pisner D. A. and Schnyer D. M.: Support vector machine. Machine Learning: Methods and Applications to Brain Disorders, Published online January 1, 2020, 101–121.
78. Calder C. A., Cressie N. Kriging and Variogram Models, 'International Encyclopedia of Human Geography', In A. Kobayashi (Ed.), International Encyclopedia of Human Geography, Elsevier, 2009, 2, 45-51.
79. Lior Rokach and Oded Maimon, 'Decision Trees'. 2005, 6, 165–192.
80. Thomas G. Dietterich: 'Ensemble Methods in Machine Learning', Multiple Classifier Systems, 978-3-540-67704-8: Springer, Berlin, Heidelberg, 2000, 1857, 1-15.
81. Xiaomeng Lu, Shuang Y., Funan P. and Weiru C.: An evolutionary algorithm for Multi-objective optimization problem based on projection plane: MOEA/P. In 2021 The 5th International conference on algorithms, computing and systems (ICACS '21). Association for computing machinery, New York, NY, USA, 98–104.
82. Sharma, M., Sindhwani, P. and Maheshwari V.: Genetic algorithm optimal approach for scheduling processes in operating system, **Int. J. Comput. Netw. Inf. Secur.,** 2014, 14(5), 91–94.
83. Bilal, Pant M., Zaheer H., Garcia-Hernandez L. and Abraham A.: Differential evolution: A review of more than two decades of research, **Eng. Appl. Artif. Intell.,** 2020, 90, 103479.
84. Grosdidier A., Zoete V. and Michielin O.: SwissDock, a protein-small molecule docking web service based on EADock DSS, **Nucleic Acids Res.,** 2011, 39, W270-7.
85. Zsoldos Z., Reid D., Simon A., Sadjad S. B. and Johnson A. P.: eHiTS: A new fast, exhaustive flexible ligand docking system, **J. Mol. Graph. Model.,** 2009, 26(1), 198–212.
86. Sinha S. M.: Simplex method, mathematical programming. 2006, 132–144.
87. Pagadala N. S., Syed K. and Tuszynski J.: Software for molecular docking: a review, **Biophys. Rev.,** 2017, 9(2), 91–102.
88. Harrison R. L.: Introduction to Monte Carlo simulation, AIP Conference Proceedings, 2009, 1204, 17–21.
89. Küçükoğlu İlker and Öztürk Nursel: Simulated Annealing Approach for Transportation Problem of Cross-docking Network Design, **Procedia Soc. Behav. Sci.,** 2014, 109, 1180-1184. 10.1016/j.sbspro.2013.12.608
90. Shin W. H., Heo L., Lee J., Ko J., Seok C. and Lee J.: LigDockCSA: protein-ligand docking using conformational space annealing, **J. Comput. Chem.,** 2011, 32(15), 3226-3232.
91. Joung I. S., Kim J. Y., Gross S. P., Joo K. and Lee J.: Conformational space annealing explained: A general optimization algorithm, with diverse applications, **Comput. Phys. Commun.,** 2018, 223, 28–33.
92. Wang Z., Sun H., Shen C., Hu X., Gao J., Li D., Cao D. and Hou T.: Combined strategies in structure-based virtual screening, **Phys. Chem. Chem. Phys.,** 2020, 22(6), 3149-3159.
93. C. F. Lagos, A. Vecchiola, F. Allende, C. A. Fuentes, J. E. Tichauer, C. Valdivia, S. Solari, C. Campino, A. Tapia-Castillo, R. Baudrand, P. Villarroel, M. Cifuentes, G. I. Owen, C. A. Carvajal and C. E. Fardella: Identification of novel 11beta-HSD1 inhibitors by combined ligand- and structure-based virtual screening, **Mol. Cell. Endocrinol.,** 2014, 384, 71–82.
94. Alcaro S., Musetti C., Distinto S., Casatti M., Zagotto G., Artese A., Parrotta L., Moraca F., Costa G., Ortuso F., MacCioni E. and Sissi C.: Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches, **J. Med. Chem.,** 2013, 56 (3), 843–855.
95. Jacobsson M., Lidén P., Stjernschantz E., Boström H. and Norinder U.: Improving Structure-Based Virtual Screening by Multivariate Analysis of Scoring Data, **J. Med. Chem.,** 2003, 46(26), 5781–5789.
96. Saudale F. Z.: Exploration of FDA-approved drug interactions with SARS-CoV2 main protease using structure-based virtual screening

and MM-GBSA binding free energy methods, AIP Conference Proceedings, 2021, 2370, 070005.

97. Kumar Y. and Singh H.: *In silico* identification and docking-based drug repurposing against the main protease of SARS-CoV-2, causative agent of COVID-19, **ChemRxiv,** 2020, 1.
98. Gahlawat A., Kumar N., Kumar R., Sandhu H., Singh I. P., Singh S., Sjöstedt A. and Garg P.: Structure-based virtual screening to discover potential lead molecules for the SARS-CoV-2 Main Protease, **J. Chem. Inf. Model.,** 2020, 60(12), 5781-5793.
99. Takeda S., Kaneko H. and Funatsu K.: Chemical-space-based *de novo* design method to generate drug-like molecules, **J. Chem. Inf. Model.,** 2016, 56, 1885–1893.
100. Yuan Y., Pei J. and Lai L.: LigBuilder V3: A multi-target *de novo* drug design approach, **Front. Chem.,** 2020, 8, 142.
101. Oladipupo T.: Types of machine learning algorithms; in Y. Zhang (ed.), New Advances in Machine Learning, IntechOpen, London. 2010
102. Lenselink E. B., Ten Dijke N. and Bongers B.: Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set, **J. Cheminform.,** 2017, 9(1), 45.
103. Gonczarek A., Tomczak J. M., Zaręba S., Kaczmar J., Dąbrowski P. and Walczak M. J.: Interaction prediction in structure-based virtual screening using deep learning, **Comput. Biol. Med.,** 2018, 100, 253–258.
104. Ballester P. J.: Selecting machine-learning scoring functions for structure-based virtual screening, **Drug Discov. Today Technol.,** 2019, 32(33), 81–87.
105. Sieg J., Flachsenberg F. and Rarey M.: In Need of Bias Control: Evaluating Chemical Data for Machine Learning in Structure-Based Virtual Screening, **J. Chem. Inf. Model.,** 2019, 59(3), 947–961.
106. Pereira J. C., Caffarena E. R. and Dos Santos C. N.: Boosting docking-based virtual screening with deep learning, **J. Chem. Inf. Model.,** 2016, 56(12), 2495–2506.
107. S. Zhang, X. Li M., Zong X., Zhu and R. Wang: Efficient kNN classification with different numbers of nearest neighbors, **IEEE Trans. Neural Netw. Learn. Syst.,** 2018, 29(5), 1774-1785.
108. Irwin J. J., Shoichet B. K. and Mysinger M. M.: Automated docking screens: a feasibility study, **J. Med. Chem.,** 2009, 52(18), 5712–5720.
109. Created with BioRender.com (Access date: 29 December 2021).
110. Fu Y., Sun Y-N., Yi K-H., Li M-Q., Cao H-F., Li J-Z. and Ye F.: Combination of virtual screening protocol by *in silico* toward the discovery of novel 4-hydroxyphenylpyruvate dioxygenase inhibitors, **Front. Chem.,** 2018, 14(6), 2296-2646.
111. Kučera, Tomáš.: Virtual screening in drug design - overview of most frequent techniques, **Medicine, Chemistry, Computer Science.,** 2016, 85(2), 50-55.