

Topical Review

VIRTUAL SCREENING STRATEGIES IN DRUG DISCOVERY – A BRIEF OVERVIEW

Pham Quoc Long^{1,2}, Pham Minh Quan^{1,2,*}

¹*Institute of Natural Products Chemistry, Vietnam Academy of Science and Technology (VAST),
18 Hoang Quoc Viet, Cau Giay, Ha Noi, Viet Nam*

²*Graduate University of Science and Technology, VAST, 18 Hoang Quoc Viet, Cau Giay,
Ha Noi, Viet Nam*

*Emails: mar.biochem@fpt.vn; pham-minh.quan@inpc.vast.vn

Received: 14 April 2021; Accepted for publication: 9 June 2021

Abstract. Computer-aided drug design has now become a compulsory tool in the drug discovery and development process. It uses computational approaches to discover, develop, and analyze drugs in order to identify potential compounds with expected biological activities. In the first part, this review provides a comprehensive introduction of the virtual screening technique, knowledge and advances in both structure-based virtual screening and ligand-based virtual screening strategies. In the second part, recent database of compounds provided worldwide and drug-like parameters which are helpful in supporting for the virtual screening process will be discussed. This information will provide a good platform to estimate the advance of applying these techniques in the new drug-lead identification and optimization.

Keywords: virtual screening, drug discovery, molecular docking, drug-likeness, ADME

Classification numbers: 1.2.1, 1.2.4.

1. INTRODUCTION

Viet Nam has a long history of traditional medicine, since long time ago, our ancestors have known how to use the surrounding plants as inexpensive but effective medicines in treating diseases [1]. Through time, the experience of using these medicinal plants has been improved, not only common colds can be treated, but even terminal illnesses such as cancer, cardiovascular diseases ... can also be treated or supported with traditional medicine. The mechanism of action behind these remedies to this day is still a question that modern science cannot fully explain.

In recent decades, millions of compounds have been isolated from plants, marine organisms and microorganisms worldwide [2, 3]. Amongst these, many compounds have potential to develop into drugs serving human life. The study of the chemical composition of these plants and animals has contributed to elucidating the ability to cure diseases of the traditional remedies [4]. In addition, it also contributes to the discovery of the main bioactive compounds that help in the treatment of diseases and avoid side effects. However, due to financial and technical

difficulties, not all the isolated compounds were tested for their therapeutic activity, even if they were tested, it was not sufficient.

Nowadays, with the robust development of information technology, complex chemical processes can be simulated with relatively high accuracy. With the advantage of saving time and money on the testing of large compound databases coupled with an increasing number of biological targets, a number of virtual screening methods and virtual biopharmaceutical assays have been developed by scientists using computerized software [5, 6].

In this review, the concept of virtual screening along with its strategies applied in drug discovery will be presented. This is followed by a brief introduction about the database of small molecules and drug-likeness parameters in supporting the virtual screening procedures.

2. OVERVIEW OF COMPUTER-AIDED DRUG DESIGN

The application of information technology in chemistry - biology - medicine research has been developed since the late 1950s in the world. In the 1960s, simple computer programs were available to simulate the NMR spectrum [7]. Using Hansch model to analyze the structure – activity relationship, multiple computers were connected to solve complex regression equations [8]. However, the actual molecules were quite complex to solve the problems of spatial structure at that time.

In the 1970s, with the improvement in processing speed and user-friendly interface, IT started to have a more significant contribution [8]. The main difficulty during this time was that there were no computer programs able to accurately describe molecules and their properties from theoretical results. This barrier was then solved by graphical programs powerful enough to represent HOMO, LUMO, MUP (molecular electrostatic potential), bipolar moment vectors off molecules [7, 8]. In the early 1990s, multi-core computers (clusters) have been developed with enough power to perform computations on chemical processes in a short time [9]. These results contributed to the increasing interest of scientists in the use of information technology in chemical research.

In traditional research of natural products chemistry in the past, compounds were mainly isolated randomly through experiments and their biological activities were then identified using simple assays such as: antibiotic, anti-inflammatory and cytotoxicity assays. Since recent decades, in developed countries, new drug generations are being discovered and developed through powerful genetic and biochemical screening tools [10]. These methods will allow the rapid and precise detection of compounds containing the desired activity in a wide variety of extracts. More importantly, these trials also provide preliminary information about the mechanism of action of the bioactive compounds in drug development, which is important for the orientation of further drug design in the later stage [5, 8].

To conduct these screening methods, it is necessary to determine the crystal structure of the targeted protein/enzyme (receptor) in which its function is responsible for the development of disease. In addition to accurate prediction and understanding of the mechanism of action of the drug, these methods also provide important knowledge for the development of new drugs when the disease has become drug-resistant [11, 12]. When the drug is used incorrectly or due to environmental conditions, chemical agents can lead to resistance due to a mutation in the structure of the DNA of the pathogenic protein. The traditional research pathway could not help to detect these changes, however, with the application of new computational technology in chemistry and biology, the problem can be solved by studying changes in the DNA structure,

changes in the interaction between receptors – bioactive molecule (ligand), thus, suggesting ideas for scientists to modify the structure of the currently used bioactive molecule to make the drug's effectiveness back. This study requires a close collaboration of researchers in three fields - biology, chemistry and medicine, in which:

- Molecular biologists are pioneers in the research and discovery of the crystal structure of the proteins/enzymes that are responsible for the emerge and development of disease.
- Chemists screen the big database of molecules based on their potential to inhibit these biological targets and then synthesize/semi-synthesize them.
- Biological experiments, pre-clinical and clinical tests in the following stage are the combination of work between chemists, biologists, doctors, and pharmacists.

In modern bioactive compounds screening models, a virtual *in silico* (virtual screening) method has emerged recently and immediately plays an important role in the drug discovery process. This method uses advances in computer science to virtual screening, describe and predict new structure of compounds that are expected to have biological activities [13, 14]. The main advantage of this method is that it minimizes the cost and time involved in drug discovery and development. It is often described as a multi-step sequential method through different screening criteria from which gradually narrows the selection of compounds with the potential to develop drugs with desired biological activities. The compounds studied do not have to be readily available, and their bioactivities are predicted virtually so it could save the expenses and material [13, 15]. Based on this principle, any compound can be assessed through virtual screening. Depending on the scale of the study, the compound database for virtual screening can reach tens of millions of compounds, and all of these compounds can be analyzed at a single screening.

Table 1. Lists of some *in silico* screening projects in the world.

Name of the project	Protein	Function of protein	No. of ligands	Ref.
<i>Malaria</i>	Plasmepsin PMII	Decomposition of Hemoglobin	1 million	[30]
	Glutathione-S-Transferase-GST	Detoxify	4.3 millions	[31]
	Dihydrofolate-Reductase-DHFR	DNA synthesis	4.3 millions	[32]
<i>Avian flu</i>	Neuraminidase	Create new virus	300 millions	[33]
<i>Diabetes</i>	Amylase/Glucoamylase	Decomposition of carbohydrates	300 millions	[34]
<i>COVID-19</i>	Main protease		1 billion	[23]
	Spike protein		10 millions	[35]

Typically, each new drug on the market costs about 800 million euros and takes 10-15 years for the research and development process [16 - 19]. Meanwhile, with modern networked computer systems (eg Grid computation) millions of structures can be virtually screened in a matter of weeks. For example, WISDOM (World-wide in Silico Docking On Malaria) is a

successful project using Grid in screening and developing anti-malarial drugs on networked machines around the world. During the three years of the project (2007 - 2010), hundreds of millions of compounds were screened and dozen of potential compounds have been tested *in vitro* followed by *in vivo* and are under clinical and preclinical testing [20 - 22]. Another typical example is the COVID-19 pandemic, since the first case appeared in December 2019 until now, there is no efficient drugs has been discovered and with the pressure to find effective drugs quickly, many research units around the world have been using virtual screening method to screen billion of compounds with the aims to repurposing drugs or find new therapeutic compounds for treatment (Table 1) [23 - 29].

The *in silico* screening methods usually uses receptor-ligand interactions to find the compounds (ligand) whose structure is best predicted to bind with the receptor (targeted protein/enzyme) here with the lowest ΔG value (Figure 1) [36]. The structure of the receptor in a three-dimensional model (3D) is determined for each study case, the ligands are developed based on the structure of chemical compounds, especially the well-known skeleton and clearly sourced.

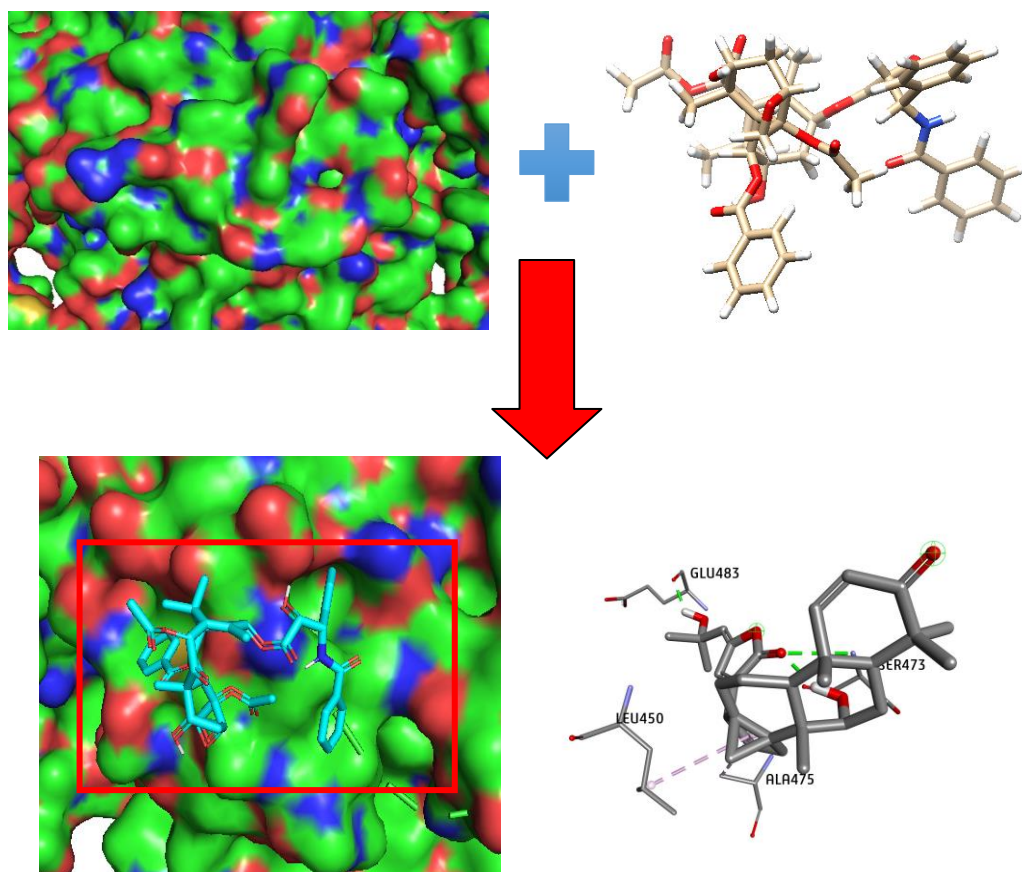


Figure 1. Diagram illustrating the interaction between protein and ligand.

