JBUON 2016; 21(6): 1337-1358
ISSN: 1107-0625, online ISSN: 2241-6293 • www.jbuon.com
E-mail: editorial_office@jbuon.com

SPECIAL ARTICLE

Overview on the current status on virtual high-throughput screening and combinatorial chemistry approaches in multi-target anticancer drug discovery; Part II

George D. Geromichalos[1], Constantinos E. Alifieris[2], Elena G. Geromichalou[2], Dimitrios T. Trafalis[2]

[1]Department of Cell Culture–Molecular Modeling and Drug Design, Symeonidion Research Center, Theagenion Cancer Hospital, Thessaloniki, Greece; [2]Laboratory of Pharmacology, Medical School, National and Kapodistrian University of Athens, Athens, Greece

Summary

Conventional drug design embraces the "one gene, one drug, one disease" philosophy. Nowadays, new generation of anticancer drugs, able to inhibit more than one pathway, is believed to play a major role in contemporary anticancer drug research. In this way, polypharmacology, focusing on multi-target drugs, has emerged as a new paradigm in drug discovery. A number of recent successful drugs have in part or in whole emerged from a structure-based research approach. Many advances including crystallography and informatics are behind these successes. In this part II we will review the role and methodology of ligand-, structure- and fragment-based computer-aided drug design computer aided drug desing (CADD), virtual high throughput screening (vHTS), de novo drug design, fragment-based design and structure-based molecular docking, homology modeling, combinatorial chemistry and library design, pharmacophore model chemistry and informatics in modern drug discovery.

Key words: *combinatorial chemistry, computational molecular docking, computer aided drug design, multi-target drug discovery, signaling networks, virtual high-throughput screening*

Introduction

Computer-aided drug design uses computational chemistry to discover, enhance, or study drugs and related biologically active molecules. The most fundamental goal is to predict whether a given molecule will bind to a target and, if so, how strongly. Molecular mechanics or molecular dynamics (MD) are most often used to predict the conformation of the small molecule and to model conformational changes in the biological target that may occur when the small molecule binds to it. CADD has become an integral part of drug discovery and development efforts in the pharmaceutical and biotechnology industry. QSAR techniques (Quantitative Structure Activity Relationship) have been used for this purpose for over 50 years [1]. The modeling of protein flexibility requires computationally intensive MD simulations. However, it is impractical to apply MD simulation to the whole structural proteome.

In this second part we will review the role of CADD, vHTS, homology modeling, combinatorial chemistry, pharmacophore chemistry and informatics in modern drug discovery.

1. Computer aided drug design (CADD)

Ideally the computational method should be able to predict affinity before a compound is syn-

Correspondence to: George D. Geromichalos, PhD. Department of Cell Culture-Molecular Modeling and Drug Design, Symeonidion Research Center, Theagenion Cancer Hospital, 2 Al. Symeonidi str., 54007 Thessaloniki, Greece. Tel: +30 231 0898237, Fax: +30 321 0845514, E-mail: geromchem@yahoo.gr

thesized and hence in theory only one compound needs to be synthesized. The reality however is that, athough present computational methods have accelerated discovery of novel molecules by reducing the number of iterations required, they are still imperfect, and it still takes several iterations of design, synthesis, and testing before an optimal molecule is discovered and provide at best only qualitatively accurate estimates of affinity.

One approach to CADD is to pre-filter the structural proteome to find the most likely cases to apply MD. Xie et al have undertaken a human structural proteome wide ligand binding site comparison using previously developed algorithms and added intensive binding free energy calculations, based on protein-ligand docking, MD simulation and MM/GBSA free energy calculations [2]. Semi-empirical, *ab initio quantum* chemistry methods, or density functional theory are often used to provide optimized parameters for the molecular mechanics calculations and also provide an estimate of the electronic properties (electrostatic potential, polarizability, etc.) of the drug candidate that will influence binding affinity.

Molecular mechanics methods may also be used to provide semi-quantitative prediction of the binding affinity. Also, knowledge-based scoring function may be used to provide binding affinity estimates. These methods use linear regression, machine learning, neural nets or other statistical techniques to derive predictive binding affinity equations by fitting experimental affinities to computationally derived interaction energies between the small molecule and the target [3].