The research using virtual screening method was first recognized and published in 1997 [37]. Since then, the application of this model has been increasingly popular and becomes a new

research trend in the pharmaceutical industry, along with that is the number of published studies related to this field is increasing dramatically (Table 2, Figure 2).

Table 2. Statistics of virtual screening studies published in some prestigious international scientific journals in 2000 in comparison to 2021.

Name of journal	No. of publication in 2000	No. of publication in 2021
Journal of Chemical Information and Modeling	438	2,332
Journal of Medicinal Chemistry	316	5,467
Bioorganic & Medicinal Chemistry Letters	196	4,095
Journal of Computer-Aided Molecular Design	151	1,096
Bioorganic & Medicinal Chemistry	145	3,875
ChemMedChem	92	4,065
European Journal of Medicinal Chemistry	84	4,241
Chemical Biology & Drug Design	77	1,067
ACS Chemical Biology	13	648
ChemBioChem	11	802
Nature Chemical Biology	8	351
Angewandte Chemie	2	1,032

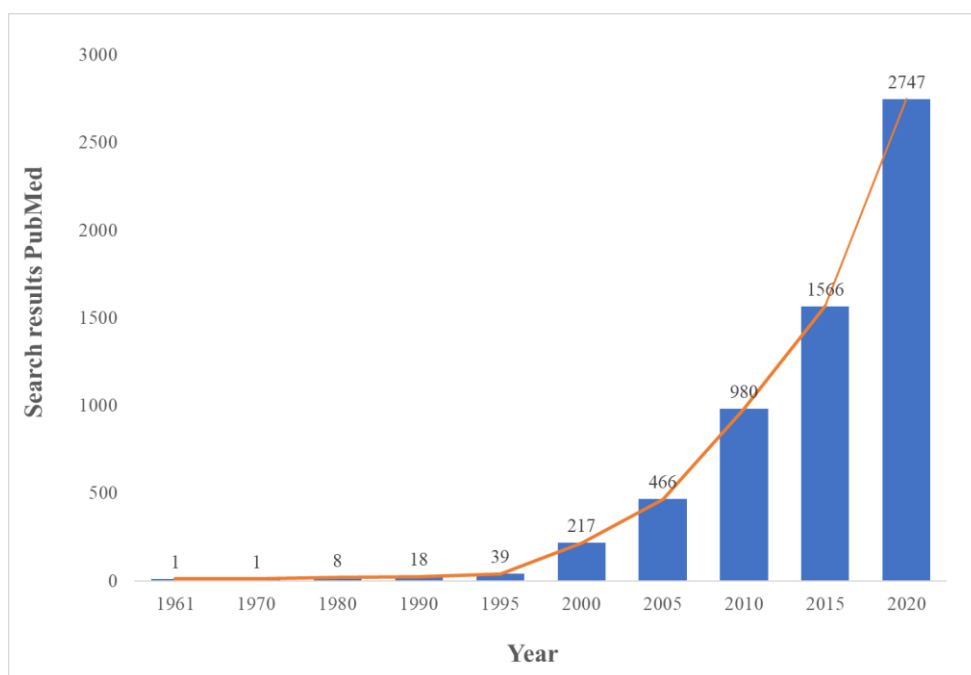


Figure 2. Total number of publications related to virtual screening since 1961 to 2020.

In Viet Nam, research directions in terms of chemistry and biological activity of natural compounds play an important role in finding sources of medicines, contributing to agricultural development and environmental protection as well as producing some functional foods. The research and application of information technology in the fields of chemical and other life sciences have been initiated and developed during recent decades. In chemistry, up to now, most of the research focuses on isolation, structure elucidation, design and synthesis of compounds, study of the relationship between structure of some series of compounds and their bioactivities. The use of information technology for *in silico* screening of new drugs recorded only few studies published. However, it is gradually becoming a new trend attracting the attention of many research groups in Viet Nam.

3. VIRTUAL SCREENING STRATEGIES

3.1. Structure-based virtual screening approach

Table 3. Examples of commonly used virtual screening softwares.

Software	Free for academia	Website
AutoDock	Yes	http://autodock.scripps.edu/
Dock	Yes	http://dock.compbio.ucsf.edu/
FlexX	No	http://www.biosolveit.de/flexx/
Glide	No	http://www.schrodinger.com/
Gold	No	http://www.ccdc.cam.ac.uk/products/life_sciences/gold/
EADock	No	http://lausanne.isb-sib.ch/agrosdid/projects/eadock/eadock_dss.php
Surflex	No	http://www.tripos.com/index.php
ICM	No	http://www.molsoft.com/docking.html
LigandFit	No	http://accelrys.com/products/discovery-studio
eHiTS	No	http://www.simbiosys.ca/ehits/index.html
SLIDE	Yes on demand	http://www.bch.msu.edu/~kuhn/software/slide/index.html
ROSETTA_DOCK	Yes on demand	http://rosettadock.graylab.jhu.edu/
Virtual Docker	No	http://www.molegro.com/mvd-product.php
Ligand_Receptor Docking	No	http://www.chemcomp.com/software-sbd.htm
FRED	Yes on demand	http://www.eyesopen.com/oedocking
ZDOCK	Yes	http://zlab.umassmed.edu/zdock/

For the structure-based virtual screening approach (SBVS), the input data included: X-ray crystal structure of the targeted protein/enzyme (receptor) and database of compounds (ligands). These compounds will be screened by docking them on the active sites of the receptor using different computation algorithms. In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form

a stable complex. Next, the docking score will be evaluated to rank the binding affinity between the ligand and the receptor. This is usually a multi-step process in which compounds are ranked and selected based on the interaction score and a number of other criteria [38]. Usually, only a handful of compounds with the highest scores are physically tested. Typically, only few compounds with the highest rank will be processed for further *in vitro* and *in vivo* experiments.

In the early years of this virtual screening method, the algorithm software used for research was called UCSF Dock, since then a lot of other softwares have been developed, for example: Gold, Dock, Glide, FlexX, AutoDock (Table 3) [39 - 63].

One of the critical steps in the SBVS model is the scoring of ligands [59, 60]. Nowadays, although prediction of binding conformation between ligands and receptor could be done with different software, the scoring and ranking compounds are still challenging. Some of the difficulties come from the fact that in some cases, molecular interactions are difficult to parameterize. Scoring function is used for the following purpose: a) to evaluate the binding pose of a compound generated by different algorithms to choose the most energetically preferred pose; b) to rank the studied compounds from which determine the most potential candidate. The scoring methods have been continuously developed over the years [61, 62], they could be grouped into three main categories: force field-based, knowledge-based and empirical [63, 64]. Some scoring models use a combination of force field-based and empirical models.

The force field scoring function [65, 66] assumes the free binding energy is the sum of molecular mechanical force fields potentials: Coulomb, Van der Waals, hydrogen bonds. Solvation [67, 68] and entropy [69] energies can also be considered. The empirical scoring function [52, 70] considers the free binding energy to be the sum of the bonds including: hydrogen bonds, hydrophobic bonds by fitting the calculated score with experimental binding affinity data for a training set of ligand-protein complexes [71]. The knowledge-based scoring function [72, 73] is based on statistical data analysis of atomic pair frequencies in ligand-protein complexes with known three-dimensional structures.

Over the past two decades, considerable efforts have been made to refine the scoring functions to accurately predict binding free energies, thus, they can be used for ranking except in the case of quantitative biological activities. However, due to the complexity of the ligand-protein binding process and the approximate calculations performed when calculating the desolvation and entropy processes, the docking score has yet to prove accurate in the binding affinity prediction [59, 74, 75]. Some methods that have been proposed to improve scoring include adding elements to calculate solvation and entropy effects [68] to give precise algorithm using high-level quantum calculations [76], target-specific scoring functions [77] and scoring simultaneously by combining multiple scoring models [78, 79]. On the other hand, it is more efficient way to use the docking score as the orientation to determine the suitability of the interaction in combination with other parameters such as tightness-of-fit by specific molecule that reflect the essence of the binding event. These parameters can be obtained by observing hydrogen bonds, which is very important parameter in docking, the spatial configuration of the π - π bond and/or the space occupancy of the hydrophobic region that pre-positions the ligand in the binding site.

Another unexploited aspect of the SBVS model is the flexibility of the target receptor [80], which will consume more computer resources and be more complex to process. In recent years, one of the biggest challenges facing many docking algorithms has been the flexible processing of target receptors. "Soft docking" (included in all docking softwares) allows small overlaps between the ligand and the receptor without large steric penalties [81]. However, this can increase the failure of outcome results because it causes more diverse structures to be bonded. It

also does not allow change of large conformation compounds, such as side-chain rotations or protein backbone motions. Some softwares such as AUTODOCK4 [46], DOCK [41], GOLD [48], EADock [49], IFREDA [51], FlexE [82] or GLIDE induced Fit [83] (Table 4) allow simulation around torsional degrees of freedom of the selected side chain using similar methods to explore the spatial conformation of flexible ligands.

Table 4. Docking programs that allow flexibility of protein.

Name of program	Ligand flexibility	Protein flexibility	Scoring function	Ref.
Autodock	Evolutionary algorithm	Flexible side chain	Force field	[46]
Dock	Incremental build	Protein side chain and flexibility	Force field or contact score	[41]
Gold	Evolutionary algorithm	Protein side chain and backbone flexibility	Empirical score	[48]
EADock	Evolutionary algorithm	Flexible side chain and backbone	Force field	[49]
ICM, IFREDA	Pseudo-Brownian sampling and local minimization	Flexible side chains	Force field and Empirical score	[51]
FlexE	Incremental build	Ensemble of protein structure	Empirical score	[82]
Glide Induced Fit	Exhaustive search	Flexible side chains	Empirical score	[83]

Currently, many other theoretical methods are being developed continuously and their applications also have great potential for virtual screening in the future. One of these theories is the Relaxed Complex Scheme (RCS). RCS uses a set of low energy structures extracted from the molecular dynamic (MD) simulation for searching in databases via molecular docking [84, 85]. It combines the advantages of the docking algorithm with the structural dynamic information obtained by MD simulation, detailed computation of the dynamic structure of both receptors and docked compounds. Longer-time of MD simulations could increase the possibility of studying the receptor's spatial configuration before docking. This model has been developed in combination with various MD software packages including: AMBER [86], NAMD [87], GROMACS [88] and AUTODOCK for ligand docking [46].

3.2. Ligand-based virtual screening approach

For ligand-based virtual screening approach (LBVS), the already known bioactive data are available in order to identify biologically active or inactive compounds and then search for more potential compounds based on structural similarity, pharmacology and other criteria.

One of the most popular models of LBVS studies is the quantitative structure–activity relationship (QSAR). The objective of QSAR is to determine the correlation between the structural/physical properties of known bioactive compound and their biological activity [89, 90]. Information on compound activity levels such as binding affinity (KD) or inhibitory concentration (IC₅₀) is essential for QSAR. Here the structure of a compound is often described by a set of structural and physical information that is considered relevant to their binding ability. The quality of the QSAR model is influenced by the compatibility of each case, structured – biological activity input data, compound description, the effect of the peripheral data, the suitability of the developed correlations, the 3D configuration, and the selection of solution directions [91].

Machine learning is increasingly being used more commonly in the algorithm for the research direction of LBVS in order to quickly and accurately establish and find the structure–activity correlation. Various technologies have been developed, each of them has its own advantages and disadvantages. Among these methods, regression models and classifications such as: Multiple Linear Regression, Nearest Neighbors, Naïve Bayesian Classification, Support Vector Machines, Neural Networks and Decision Trees have been applied successfully. These algorithms rely on certain different properties between active and inactive compounds to filter out potential candidate [92].

The efficiency of machine learning technology depends on many factors such as: diversity of data, ability to handle imbalances in data files (the number of inactive compounds is often superior to bioactive compounds) and parameters of the bioactivity of the compounds.

4. DATABASE OF SMALL MOLECULES

One of the prerequisites in traditional drug development is the identification of a specific biological target, for example, a compound that has been studied and demonstrated that its ability to interact with that target leads to the possibility to cure or improve symptoms. This first step involves the identification of potential biological targets and then validate them. Potential biological targeting requires research in the "Biological Space" (Figure 3) through human genetic sequencing, depending on high-speed sequencing technology and computer algorithms to process large amounts of output. Once a biological target has been found and validated, the next step is to identify an entity that can selectively interact with that target in a way that can induce a healing effect. According to the concept of the field of drug research, this entity is a small molecule chemical compound. Finding a compound that selectively binds to the active site of the receptor is not an easy task. To increase the chance of success, it is necessary to search in the "Chemical Space". In theory, the total number of compounds in the Chemical Zone can be estimated up to 10 million compounds [93 - 95]. This is a very large number and is beyond the capabilities of scientists currently.

Although there have been many attempts to establish such super-large databases, obtaining sufficient compounds for the "Chemical Zones" are not possible at present. In addition, only few pharmaceutical corporations are known to possess database of more than 2 million compounds. However, only a small amount of compounds in those databases are stable, water-soluble, have functional groups suitable for binding to biological targets such as proteins or nucleic acids and have sufficient structural complexity [96] to be classified in the "Medicinal Chemistry Space" region. It is argued that the compounds in the "Chemical Zone" resulting from traditional screening collection are insufficient to solve unvalidated biological targets, thus, further extensive research is needed outside of this "Chemical Zone". A feasible source for research

could be constructed from natural compound derivatives which are obtained from bacteria, plants, animals, and marine organisms through emerging technologies. These compounds form the natural product-like combinatorial libraries [3, 97].

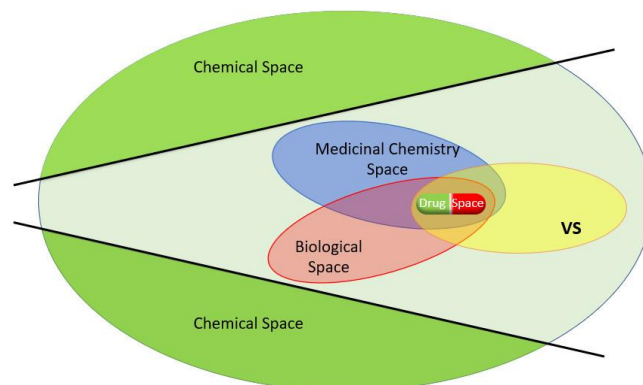


Figure 3. Model of bioactive compound search in pharmacological research.

The drug-like compound concept was devised to define the properties required for a compound to be developed successfully to drug. Over time, more stringent regulations along with procedures with drug-oriented properties have been applied to compounds during database screening. Table 5 shows some criteria defined by Hann and Oprea [98].

Table 5. Properties used for drug-like criteria by Hann and Oprea [98].

Properties	Drug-likeness
Molecular Weight (MW)	200-460
Lipophilicity (ClogP)	-4/4.2
H-bond donor (sum of NH and OH)	≤ 5
H-bond acceptor (sum of N and O)	≤ 9
Polar surface area (PSA)	$\leq 170 \text{ \AA}^2$
Number of rotatable bonds	≤ 10
CACO-2 membrane permeability	≥ 100
Solubility in water (logS)	-5/0.5
Others	No toxic and reactive fragments

There are many *in silico* tools available today that can be used to build compound databases with drug-like properties. These are the features based on empirical principles. A typical example is the Lipinski's Rule of Five [99] which states that a compound is considered non-drug-like if there are more than 5 given hydrogen bonds, more than 10 received hydrogen bonds, a molecular mass greater than 500 and the hydrophilic index was greater than 5. This principle was recently revised using the pharmacokinetic data in rats [100]. Many of the relevant rules have also been changed and the new "Rule of Three" [101] proposition defines fragments

properties with an average molecular mass ≤ 300 Da, Clog P value \leq , quantity hydrogen bonding for ≤ 3 , the number of hydrogen bonds received ≤ 3 . Recently, the Pfizer rule “Rule of 3/75” has described that compounds with Clog P ≤ 3 and surface polarization area (TPSA) > 75 are highly resistant to *in vivo* tests [102]. Table 6 provides information on database of compounds containing drug-like properties that comply with the "Rule of Three" and "Rule of Five" rules.

Table 6. Example of databases containing compounds with drug-like properties that comply with the rules of "Rule of Three" and "Rule of Five".

Database of compound	Rule	Website
Vitas-M Allium Library	3	http://www.vitasmlab.com/
TimTec Fragment-Based Library	3 and 5	http://www.timtec.net/
ChemBridge Fragment Library	3	http://www.chembridge.com/
Lifechemicals General Fragments Library	3	http://www.lifechemicals.com/
ASINEX's BioFragments	3	http://www.asinex.com/
Enamine Fragment Library	3	http://www.enamine.net/
Keyorganics BIONET Fragment Library	3	http://www.keyorganics.co.uk/
Maybridge Ro3 Library	3	http://www.maybridge.com/
Maybridge Screening Collection	5	http://www.maybridge.com/
OTAVA Fragment Library	3	http://www.otavachemicals.com/
Prestwick Fragment Library	3	http://www.prestwickchemical.com/
ChemDiv Fragment-Based Library	3	http://us.chemdiv.com/

Along with the screening method using traditional physical and chemical parameters, today there are many additional options available. For example, when preparing the database for a virtual screening process, it is necessary to remove compounds that contain undesired active groups or substituents that could interfere with and produce erroneous results [103 - 108]. Compounds are less potent if they contain factors such as hydantoin, nitro, alkyl, aniline and carbazide, which are involved in toxic metabolites. Besides, groups such as aldehydes and epoxide can be considered unsuitable for electrophoresis, the thiol group is considered to be oxidizing agents.

The screening process may be affected by high energy levels or unrealistic configurations of the compound. Some configuration building methods do not provide the lowest energy level for shaping and ranking spatial configurations, which in turn leads to configurations with high energy levels. If these configurations are not removed, it will lead to erroneous results in docking.

Compound databases are often distributed free of charge by commercial companies or research institutes. These include drugs, carbohydrates, synthetic compounds, natural compounds, etc. (Table 7) [109 - 116]. ZINC [109] is a free online database with the capacity of up to 13 million compounds in the current version with information on biological activity

(molecular weight, ClogP and number of rotational bonds). Other database files such as drug-like compounds, potency and fragments have also been introduced.