Drug design with the help of computers may be used at any of the following stages of drug discovery: (1) hit identification using virtual screening (structure- or ligand-based design); (2) hit-to-lead optimization of affinity and selectivity (structure-based design, QSAR, etc.); and (3) lead optimization of other pharmaceutical properties while maintaining affinity (Figure 1 & 2).

1.1 In silico drug design

In silico methods with the necessary bioinformatics tools can assist in the identification of drug targets, in analyzing target structures for possible binding/active sites, generating candidate molecules, checking for their drug likeness, docking these molecules with the target, ranking them according to their binding affinities, and further optimizing the molecules to improve binding characteristics [4]. This is simplified with high performance computing and data management software. Multidisciplinary research teams using complementary experimental and informatics techniques increase the chances of success more quickly and at a lower cost [5]. Major roles of computation in drug discovery are: (1) virtual screening & *de novo* design, (2) *in silico* ADME/T prediction and (3) advanced methods for determining protein-ligand binding [6].

1.2 Computer-aided drug design (CADD) methods: Ligand-based & Structure-based

Computational methods comprise virtual screening of available chemical databases and *de novo* drug design. The two approaches assist chemogenomics in vHTS by aiming to select a set of molecules that are predicted to exhibit a biological activity on a given target. The estimation of the activity, usually by the mean of a score, is addressed by two major types of drug design. The first is referred to as ligand-based drug design and the second, structure-based drug design, depending on the data on known ligands and the availability of the 3D experimental structure of the target [7].

Ligand-based drug design (LBDD; indirect drug design) relies on knowledge of other molecules that bind to the biological target of interest. A candidate ligand can then be compared to the pharmacophore model to determine whether it is compatible and likely to bind [8]. eHiTS LASSO is a LBDD filter, an accurate fragment-based docking program for both virtual screening and binding pose prediction of ligands that uses the chemical features of a ligand surface to create a pseudo-pharmacophore for rapid screening of large databases. The eHiTS integrated Chemical Visualizer (CheVi) is an advanced visualization package that runs LASSO and is specifically designed to help users analyze how ligands interact with receptors [9]. Alternatively, a QSAR can be used from which a correlation between calculated properties of molecules and their experimentally determined biological activity can be derived [10].

Structure-based drug design (SBDD; direct drug design) relies on knowledge of the 3D structure of the biological target obtained through methods such as x-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy [11]. SBDD virtual screening involves molecular docking of candidate ligands into a protein target followed by applying a scoring function to estimate the likelihood that the ligand will bind to the protein with high affinity [12]. If an experimental