Table 7. Example of compound databases distributed by commercial companies and research institutes worldwide.

Database	Type	No. of compounds	Website
ZINC	Free	13 million	http://zinc.docking.org
ChemDB	Free	5 million	http://cdb.ics.uci.edu
eMolecules	Commercial	7 million	http://www.emolecules.com
ChemSpider	Free	26 million	http://www.chemspider.com
PubChem	Free	30 million	http://pubchem.ncbi.nlm.nih.gov
ChemBank	Free	1.2 million	http://chembank.broadinstitute.org
DrugBank	Free	4,800 drugs; 2,500 biological targets	http://www.drugbank.ca
NCI Open Database	Free	265,000	http://cactus.nci.nih.gov/ncidb2.2/
Chimiotheque Nationale	Commercial	48,370	http://chimiotheque-nationale.enscm.fr/?lang=fr
Drug Discovery Center Collection	Commercial	340,000	http://www.drugdiscovery.uc.edu/
ChEMBL	Free	1 million	http://www.ebi.ac.uk/chembl/index.php
WOMBAT	Commercial	263,000	http://www.sunsetmolecular.com
ChemBridge	Commercial	700,000	http://www.chembridge.com
Specs	Commercial	240,000	http://www.specs.net
CoCoCo	Free	7 million	http://cococo.unimore.it/tiki-index.php
Asinex	Commercial	550,000	http://www.asinex.com
Enamine	Commercial	1.7 million	http://www.enamine.net
Maybridge	Commercial	56,000	http://www.maybridge.com
ChemDiv	Commercial	1.5 million	http://www.chemdiv.com
ACD	Commercial	3.9 million	http://accelrys.com/products/databases/sourcing/available-chemicalsdirectory.html
MDDR	Commercial	150,000	http://accelrys.com/products/databases/bioactivity/mddr.html

Table 8 presents some of the commercial databases provided by other distributors [92].

Table 8. Database of small molecules provided by commercial distributors [92].

Company	Database	Website
Asinex	Antibacterials	http://www.asinex.com
SPECS	Kinase-targeted Library	http://www.specs.net/
Timtec	GPCR Ligands	http://www.timtec.net
	Kinase Modulators	
	Protease Inhibitors	
	Potassium Channels Modulators	
	Nuclear Receptors Ligands	
ChemBridge	Kinase-Biased Sets	http://www.chembridge.com
	GPCR Library	
	Channel-Biased Sets	
ChemDiv	GPCRs	http://www.chemdiv.com/main.phtml
	Kinases	
InterBioScreen	Analgesics	http://www.ibscreen.com
	Antibacterials	
	Antidiabetics	
	Cancerostatics	
	regulators	
MayBridge		http://www.maybridge.com
Key Organics	Bionet	http://www.keyorganics.ltd.uk
	Antimalarial Agents	
	Active Compounds for Cancer Research	
	Active Compounds for CNS Research	
Life Chemicals	GPCR Library	http://lifechemicals.emolecules.com/
	Kinase Library	
	Anticancer Library	

5. INTRODUCTION TO DRUG-LIKENESS PARAMETERS IN DRUG DISCOVERY

5.1. Lipinski's Rule of Five

Lipinski's Rule of Five helps distinguish between molecules that have a drug-like potential and those do not have potential as an oral drug [99, 100]. It predicts the drug-likeness of compounds based on whether or not they meet the following rules: a) molecular weight below 500 Dalton; b) High lipophilicity (expressed as LogP less than 5); c) Less than 5 hydrogen bond donors; d) Less than 10 hydrogen bond acceptors; e) Molar refractivity should be between 40-130. In which:

LogP value (partition coefficient between octanol and water) represents the ratio at equilibrium of the concentration of a compound between two phases, an oil and a liquid phase. The LogP value plays an important role in assessing the absorption, transport, distribution of substances and drug interactions with receptors [117]. This is one of the basic parameters that can be used to evaluate whether or not a compound has the potential to develop into drug. Molar refractivity is a measure of the total polarizability of a mole of a molecule [118].

In general, compounds that violate two or more criteria are predicted to be less likely to be developed as oral medications. Based on literature studies, several suggestions should be noted for drug development orientation such as: The higher the LogP value suggests that the more easily the compound disperses across the cell membrane and dissolves well in the lipid medium; Drug used orally, absorbed in the intestine should have a value of $1.35 \leq \text{LogP} \leq 1.8$; Drugs targeting the central nervous system should have a value of $\text{LogP} \sim 2$; Most metal complexes with good permeability have $\text{LogP} \leq 6$, the number of groups receiving hydrogen bonds 10 and the number of groups giving hydrogen bonds 5; Drugs used sublingually should have $\text{LogP} \geq 5$ [98 - 100, 119, 120].

5.2. Introduction to ADME

Depending on the nature of the drug and the treatment goals, people may deliver drugs into the body in different ways. Either way, drugs eventually enter the bloodstream at varying degrees to where it takes effect. ADME (Absorption, Distribution, Metabolism, Excretion) meaning absorption, dispersion, metabolism, and excretion are drug interactions with the body through the influence of molecular biology¹²¹⁻¹²³. Determining these parameters is complicated because the body is a system equipped with a myriad of mechanisms to remove foreign entity that enters inside it during metabolism or excretion. The body uses a set of enzymes with metabolic functions (the most important in the cells are the family of hemoprotein cytochromes P450 which present in the liver), transporters, excretion, the cavity will absorb and then metabolize drugs, etc.

5.2.1. Absorption

Absorption is the entry of the drug into the general circulation of the body. In order to choose the appropriate way to introduce drugs into the body, it is necessary to base on the treatment purpose, properties of the drug, dosage form, and pathological state of the patient ... The route of drug delivery into the body greatly affects the absorption and effects of the drug. There are many ways to bring drugs into the body such as gastrointestinal tract, injection route, respiratory tract and skin [124].

The biological barrier is the body's self-defense mechanism from the penetration of toxins as well as exogenous substances. Drugs are identified as exogenous substances, thus, biological barriers significantly prevent the penetration of the drug to the desired destination. Many drugs are effective in laboratory studies (*in vitro*) but have failed in animal or human trials, mostly due to the inability to penetrate the biological barrier of the body to reach the target [125]. From the perspective of the organs in the body, their biological barrier is the outermost layer of epithelial cells of the organs and the endothelial barrier (the compartment between the capillaries and endothelial cells). From a cellular perspective, the biological barrier of a cell is the cell membrane separating the intracellular and extracellular environment (the cell membrane) [125, 126].

Cell membranes (biofilms) are composed of plaques, consisting of lipid layers with two molecular rows, considered as soft structure, which is a dense liquid. In the lipid layer, there are membrane-transported albumin and lipoprotein particles, the two sides of the membrane together containing polar groups. The membrane is characterized by a rapid change in structure, albumin molecules are floating in the membrane, the spatial structure is also so altered that the membrane can form channels for small molecules, water-soluble substances, and ions to pass through to enter the cell. The membrane's barrier function is also capable of creating frameworks for receptor molecules or enzymes to attach to on its face or inside.

5.2.2. Distribution

Once absorbed, the drug enters the bloodstream to be transported to its target of action. In the blood, drugs can exist in two forms: free form and protein-associated form of plasma. Some drugs may be partially decomposed in the bloodstream [121, 126].

5.2.3. Metabolism

Metabolism is the process of transforming drugs in the body under the effect of enzymes. Through metabolism, the majority of drugs are often reduced, lost effect or toxicity [127]. Some drugs still retain the same pharmacological effect, some drugs only work after being metabolized. Therefore, metabolism is the body's detoxification process for drugs.

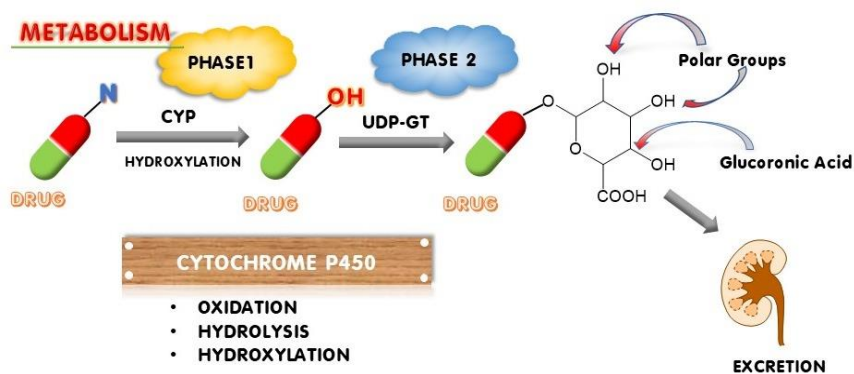


Figure 4. Stage of drug metabolism by enzyme CYP in the liver [128].

The liver is the most important organ in drug metabolism. In addition, drug metabolism can also occur in other organizations such as kidneys, intestines, lungs, blood... Oral medications must undergo initial metabolism in the liver before entering the circulatory system of

distribution in the body. This initial process of metabolism is often so strong that the drug loses its effectiveness and sometimes it is necessary to turn it into an intravenous drug to ensure its activity.

Most of the drug metabolism reactions in the body, especially in the liver involve the participation of many different enzymes. Among them, cytochrome P450 (CYP) is an enzyme that plays a major role in drug metabolism [129]. Cytochrome P450 performs drug metabolism in 3 ways: oxidation, hydrolysis, hydroxylation (step 1), then the enzyme glucuronosyltransferase (UDP-GT) will attach glucuronic acid to the drug (step 2). Glucuronic acid group contains more OH and COOH so it is easily filtered and eliminated by the kidneys (Figure 4).