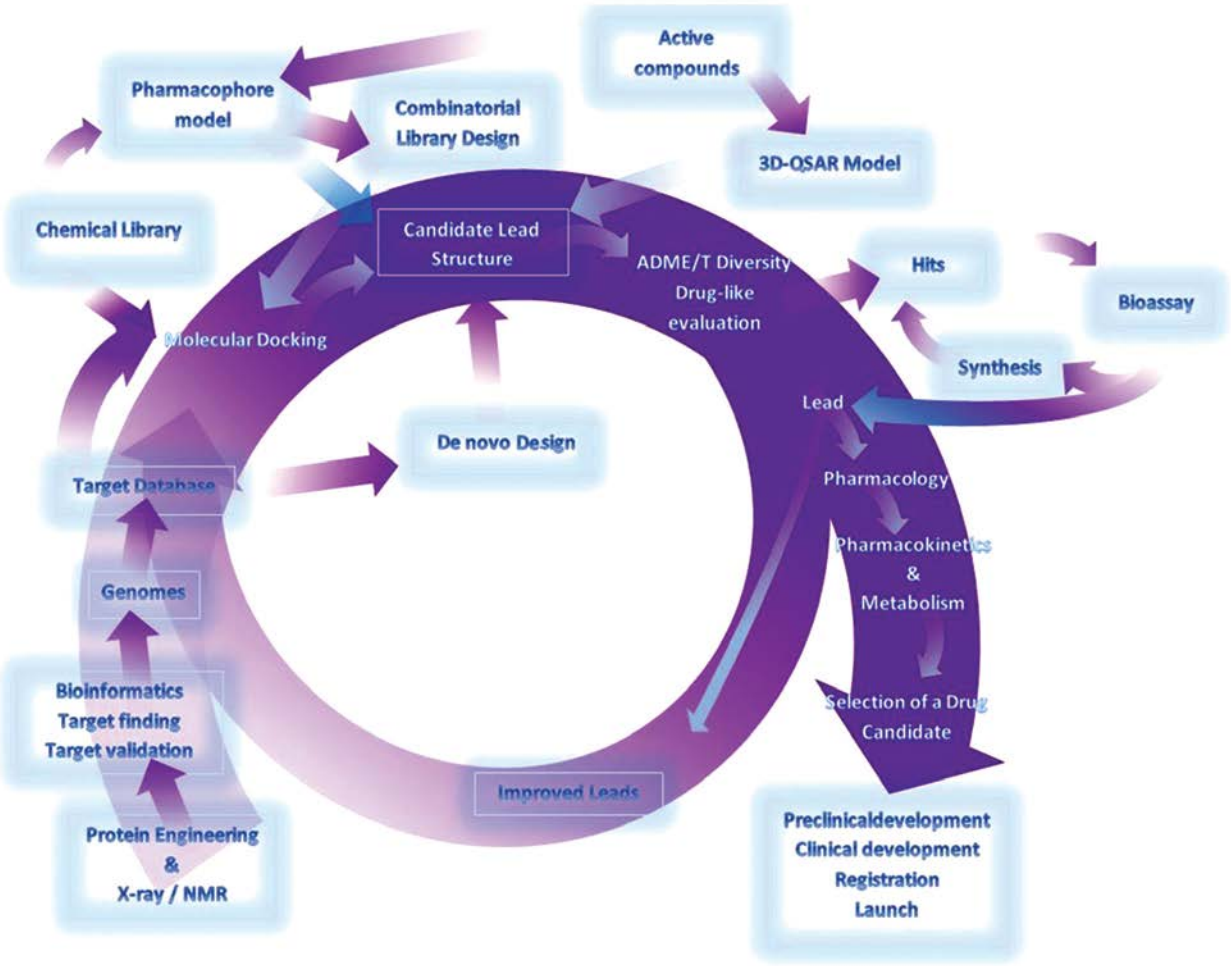

Figure 1. Flow chart of rational designing and reverse designing plus *in silico* virtual high throughput screening techniques used in drug designing method. Flow begins from genes towards drugs with the aid of protein engineering, bioinformatics and chemoinformatics (computational chemistry and modeling).

structure of a target is not available, it may be possible to create a homology model of the target based on the experimental structure of a related protein that is predicting affinity and selectivity using interactive graphics and scientist intuition or using automated computational procedures.

1.2.1 Data searching, "Building" Ligands and application of quantum mechanics in SBDD

SBDD can be divided into database searching and ligand formation. Database searching involves "finding" ligands for a given receptor by screening a large number of potential ligand molecules to find those fitting the binding pocket of the receptor. The key advantage of database searching is that it saves synthetic effort to obtain new lead compounds [13]. In "building" ligands method, ligand molecules are built up within the constraints of the binding pocket by assembling small pieces (atoms or molecular fragments) in a stepwise manner. The advantage of such a method is that novel structures not contained in any database can be suggested [14].

For SBDD, several post-screening analysis methods focusing on protein-ligand interaction have been developed for mining potential candidates: (1) consensus scoring [15]: selecting candidates by multiple scoring functions which may lose the relationship between protein-ligand structural information and scoring criterion; (2) geometric analysis: comparing protein-ligand interactions by visually inspecting individual structures [16] and (3) cluster analysis; according to protein-ligand 3D information which needs meaningful representation of protein-ligand interactions [17]. The crystal structure of a ligand bound to a protein provides a detailed insight into the interactions made between the protein and the ligand. Structure design can be used to identify where the ligand can be changed to modulate the physicochemical and ADME properties of the compound, by showing which parts of the compound are important to affinity and which parts can be altered without affecting the binding.

The last decade has seen great advances in the use of quantum mechanics (QM) to solve bio-

logical problems of pharmaceutical relevance [18]. Linear scaling QM methods have been better used in academic environments as industrial applications are sometimes difficult to access. Recently, QuantumBio (http://www.quantumbioinc.com/) has bridged that gap through the development of several QM-based interaction profiling tools specifically tailored to the SBDD process [19]. When plugged into the Molecular Operating Environment (MOE, http://www.chemcomp.com/software.htm), these tools – including scoring, pair-wise interaction energy decomposition, and QSAR – become better integrated in the field. Three major MOE svl plugins have been developed: MOE/QMScore, MOE/NMRScore, and MOE/QM-PWD. With these tools it is feasible to prepare any number of QM simulations using the MOE graphical user interface (GUI), execute the simulations in parallel using MOE's message passing infrastructure, and finally import the results back into the MOE GUI for further analysis. These QM simulations have been carried out for a series of protein kinase B inhibitors derived from fragment-based drug design (FBDD) and structure-based drug design (SBDD) [20]. Seven scoring functions were constructed based on a mixture of several quantum-and molecular-mechanical methods. The optimal models obtained by statistical analysis of the aligned poses provide residue-based contributions to the overall binding affinity, and these results are processed using both native MOE analytical methodologies and customized widgets including the QM-PWD Interaction Energy (IE) Map, Structure activity relationship (SAR) Map, and results tables. The IE map highlights the most important residues for ligand binding, while the SAR Map highlights residues that are most critical to discriminating between more and less potent ligands. Taken together the IE and SAR Maps provide useful insights into drug design that would be difficult to garner in any other way [19,20].

1.2.2 Active site directed drug design - active site identification

Active site identification analyzes a specific protein to find its binding pocket, finds key interaction sites within the pocket and then prepares the necessary data for ligand fragment link. The basic inputs for this step are the 3D structure of the protein, a pre-docked ligand in PDB format, as well as their atomic properties. Both ligand and protein atoms need to be classified and their atomic properties into four atomic types: hydrophobic, H-bond donor, H-bond acceptor and polar. The space inside the ligand binding region would be studied with virtual probe atoms of the four types above so the chemical environment of all spots in the ligand binding region can be known and correct chemical fragments can be used [21].

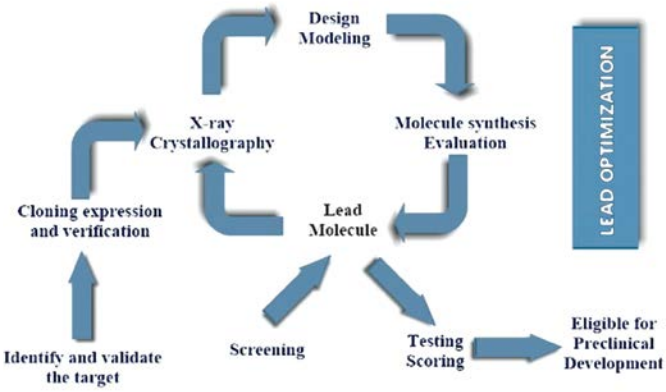

Figure 2. Typical sequence of lead optimization.

1.3 Fragment-Based Ligand Design

Extraction of fragments from molecules and their contributions to anticancer activity through inhibition of a number of target-proteins can lead to the design of new molecular entities designed from those fragments [10]. Development into a lead structure can involve three possible scenarios: (1) to grow from these "needles" into the depths of the pocket; (2) merging multiple overlapping binders into a single potent lead; or (3) linking two or more fragments into one compound with optimized potency [22]. These tasks can now be accomplished computationally with a novel software tool, LeadIT, which was primarily designed for mixed medicinal and computational chemistry teams [23].

Synthetically accessible compounds can be generated in seconds using fragment-based design by using an indexed 3D fragment library of fragments. When we want to plant "seeds" into different regions, we need a fragment or "building block" database to choose from. The combined method of "grow" and "link" is used in order to make the construction result more reliable. Although the diversity of organic structures is infinite, the number of basic fragments is rather limited. Before the first fragment (seed) is put into the binding pocket, and other fragments can be added one by one, it is useful to identify potential problems. First, the possibility for the fragment combinations is enormous. A small perturbation of the previous fragment conformation would cause great difference in the following construc-

tion process. At the same time, in order to find the lowest binding energy on the potential energy surface (PES) between planted fragments and receptor pocket, the scoring function calculation should be done for every step of conformation change of the fragments derived from every type of possible fragments combination. Since this requires a large amount of computation, one may think using other possible strategies to let the program work more efficiently. When a ligand is inserted into the pocket site of a receptor, priority is given to these groups of the ligand that favor conformation to bind tightly. After the niche of the important blocks is constructed ("grow"), then a continuous ligand will be made in a manner that makes the rest part of the ligand having the lowest energy ("link"). This strategy effectively reduces calculation burden for fragment construction and reduces the possible combinations of fragments, which minimizes the number of possible ligands that can be derived from the program [14,24].