For example, aspirin, after being hydrolyzed by CYP in the liver, is converted to Salicylic acid and subsequently, a glucuronic acid group is added to the UDP-GT to go to the next process (Figure 5). The result of this metabolism by CYP leads to a new salicylic acid being practically the active compounds, so aspirin is also known as a pro-drug (precursor).

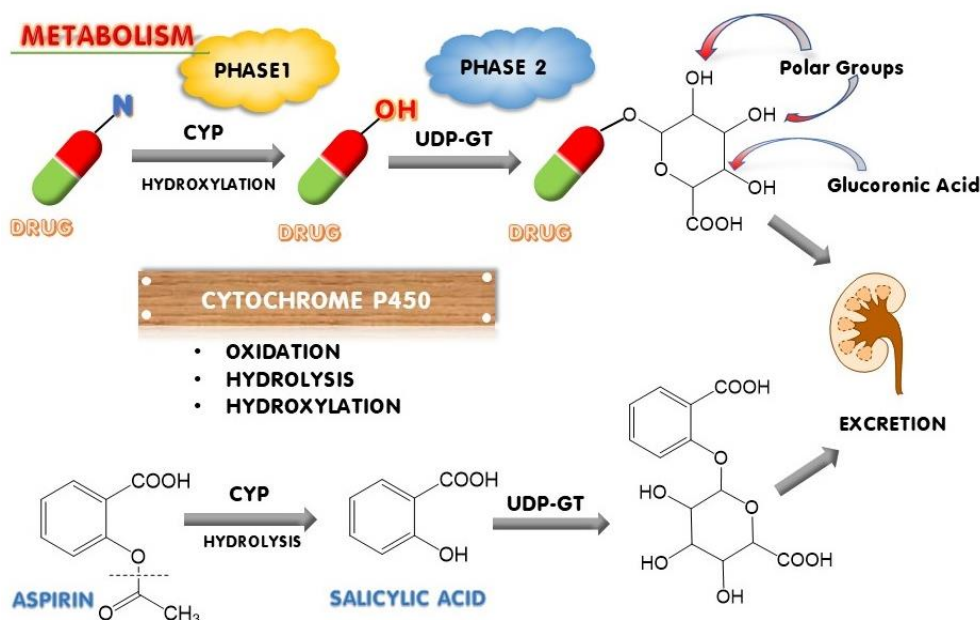


Figure 5. Metabolism of aspirin by CYP in the liver [128].

A precursor is a drug or compound that, after administration, is converted (in the body) into a drug with pharmacological activity. Inactivated precursors are pharmacologically inactive drugs that are converted into an active form in the body. Instead of using the drug directly, a corresponding precursor can be used to improve the way the drug is absorbed, distributed, metabolized, and excreted (ADME). The precursor is often designed to improve bioavailability when the drug itself is poorly absorbed from the gastrointestinal tract. A precursor that can be used to improve how a drug selectively interacts with cells or processes which are not its intended targets. This helps reduce drug side effects or unwanted effects, especially important in treatments such as chemotherapy, that can cause serious and unwanted side effects.

5.2.4. Excretion

Drug elimination is the process that leads to a decrease in drug concentration in the body. Drugs are excreted from the body mainly through the kidneys. In addition, they can also be eliminated through other routes such as gastrointestinal tract, respiratory tract, skin, sweat, breast milk or tears [130].

Some drugs can be eliminated at the same time in different ways, but normally each drug has its main elimination pathway depending on the nature and its chemical structure, on dosage form and administration route, etc.

6. CONCLUDING REMARKS

In this review, we have briefly introduced the concept of computer-aided drug design which is a new research trend worldwide in recent years. The virtual screening method has provided itself as an effective and important method for drug discovery process through two main strategies including SBVS and LBVS approaches. In addition, an overview of current databases of small molecules and information on drug-likeness parameters also presented. In conclusion, we suggest that VS methods play a pivotal role in drug discovery research and there are obvious opportunities to utilize this computational screening technology in the future.

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

Conflict statement: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

1. Do Tat L. and Nguyen Xuan D. - Native drugs of Viet Nam: which traditional and scientific approaches?, *J. Ethnopharmacol.* **32** (1-3) (1991) 51-56.
[https://doi.org/10.1016/0378-8741\(91\)90103-k](https://doi.org/10.1016/0378-8741(91)90103-k).
2. Dias D. A., Urban S. and Roessner U. - A Historical Overview of Natural Products in Drug Discovery, *Metabolites* **2** (2) (2012) 303-336.
<https://doi.org/10.3390/metabo2020303>.
3. Newman D. J., Cragg G. M. and Snader K. M. - Natural Products as Sources of New Drugs over the Period 1981-2002, *J. Nat. Prod.* **66** (7) (2003) 1022-1037.
<https://doi.org/10.1021/np030096l>.
4. Pan S. Y., Zhou S. F., Gao S. H., Yu Z. L., Zhang S. F., Tang M. K., Sun J. N., Ma D. L., Han Y. F., Fong W. F. and Ko K. M. - New Perspectives on How to Discover Drugs from Herbal Medicines: CAM's Outstanding Contribution to Modern Therapeutics, *Evid. Based Complement. Alternat. Med.* **2013** (2013) 1-25. <https://doi.org/10.1155/2013/627375>.
5. Sliwoski G., Kothiwale S., Meiler J., Lowe E. W. and Barker E. L. - Computational Methods in Drug Discovery, *Pharmacol. Rev.* **66** (1) (2013) 334-395.
<https://doi.org/10.1124/pr.112.007336>.

6. Réda C., Kaufmann E. and Delahaye-Duriez A. - Machine learning applications in drug development, *Computational and Structural Biotechnology Journal* **18** (2020) 241-252. <https://doi.org/10.1016/j.csbj.2019.12.006>.
7. Ooms F. - Molecular Modeling and Computer Aided Drug Design. Examples of their Applications in Medicinal Chemistry, *Curr. Med. Chem.* **7** (2) (2000) 141-158. <https://doi.org/10.2174/0929867003375317>.
8. Gasteiger J. - Chemoinformatics: a new field with a long tradition, *Anal. Bioanal. Chem.* **384** (1) (2005) 57-64. <https://doi.org/10.1007/s00216-005-0065-y>.
9. Downs G. M. and Barnard J. M. Clustering Methods and Their Uses in Computational Chemistry, In: *Reviews in Computational Chemistry*, Vol. 18, 2002, pp. 1-40.
10. Song C. M., Lim S. J. and Tong J. C. - Recent advances in computer-aided drug design, *Brief. Bioinform.* **10** (5) (2009) 579-591. <https://doi.org/10.1093/bib/bbp023>.
11. Yu W. and MacKerell A. D. Computer-Aided Drug Design Methods, In: *Antibiotics*, 2017, Chapter Chapter 5, pp. 85-106.
12. Singh B. K. and Surabhi S. - Computer Aided Drug Design: An Overview, *Journal of Drug Delivery and Therapeutics* **8** (5) (2018) 504-509. <https://doi.org/10.22270/jddt.v8i5.1894>.
13. da Silva Rocha S. F. L., Olanda C. G., Fokoue H. H. and Sant'Anna C. M. R. - Virtual Screening Techniques in Drug Discovery: Review and Recent Applications, *Curr. Top. Med. Chem.* **19** (19) (2019) 1751-1767. <https://doi.org/10.2174/1568026619666190816101948>.
14. Slater O. and Kontoyianni M. - The compromise of virtual screening and its impact on drug discovery, *Expert Opinion on Drug Discovery* **14** (7) (2019) 619-637. <https://doi.org/10.1080/17460441.2019.1604677>.
15. Kontoyianni M. Docking and Virtual Screening in Drug Discovery, In: *Proteomics for Drug Discovery*, 2017, Chapter Chapter 18, pp. 255-266.
16. Paul S. M., Mytelka D. S., Dunwiddie C. T., Persinger C. C., Munos B. H., Lindborg S. R. and Schacht A. L. - How to improve R&D productivity: the pharmaceutical industry's grand challenge, *Nature Reviews Drug Discovery* **9** (3) (2010) 203-214. <https://doi.org/10.1038/nrd3078>.
17. Kale V. P., Habib H., Chitren R., Patel M., Pramanik K. C., Jonnalagadda S. C., Challagundla K. and Pandey M. K. - Old drugs, new uses: Drug repurposing in hematological malignancies, *Semin. Cancer Biol.* **68** (2021) 242-248. <https://doi.org/10.1016/j.semcancer.2020.03.005>.
18. Tamimi N. A. M. and Ellis P. - Drug Development: From Concept to Marketing!, *Nephron Clinical Practice* **113** (3) (2009) c125-c131. <https://doi.org/10.1159/000232592>.
19. Moridani M. and Harirforoosh S. - Drug development and discovery: challenges and opportunities, *Drug Discovery Today* **19** (11) (2014) 1679-1681. <https://doi.org/10.1016/j.drudis.2014.06.003>.
20. Jacq N., Salzemann J., Jacq F., Legré Y., Medernach E., Montagnat J., Maaß A., Reichstadt M., Schwichtenberg H., Sridhar M., Kasam V., Zimmermann M., Hofmann M. and Breton V. - Grid-enabled Virtual Screening Against Malaria, *Journal of Grid Computing* **6** (1) (2007) 29-43. <https://doi.org/10.1007/s10723-007-9085-5>.
21. Kasam V., Salzemann J., Botha M., Dacosta A., Degliesposti G., Isea R., Kim D., Maass A., Kenyon C., Rastelli G., Hofmann-Apitius M. and Breton V. - WISDOM-II: Screening

- against multiple targets implicated in malaria using computational grid infrastructures, *Malar. J.* **8** (1) (2009). <https://doi.org/10.1186/1475-2875-8-88>.
22. de Beer T. A. P., Wells G. A., Burger P. B., Joubert F., Marechal E., Birkholtz L. and Louw A. I. - Antimalarial Drug Discovery: In Silico Structural Biology and Rational Drug Design, *Infectious Disorders - Drug Targets* **9** (3) (2009) 304-318. <https://doi.org/10.2174/1871526510909030304>.
 23. Ton A. T., Gentile F., Hsing M., Ban F. and Cherkasov A. - Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds, *Mol. Inform.* **39** (8) (2020). <https://doi.org/10.1002/minf.202000028>.
 24. Rutwick Surya U. and Praveen N. - A molecular docking study of SARS-CoV-2 main protease against phytochemicals of *Boerhavia diffusa* Linn. for novel COVID-19 drug discovery, *VirusDisease* (2021). <https://doi.org/10.1007/s13337-021-00683-6>.
 25. Soumia M., Hanane Z., Benaissa M., Younes F. Z., Chakib A., Mohammed B. and Mohamed B. - Towards potential inhibitors of COVID-19 main protease Mpro by virtual screening and molecular docking study, *Journal of Taibah University for Science* **14** (1) (2020) 1626-1636. <https://doi.org/10.1080/16583655.2020.1850002>.
 26. Talluri S. - Molecular Docking and Virtual Screening based prediction of drugs for COVID-19, *Comb. Chem. High Throughput Screen.* **23** (2020). <https://doi.org/10.2174/1386207323666200814132149>.
 27. Keretsu S., Bhujbal S. P. and Cho S. J. - Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation, *Sci. Rep.* **10** (1) (2020). <https://doi.org/10.1038/s41598-020-74468-0>.
 28. Ngo S. T., Nguyen H. M., Thuy Huong L. T., Quan P. M., Truong V. K., Tung N. T. and Vu V. V. - Assessing potential inhibitors of SARS-CoV-2 main protease from available drugs using free energy perturbation simulations, *RSC Advances* **10** (66) (2020) 40284-40290. <https://doi.org/10.1039/d0ra07352k>.
 29. Pham M. Q., Vu K. B., Han Pham T. N., Thuy Huong L. T., Tran L. H., Tung N. T., Vu V. V., Nguyen T. H. and Ngo S. T. - Rapid prediction of possible inhibitors for SARS-CoV-2 main protease using docking and FPL simulations, *RSC Advances* **10** (53) (2020) 31991-31996. <https://doi.org/10.1039/d0ra06212j>.
 30. Pranav Kumar S. K. and Kulkarni V. M. - Molecular dynamics simulations of the three dimensional model of plasmepsin II-peptidic inhibitor complexes, *Drug Des. Discov.* **17** (4) (2001) 293-313.
 31. Dong G. Q., Calhoun S., Fan H., Kalyanaraman C., Branch M. C., Mashiyama S. T., London N., Jacobson M. P., Babbitt P. C., Shoichet B. K., Armstrong R. N. and Sali A. - Prediction of Substrates for Glutathione Transferases by Covalent Docking, *J. Chem. Inf. Model.* **54** (6) (2014) 1687-1699. <https://doi.org/10.1021/ci5001554>.
 32. Choowongkamon K., Theppabutr S., Songtawee N., Day N. P. J., White N. J., Woodrow C. J. and Imwong M. - Computational analysis of binding between malarial dihydrofolate reductases and anti-folates, *Malar. J.* **9** (1) (2010). <https://doi.org/10.1186/1475-2875-9-65>.
 33. Liu Z., Zhao J., Li W., Wang X., Xu J., Xie J., Tao K., Shen L. and Zhang R. - Molecular Docking of Potential Inhibitors for Influenza H7N9, *Comput. Math. Methods Med.* **2015** (2015) 1-8. <https://doi.org/10.1155/2015/480764>.
 34. Chenafa H., Mesli F., Daoud I., Achiri R., Ghalem S. and Neghra A. - In silico design of enzyme α -amylase and α -glucosidase inhibitors using molecular docking, molecular

- dynamic, conceptual DFT investigation and pharmacophore modelling, *J. Biomol. Struct. Dyn.* (2021) 1-22. <https://doi.org/10.1080/07391102.2021.1882340>.
35. Basu A., Sarkar A. and Maulik U. - Molecular docking study of potential phytochemicals and their effects on the complex of SARS-CoV2 spike protein and human ACE2, *Sci. Rep.* **10** (1) (2020). <https://doi.org/10.1038/s41598-020-74715-4>.
 36. Du X., Li Y., Xia Y.-L., Ai S.-M., Liang J., Sang P., Ji X.-L. and Liu S.-Q. - Insights into Protein–Ligand Interactions: Mechanisms, Models, and Methods, *Int. J. Mol. Sci.* **17** (2) (2016). <https://doi.org/10.3390/ijms17020144>.
 37. Muegge I. and Oloff S. - Advances in virtual screening, *Drug Discovery Today: Technologies* **3** (4) (2006) 405-411. <https://doi.org/10.1016/j.ddtec.2006.12.002>.
 38. Reddy A. S., Pati S. P., Kumar P. P., Pradeep H. N. and Sastry G. N. - Virtual Screening in Drug Discovery - A Computational Perspective, *Current Protein & Peptide Science* **8** (4) (2007) 329-351. <https://doi.org/10.2174/138920307781369427>.
 39. Kuntz I. D., Blaney J. M., Oatley S. J., Langridge R. and Ferrin T. E. - A geometric approach to macromolecule-ligand interactions, *J. Mol. Biol.* **161** (2) (1982) 269-288. [https://doi.org/10.1016/0022-2836\(82\)90153-x](https://doi.org/10.1016/0022-2836(82)90153-x).
 40. Jones G., Willett P., Glen R. C., Leach A. R. and Taylor R. - Development and validation of a genetic algorithm for flexible docking 1 Edited by F. E. Cohen, *J. Mol. Biol.* **267** (3) (1997) 727-748. <https://doi.org/10.1006/jmbi.1996.0897>.
 41. Ewing T. J. A., 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.* **15** (5) (2001) 411-428. <https://doi.org/10.1023/a:1011115820450>.
 42. Friesner R. A., Banks J. L., Murphy R. B., Halgren T. A., Klicic J. J., Mainz D. T., Repasky M. P., Knoll E. H., Shelley M., Perry J. K., Shaw D. E., Francis P. and Shenkin P. S. - Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, *J. Med. Chem.* **47** (7) (2004) 1739-1749. <https://doi.org/10.1021/jm0306430>.
 43. 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.* **47** (7) (2004) 1750-1759. <https://doi.org/10.1021/jm030644s>.
 44. Kramer B., Rarey M. and Lengauer T. - Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking, *Proteins: Structure, Function, and Genetics* **37** (2) (1999) 228-241. [https://doi.org/10.1002/\(sici\)1097-0134\(19991101\)37:2<228::Aid-prot8>3.0.Co;2-8](https://doi.org/10.1002/(sici)1097-0134(19991101)37:2<228::Aid-prot8>3.0.Co;2-8).
 45. Buzko O. V., Bishop A. C. and Shokat K. M. - Modified AutoDock for accurate docking of protein kinase inhibitors, *J. Comput. Aided Mol. Des.* **16** (2) (2002) 113-127. <https://doi.org/10.1023/a:1016366013656>.
 46. Morris G. M., Goodsell D. S., Halliday R. S., Huey R., Hart W. E., Belew R. K. and Olson A. J. - Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, *J. Comput. Chem.* **19** (14) (1998) 1639-1662. [https://doi.org/10.1002/\(sici\)1096-987x\(19981115\)19:14<1639::Aid-jcc10>3.0.Co;2-b](https://doi.org/10.1002/(sici)1096-987x(19981115)19:14<1639::Aid-jcc10>3.0.Co;2-b).
 47. Warren G. L., Andrews C. W., Capelli A.-M., Clarke B., LaLonde J., Lambert M. H., Lindvall M., Nevins N., Semus S. F., Senger S., Tedesco G., Wall I. D., Woolven J. M., Peishoff C. E. and Head M. S. - A Critical Assessment of Docking Programs and Scoring Functions, *J. Med. Chem.* **49** (20) (2006) 5912-5931. <https://doi.org/10.1021/jm050362n>.

48. Verdonk M. L., Cole J. C., Hartshorn M. J., Murray C. W. and Taylor R. D. - Improved protein-ligand docking using GOLD, *Proteins: Structure, Function, and Bioinformatics* **52** (4) (2003) 609-623. <https://doi.org/10.1002/prot.10465>.
49. Grosdidier A., Zoete V. and Michielin O. - EADock: Docking of small molecules into protein active sites with a multiobjective evolutionary optimization, *Proteins: Structure, Function, and Bioinformatics* **67** (4) (2007) 1010-1025. <https://doi.org/10.1002/prot.21367>.
50. Jain A. N. - Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search, *J. Comput. Aided Mol. Des.* **21** (5) (2007) 281-306. <https://doi.org/10.1007/s10822-007-9114-2>.
51. Abagyan R., Totrov M. and Kuznetsov D. - ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation, *J. Comput. Chem.* **15** (5) (1994) 488-506. <https://doi.org/10.1002/jcc.540150503>.
52. Bohm H.-J. - The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure, *J. Comput. Aided Mol. Des.* **8** (3) (1994) 243-256. <https://doi.org/10.1007/bf00126743>.
53. Ravitz O., Zsoldos Z. and Simon A. - Improving molecular docking through eHiTS' tunable scoring function, *J. Comput. Aided Mol. Des.* **25** (11) (2011) 1033-1051. <https://doi.org/10.1007/s10822-011-9482-5>.
54. Zavodszky M. I., Rohatgi A., Van Voorst J. R., Yan H. and Kuhn L. A. - Scoring ligand similarity in structure-based virtual screening, *J. Mol. Recognit.* **22** (4) (2009) 280-292. <https://doi.org/10.1002/jmr.942>.
55. Gray J. J., Moughon S., Wang C., Schueler-Furman O., Kuhlman B., Rohl C. A. and Baker D. - Protein-Protein Docking with Simultaneous Optimization of Rigid-body Displacement and Side-chain Conformations, *J. Mol. Biol.* **331** (1) (2003) 281-299. [https://doi.org/10.1016/s0022-2836\(03\)00670-3](https://doi.org/10.1016/s0022-2836(03)00670-3).
56. Thomsen R. and Christensen M. H. - MolDock: A New Technique for High-Accuracy Molecular Docking, *J. Med. Chem.* **49** (11) (2006) 3315-3321. <https://doi.org/10.1021/jm051197e>.
57. McGann M. - FRED and HYBRID docking performance on standardized datasets, *J. Comput. Aided Mol. Des.* **26** (8) (2012) 897-906. <https://doi.org/10.1007/s10822-012-9584-8>.
58. Chen R., Li L. and Weng Z. - ZDOCK: An initial-stage protein-docking algorithm, *Proteins: Structure, Function, and Genetics* **52** (1) (2003) 80-87. <https://doi.org/10.1002/prot.10389>.
59. Kitchen D. B., Decornez H., Furr J. R. and Bajorath J. - Docking and scoring in virtual screening for drug discovery: methods and applications, *Nature Reviews Drug Discovery* **3** (11) (2004) 935-949. <https://doi.org/10.1038/nrd1549>.
60. Brooijmans N. and Kuntz I. D. - Molecular Recognition and Docking Algorithms, *Annu. Rev. Biophys. Biomol. Struct.* **32** (1) (2003) 335-373. <https://doi.org/10.1146/annurev.biophys.32.110601.142532>.
61. Moitessier N., Englebienne P., Lee D., Lawandi J. and Corbeil C. R. - Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go, *Br. J. Pharmacol.* **153** (S1) (2008) S7-S26. <https://doi.org/10.1038/sj.bjp.0707515>.