PASS algorithm estimates qualitative (yes/no) prediction of biological activity spectra for over 4000 biological activities and, therefore, provides the basis for the preparation of a fragment library with positive biological activity corresponding to multiple criteria [25]. The algorithm has been validated using the fractions of intermolecular interactions calculated for known inhibitors of nine enzymes extracted from the Protein Data Bank database [26].

1.4 Lead "drug-like" compounds

The phrase "drug-like" generally means molecules which contain functional groups and/or have properties consistent with the majority of known drugs. Lead structures are ligands that typically exhibit suboptimal target binding affinity. Lead structures exhibit, on average, less molecular complexity (less molecular weight, less number of rings and rotatable bonds), are less hydrophobic (lower ClogP and LogD) and have lower polarizability when compared to drugs. Leads are useful for further study when they are "drug-like" and specifically when the have: (1) relatively simple chemical features, amenable for combinatorial and medicinal chemistry optimization efforts; (2) membership to a well-established SAR series, wherein compounds with similar structures exhibit similar target binding affinity; (3) favorable patent situation; and (4) good ADME (absorption, distribution, metabolism and excretion) properties [27]. Leads discovered using virtual screening and *de novo* design methodologies need to be optimized to produce candidates with improved bioavailability and low toxicity. Studies have indicated that poor pharmacokinetics and toxicity are the most important causes of high attrition rates in drug development and it has been widely accepted that these areas should be considered as early as possible in the drug discovery process, thus improving the efficiency and cost-effectiveness of the industry [28].

1.4.1 In silico ADME/T prediction in the drug discovery process

Of all the attributes which determine a drug's ultimate *in vivo* efficacy, physicochemical behavior is perhaps the most fundamental. Historically, chemical synthetic efforts have been guided almost exclusively by SAR on intention to maximize interaction with a therapeutic target (e.g. receptor, enzyme). ADME information is critical in all phases of a fully integrated drug development program. Moreover, as solubility, pKa, and lipophilicity are so integrally linked with chemical structure, physicochemical properties in some respects can be relatively predictably manipulated by chemical modifications [29].

Both logP and pKa are the primary determinants of compound solubility and affect both gastrointestinal (GI) absorption and renal tubular re-absorption. Compounds with lipophilicity logP values between 0 and 3 or 4 (e.g. 0.5 to 2) are the most suitable candidates for passive transcellular absorption across intestinal epithelia and are likely to undergo substantial renal tubular reabsorption (depending also on urine pH) [30]. The latter effect will tend to prolong t1/2, however the greater lipophilicity and minimal renal excretion will also render the compound more susceptible to metabolism.. In contrast, more hydrophilic compounds with logP values less than 0 are likely to traverse the epithelium more slowly via paracellular channels, although it should be appreciated that molecular size is also an important determinant of transit through the narrow paracellular channels. Mathematical predictions based on a weighting of functional groups within a molecule have been performed with impressive accuracy [31].

Evaluation of drug-likeness involves prediction of ADMET properties (or ADME-Tox, incorporating the potential or real toxicity of the compound) [32] and these predictions can be attempted with various studies as (1) *in vitro–in vivo* using data obtained from tissue or recombinant material from human and pre-clinical species;

(2) inter-species, *in vivo-in vivo* using data from pre-clinical species; and (3) *in silico* or computational predictions projecting *in vitro* or *in vivo* data. *In silico* prediction of drug-likeness at an early stage involves evaluation of various ADMET properties using computational approaches like QSAR or molecular modeling. Availability of large databases of drug or drug-like molecules provides useful information

Elucidation of chemical-protein interactions (CPI) is also fundamental in target identification and drug discovery. It is time-consuming and costly to determine CPI experimentally, and computational methods will facilitate the determination of CPI [10]. As a result, the screening, design, and optimization of pharmacokinetic properties have become the bottleneck and a major challenge in drug research. In order to focus in optimal molecules saving time, costs and reducing the high-attrition rates of active compounds, powerful biological, physicochemical, and computational approaches are used to incorporate structure-permeation, structure-distribution, structure-metabolism, and structure-toxicity relations into drug-design strategies [29].