62. Huang S.-Y., Grinter S. Z. and Zou X. - Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions, *Phys. Chem. Chem. Phys.* **12** (40) (2010). <https://doi.org/10.1039/c0cp00151a>.
63. Krovat E., Steindl T. and Langer T. - Recent Advances in Docking and Scoring, *Current Computer Aided-Drug Design* **1** (1) (2005) 93-102. <https://doi.org/10.2174/1573409052952314>.
64. Bentham Science Publisher B. S. P. - Scoring Functions for Protein-Ligand Docking, *Current Protein & Peptide Science* **7** (5) (2006) 407-420. <https://doi.org/10.2174/138920306778559395>.
65. Meng E. C., Shoichet B. K. and Kuntz I. D. - Automated docking with grid-based energy evaluation, *J. Comput. Chem.* **13** (4) (1992) 505-524. <https://doi.org/10.1002/jcc.540130412>.
66. Goodsell D. S. and Olson A. J. - Automated docking of substrates to proteins by simulated annealing, *Proteins: Structure, Function, and Genetics* **8** (3) (1990) 195-202. <https://doi.org/10.1002/prot.340080302>.
67. Shoichet B. K., Leach A. R. and Kuntz I. D. - Ligand solvation in molecular docking, *Proteins: Structure, Function, and Genetics* **34** (1) (1999) 4-16. [https://doi.org/10.1002/\(sici\)1097-0134\(19990101\)34:1<4::Aid-prot2>3.0.Co;2-6](https://doi.org/10.1002/(sici)1097-0134(19990101)34:1<4::Aid-prot2>3.0.Co;2-6).
68. Zou X., Yaxiong and Kuntz I. D. - Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model, *J. Am. Chem. Soc.* **121** (35) (1999) 8033-8043. <https://doi.org/10.1021/ja984102p>.
69. Wang J., Morin P., Wang W. and Kollman P. A. - Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA, *J. Am. Chem. Soc.* **123** (22) (2001) 5221-5230. <https://doi.org/10.1021/ja003834q>.
70. Weng Z., Vajda S. and Delisi C. - Prediction of protein complexes using empirical free energy functions, *Protein Sci.* **5** (4) (1996) 614-626. <https://doi.org/10.1002/pro.5560050406>.
71. Schulz-Gasch T. and Stahl M. - Scoring functions for protein–ligand interactions: a critical perspective, *Drug Discovery Today: Technologies* **1** (3) (2004) 231-239. <https://doi.org/10.1016/j.ddtec.2004.08.004>.
72. Kulharia M., Goody R. S. and Jackson R. M. - Information Theory-Based Scoring Function for the Structure-Based Prediction of Protein–Ligand Binding Affinity, *J. Chem. Inf. Model.* **48** (10) (2008) 1990-1998. <https://doi.org/10.1021/ci800125k>.
73. Huang S.-Y. and Zou X. - Inclusion of Solvation and Entropy in the Knowledge-Based Scoring Function for Protein–Ligand Interactions, *J. Chem. Inf. Model.* **50** (2) (2010) 262-273. <https://doi.org/10.1021/ci9002987>.
74. Leach A. R., Shoichet B. K. and Peishoff C. E. - Prediction of Protein–Ligand Interactions. Docking and Scoring: Successes and Gaps, *J. Med. Chem.* **49** (20) (2006) 5851-5855. <https://doi.org/10.1021/jm060999m>.
75. O’Boyle N. M., Liebeschuetz J. W. and Cole J. C. - Testing Assumptions and Hypotheses for Rescoring Success in Protein–Ligand Docking, *J. Chem. Inf. Model.* **49** (8) (2009) 1871-1878. <https://doi.org/10.1021/ci900164f>.
76. Raub S., Steffen A., Kämper A. and Marian C. M. - AIScore — Chemically Diverse Empirical Scoring Function Employing Quantum Chemical Binding Energies of

- Hydrogen-Bonded Complexes, *J. Chem. Inf. Model.* **48** (7) (2008) 1492-1510. <https://doi.org/10.1021/ci7004669>.
77. Seifert M. H. J. - Robust optimization of scoring functions for a target class, *J. Comput. Aided Mol. Des.* **23** (9) (2009) 633-644. <https://doi.org/10.1007/s10822-009-9276-1>.
78. Charifson P. S., Corkery J. J., Murcko M. A. and Walters W. P. - Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins, *J. Med. Chem.* **42** (25) (1999) 5100-5109. <https://doi.org/10.1021/jm990352k>.
79. Feher M. - Consensus scoring for protein–ligand interactions, *Drug Discovery Today* **11** (9-10) (2006) 421-428. <https://doi.org/10.1016/j.drudis.2006.03.009>.
80. Evers A., Hessler G., Matter H. and Klabunde T. - Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols, *J. Med. Chem.* **48** (17) (2005) 5448-5465. <https://doi.org/10.1021/jm050090o>.
81. Jiang F. and Kim S.-H. - “Soft docking”: Matching of molecular surface cubes, *J. Mol. Biol.* **219** (1) (1991) 79-102. [https://doi.org/10.1016/0022-2836\(91\)90859-5](https://doi.org/10.1016/0022-2836(91)90859-5).
82. Perez A., Yang Z., Bahar I., Dill K. A. and MacCallum J. L. - FlexE: Using Elastic Network Models to Compare Models of Protein Structure, *J. Chem. Theory Comput.* **8** (10) (2012) 3985-3991. <https://doi.org/10.1021/ct300148f>.
83. Sherman W., Day T., Jacobson M. P., Friesner R. A. and Farid R. - Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects, *J. Med. Chem.* **49** (2) (2006) 534-553. <https://doi.org/10.1021/jm050540c>.
84. Lin J.-H., Perryman A. L., Schames J. R. and McCammon J. A. - The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme, *Biopolymers* **68** (1) (2003) 47-62. <https://doi.org/10.1002/bip.10218>.
85. Amaro R. E., Baron R. and McCammon J. A. - An improved relaxed complex scheme for receptor flexibility in computer-aided drug design, *J. Comput. Aided Mol. Des.* **22** (9) (2008) 693-705. <https://doi.org/10.1007/s10822-007-9159-2>.
86. Case D. A., Cheatham T. E., Darden T., Gohlke H., Luo R., Merz K. M., Onufriev A., Simmerling C., Wang B. and Woods R. J. - The Amber biomolecular simulation programs, *J. Comput. Chem.* **26** (16) (2005) 1668-1688. <https://doi.org/10.1002/jcc.20290>.
87. Phillips J. C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R. D., Kalé L. and Schulten K. - Scalable molecular dynamics with NAMD, *J. Comput. Chem.* **26** (16) (2005) 1781-1802. <https://doi.org/10.1002/jcc.20289>.
88. Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E. and Berendsen H. J. C. - GROMACS: Fast, flexible, and free, *J. Comput. Chem.* **26** (16) (2005) 1701-1718. <https://doi.org/10.1002/jcc.20291>.
89. Dudek A., Arodz T. and Galvez J. - Computational Methods in Developing Quantitative Structure-Activity Relationships (QSAR): A Review, *Comb. Chem. High Throughput Screen.* **9** (3) (2006) 213-228. <https://doi.org/10.2174/138620706776055539>.
90. Zou J., Xie H.-Z., Yang S.-Y., Chen J.-J., Ren J.-X. and Wei Y.-Q. - Towards more accurate pharmacophore modeling: Multicomplex-based comprehensive pharmacophore map and most-frequent-feature pharmacophore model of CDK2, *J. Mol. Graph. Model.* **27** (4) (2008) 430-438. <https://doi.org/10.1016/j.jmglm.2008.07.004>.

91. Verma J., Khedkar V. and Coutinho E. - 3D-QSAR in Drug Design - A Review, *Curr. Top. Med. Chem.* **10** (1) (2010) 95-115. <https://doi.org/10.2174/156802610790232260>.
92. Lavecchia A. and Giovanni C. - Virtual Screening Strategies in Drug Discovery: A Critical Review, *Curr. Med. Chem.* **20** (23) (2013) 2839-2860. <https://doi.org/10.2174/09298673113209990001>.
93. Dobson C. M. - Chemical space and biology, *Nature* **432** (7019) (2004) 824-828. <https://doi.org/10.1038/nature03192>.
94. Bohacek R. S., McMartin C. and Guida W. C. - The art and practice of structure-based drug design: A molecular modeling perspective, *Med. Res. Rev.* **16** (1) (1996) 3-50. [https://doi.org/10.1002/\(sici\)1098-1128\(199601\)16:1<3::Aid-med1>3.0.Co;2-6](https://doi.org/10.1002/(sici)1098-1128(199601)16:1<3::Aid-med1>3.0.Co;2-6).
95. Ertl P. - Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups, *J. Chem. Inf. Comput. Sci.* **43** (2) (2002) 374-380. <https://doi.org/10.1021/ci0255782>.
96. Selzer P., Roth H.-J., Ertl P. and Schuffenhauer A. - Complex molecules: do they add value?, *Curr. Opin. Chem. Biol.* **9** (3) (2005) 310-316. <https://doi.org/10.1016/j.cbpa.2005.04.001>.
97. Feher M. and Schmidt J. M. - Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry, *J. Chem. Inf. Comput. Sci.* **43** (1) (2002) 218-227. <https://doi.org/10.1021/ci0200467>.
98. Hann M. M. and Oprea T. I. - Pursuing the leadlikeness concept in pharmaceutical research, *Curr. Opin. Chem. Biol.* **8** (3) (2004) 255-263. <https://doi.org/10.1016/j.cbpa.2004.04.003>.
99. Lipinski C. A. - Drug-like properties and the causes of poor solubility and poor permeability, *J. Pharmacol. Toxicol. Methods* **44** (1) (2000) 235-249. [https://doi.org/10.1016/s1056-8719\(00\)00107-6](https://doi.org/10.1016/s1056-8719(00)00107-6).
100. Ridder L., Wang H., de Vlieg J. and Wagener M. - Revisiting the Rule of Five on the Basis of Pharmacokinetic Data from Rat, *ChemMedChem* **6** (11) (2011) 1967-1970. <https://doi.org/10.1002/cmdc.201100306>.
101. Congreve M., Carr R., Murray C. and Jhoti H. - A 'Rule of Three' for fragment-based lead discovery?, *Drug Discovery Today* **8** (19) (2003) 876-877. [https://doi.org/10.1016/s1359-6446\(03\)02831-9](https://doi.org/10.1016/s1359-6446(03)02831-9).
102. 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. - Physicochemical drug properties associated with in vivo toxicological outcomes, *Bioorg. Med. Chem. Lett.* **18** (17) (2008) 4872-4875. <https://doi.org/10.1016/j.bmcl.2008.07.071>.
103. Axerio-Cilies P., Castañeda I. P., Mirza A. and Reynisson J. - Investigation of the incidence of "undesirable" molecular moieties for high-throughput screening compound libraries in marketed drug compounds, *Eur. J. Med. Chem.* **44** (3) (2009) 1128-1134. <https://doi.org/10.1016/j.ejmech.2008.06.013>.
104. Benigni R. and Bossa C. - Mechanisms of Chemical Carcinogenicity and Mutagenicity: A Review with Implications for Predictive Toxicology, *Chem. Rev.* **111** (4) (2011) 2507-2536. <https://doi.org/10.1021/cr100222q>.

105. Enoch S. J., Ellison C. M., Schultz T. W. and Cronin M. T. D. - A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity, *Crit. Rev. Toxicol.* **41** (9) (2011) 783-802.
<https://doi.org/10.3109/10408444.2011.598141>.
106. Erve J. C. L. - Chemical toxicology: reactive intermediates and their role in pharmacology and toxicology, *Expert Opin. Drug Metab. Toxicol.* **2** (6) (2006) 923-946.
<https://doi.org/10.1517/17425255.2.6.923>.
107. Kazius J., McGuire R. and Bursi R. - Derivation and Validation of Toxicophores for Mutagenicity Prediction, *J. Med. Chem.* **48** (1) (2005) 312-320.
<https://doi.org/10.1021/jm040835a>.
108. Rishton G. M. - Nonleadlikeness and leadlikeness in biochemical screening, *Drug Discovery Today* **8** (2) (2003) 86-96. <https://doi.org/10.1016/s1359644602025722>.
109. Irwin J. J. and Shoichet B. K. - ZINC--a free database of commercially available compounds for virtual screening, *J. Chem. Inf. Model.* **45** (1) (2005) 177-182.
<https://doi.org/10.1021/ci049714+>.
110. Chen J. H., Linstead E., Swamidass S. J., Wang D. and Baldi P. - ChemDB update full-text search and virtual chemical space, *Bioinformatics* **23** (17) (2007) 2348-2351.
<https://doi.org/10.1093/bioinformatics/btm341>.
111. Seiler K. P., George G. A., Happ M. P., Bodycombe N. E., Carrinski H. A., Norton S., Brudz S., Sullivan J. P., Muhlich J., Serrano M., Ferraiolo P., Tolliday N. J., Schreiber S. L. and Clemons P. A. - ChemBank: a small-molecule screening and cheminformatics resource database, *Nucleic Acids Res.* **36** (Database) (2007) D351-D359.
<https://doi.org/10.1093/nar/gkm843>.
112. Wishart D. S. - DrugBank: a comprehensive resource for in silico drug discovery and exploration, *Nucleic Acids Res.* **34** (90001) (2006) D668-D672.
<https://doi.org/10.1093/nar/gkj067>.
113. Knox C., Law V., Jewison T., Liu P., Ly S., Frolkis A., Pon A., Banco K., Mak C., Neveu V., Djoumbou Y., Eisner R., Guo A. C. and Wishart D. S. - DrugBank 3.0: a comprehensive resource for 'Omics' research on drugs, *Nucleic Acids Res.* **39** (Database) (2010) D1035-D1041. <https://doi.org/10.1093/nar/gkq1126>.
114. Gaulton A., Bellis L. J., Bento A. P., Chambers J., Davies M., Hersey A., Light Y., McGlinchey S., Michalovich D., Al-Lazikani B. and Overington J. P. - ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.* **40** (D1) (2011) D1100-D1107. <https://doi.org/10.1093/nar/gkr777>.
115. Olah M., Mracec M., Ostopovici L., Rad R., Bora A., Hadaruga N., Olah I., Banda M., Simon Z., Mracec M. and Oprea T. I. WOMBAT: World of Molecular Bioactivity. In *Chemoinformatics in Drug Discovery*; 2005, pp 221-239.
116. Del Rio A., Barbosa A. J. M., Caporuscio F. and Mangiatordi G. F. - CoCoCo: a free suite of multiconformational chemical databases for high-throughput virtual screening purposes, *Mol. Biosyst.* **6** (11) (2010). <https://doi.org/10.1039/c0mb00039f>.
117. Amézqueta S., Subirats X., Fuguet E., Rosés M. and Ràfols C. Octanol-Water Partition Constant. In *Liquid-Phase Extraction*; 2020, pp 183-208.
118. Le Fèvre R. J. W. Molecular Refractivity and Polarizability. In; 1965, pp 1-90.
119. Price D. A., Blagg J., Jones L., Greene N. and Wager T. - Physicochemical drug properties associated within vivotoxicological outcomes: a review, *Expert Opin. Drug Metab. Toxicol.* **5** (8) (2009) 921-931. <https://doi.org/10.1517/17425250903042318>.

120. Wenlock M. C. - Designing safer oral drugs, *MedChemComm* **8** (3) (2017) 571-577. <https://doi.org/10.1039/c6md00706f>.
121. Ekins S. and Williams A. J. - Precompetitive preclinical ADME/Tox data: set it free on the web to facilitate computational model building and assist drug development, *Lab Chip* **10** (1) (2010) 13-22. <https://doi.org/10.1039/b917760b>.
122. Ertl P. and Jelfs S. - Designing Drugs on the Internet? Free Web Tools and Services Supporting Medicinal Chemistry, *Curr. Top. Med. Chem.* **7** (15) (2007) 1491-1501. <https://doi.org/10.2174/156802607782194707>.
123. Richard A. M., Gold L. S. and Nicklaus M. C. - Chemical structure indexing of toxicity data on the internet: moving toward a flat world, *Curr Opin Drug Discov Devel* **9** (3) (2006) 314-325.
124. Szakács G., Váradi A., Özvegy-Laczka C. and Sarkadi B. - The role of ABC transporters in drug absorption, distribution, metabolism, excretion and toxicity (ADME-Tox), *Drug Discovery Today* **13** (9-10) (2008) 379-393. <https://doi.org/10.1016/j.drudis.2007.12.010>.
125. Tsaïoun K. - Evidence-based absorption, distribution, metabolism, excretion (ADME) and its interplay with alternative toxicity methods, *Altex* (2016) 343-358. <https://doi.org/10.14573/altex.1610101>.
126. Li A. P. - Screening for human ADME/Tox drug properties in drug discovery, *Drug Discovery Today* **6** (7) (2001) 357-366. [https://doi.org/10.1016/s1359-6446\(01\)01712-3](https://doi.org/10.1016/s1359-6446(01)01712-3).
127. Guengerich F. P. - Cytochrome P450s and other enzymes in drug metabolism and toxicity, *The AAPS Journal* **8** (1) (2006) E101-E111. <https://doi.org/10.1208/aapsj080112>.
128. <http://www.handwrittentutorials.com>.
129. Zanger U. M. and Schwab M. - Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation, *Pharmacol. Ther.* **138** (1) (2013) 103-141. <https://doi.org/10.1016/j.pharmthera.2012.12.007>.
130. Caldwell J., Gardner I. and Swales N. - An Introduction to Drug Disposition: The Basic Principles of Absorption, Distribution, Metabolism, and Excretion, *Toxicol. Pathol.* **23** (2) (2016) 102-114. <https://doi.org/10.1177/019262339502300202>.