1.4.1.1 Lipinski's rule of five and improvements to the rule

Lipinski's "rule of five" is a rule of thumb to evaluate druglikeness or determine if a chemical compound has optimal physico-chemical properties that would make it a likely orally active drug in humans. The rule was formulated by Lipinski et al., based on the observation that most medication drugs are relatively small and lipophilic molecules [33,34].

The rule describes molecular properties important for a drug's pharmacokinetics in the human body, including their absorption, distribution, metabolism, and excretion ("ADME", with all parameter cut-off values containing "5"s). However, the rule does not predict if a compound is pharmacologically active. The rule is important for drug development where a pharmacologically active lead structure is optimized step-wise for increased activity and selectivity, as well as drug-like properties as described by Lipinski's rule [35]. Poor absorption or permeation is more likely when a compound fails to fulfill or violates more than one of the following criteria: (1) Not more than 5 hydrogen bond donors (oxygen or nitrogen atoms with one or more hydrogen atoms); (2) Not more than 10 hydrogen bond acceptors (estimated by the sum of oxygen and nitrogen atoms); (3) A molecular mass less than 500 daltons; and (4) The compound's lipophilicity logP not greater than 5 (or MLogP lower than 4.15)

The rule is useful but it has some limitations and there are many exceptions such as compound classes that are substrates for biological transporters and natural products: antibiotics; fungicides-protozoacides-antiseptics; vitamins and cardiac glycosides. It is too naive to detect all patterns of drugs. To evaluate druglikeness better, the rules have spawned many extensions [36]: (1) Partition coefficient logP in −0.4 to +5.6 range; (2) Molar refractivity from 40 to 130; (3) Molecular weight from 160 to 500; (4) Number of atoms from 20 to 70 (includes H-bond donors [e.g. OH's and NH's] and H-bond acceptors [e.g. N's and O's]); and (5) Polar surface area no greater than 140 Å2. Further investigations by scientists extend profiling tools to lead-like properties of compounds in the hope that a better starting point in early discovery can save time and cost. The ASDI website (http://www.frontierssi.com/), now allows using Rule-of-5 and other properties to rapidly identify compounds that may be more desirable.

A deeper understanding of the relationships between important ADME parameters and molecular structure and properties is needed to develop better *in silico* models to predict ADMET properties. ADME properties evaluated using *in silico* models are: intestinal permeability, aqueous solubility, human intestinal absorption, human oral bioavailability, active transport, efflux by P-glycoprotein, blood-brain barrier permeation, plasma protein binding, metabolic stability, interactions with cytochrome P450s and toxicity. To calculate the ADMET properties various pharmaceutical, biotech or software companies and some academic research laboratories have launched their software products like: C2-ADME (www.accelrys.com), TOPKAT (www.accelrys.com), CLOGP (www.biobyte.com), DrugMatrix (www.iconix-pharm.com), AbSolv (www.sirius-analytical.com), Bioprint (www.cerep.fr), GastroPlus (www.simulations-plus.com).

1.4.1.2 Physiologically-based pharmacokinetic (PBPK) modeling and pharmacokinetics (PK) prediction software

PBPK modeling is a mathematical modeling technique for predicting the ADME of synthetic or natural chemical substances in humans and pre-clinical animals. It provides a powerful means of integrating ADME and physicochemical screen data –either *in vitro* or *in silico*- to predict *in vivo* PK [37]. The determination of physicochemical and ADME properties during early drug discov-

ery ("early ADME data"') enables PK prediction to be performed at any stage from lead identification onwards. Sensitivity analysis - in which the effect of uncertainty in an input property on the value of an output (predicted) property is quantified - is a powerful tool for informing, and helping to direct - chemistry during lead optimization [38].

The investigation of various aspects of PK prediction may be feasible with the employment of Cloe® PK and Cloe Predict® Human Intestinal Absorption (HIA) software (https://www.cloegateway.com/). Cloe® PK is a physiologically based pharmacokinetic model for prediction of the PK of a compound in rat, mouse or humans using simple physicochemical and in vitro ADME data. Cloe Predict® HIA software assesses a compound's potential as an oral therapy by predicting absorption from the human GI tract, using Caco-2 permeability data and simple physicochemical properties. AcslX (http://www.acslx.com/solutions/pharmaceutical.shtml) is a powerful and flexible modeling tool that provides ability to perform classical PBPK and PD modeling at each stage of the drug development process and predicts drug dispersion and residual drug levels for toxicity and therapeutic threshold analysis. The future aim is to use univariate and multivariate sensitivity analyses as an aid to direct chemistry optimization.

2. Virtual high throughput screening (vHTS)

As previously mentioned, the major breakthrough in lead identification in the recent years occurred with the availability of fast and cheap computers on one hand and commercially available databases of compounds such as the public database GDB-13 which contains 970 million of drug-like molecules [39]. This resulted in virtual screening technologies using high throughput docking, homology searching and pharmacophore searches of 3D databases. The routes for the virtual screening (VS) go back to the SBDD & molecular modeling and this is perhaps the cheapest way to identify a lead and several cases have already proven successful using this technology [40].

Although combinatorial chemistry and HTS have offered medicinal chemists a much broader range of possibilities for lead discovery and optimization, the number of chemical compounds that can be reasonably synthesized ("virtual chemistry space"), is still far beyond today's capability of chemical synthesis and biological assay. The basic goal of the vHTS is the reduction of the enormous virtual chemical space of small organic molecules, screen against a specific target protein, and to synthesize a manageable number of compounds [41]. Informatics in chemistry develop and utilize various computer programs to evaluate a very large number of chemical compounds and recommend the most promising ones for bench medicinal chemists collectively known as "chemical database searching" [42].

Specific binding interactions are central to many biological processes and pathways. Similarly, most drugs act by binding specifically to a site on a target protein, thereby modulating protein activity. Over the past fifteen years, in parallel with the exponential increase in the number of available high-resolution protein structures, many computer-based VS and docking methods have emerged to manage large data of thousands of bioactive compounds ("hits") [40]. Collectively, vHTS methods include VS, *de novo* design and fragment based discovery.

2.1. Virtual screening (VS)

This approach refers to computational screening of large libraries of chemicals for compounds that complement targets of known structure which could be tested experimentally. Since the virtual screening takes place in the 3D active site of the target it is also known as "structure-based virtual screening" [43]. VS has become an integral part of the drug discovery process. Walters, et al. define VS as "automatically evaluating very large libraries of compounds" using computer programs [42]. Although filtering the entire chemical universe might sound fascinating, more practical VS scenarios focus on designing and optimizing targeted combinatorial libraries and enriching libraries of available compounds from in-house compound repositories or vendor offerings.

The aim of VS is to identify molecules of novel chemical structure that bind to the macromolecular target of interest. Thus, success of a VS is defined in terms of finding interesting new scaffolds. Interpretations of VS accuracy should therefore be considered with caution. "Low hit rates" of interesting scaffolds are preferable over "high hit rates" of already known scaffolds.

Most VS studies in the literature are retrospective. In these studies, the performance of a VS technique is measured by its ability to retrieve a small set of previously known molecules with affinity to the target of interest ("active molecules" or just "actives") from a library containing a much higher proportion of assumed inactive or decoys. By contrast, in prospective applications of VS,

the resulting hits are subjected to experimental confirmation (e.g., IC50 measurements). There is consensus that retrospective benchmarks are not good predictors of prospective performance and consequently only prospective studies constitute conclusive proof of the suitability of a technique for a particular target [44,45].

A large number of computational methods exists for VS methods; ligand-based and structure-based techniques (detailed under computer-aided drug design (CADD) in Part I published by J BUON) but which one is chosen depends on the information available and the task. Another approach to ligand-based VS is to use 2D chemical similarity analysis methods to scan a database of molecules against one or more active ligand structure [46]. Nevertheless, the VS of thousands of compounds is still time-consuming even with high-throughput methods. Alternative approaches as *de novo* drug design use search strategies to efficiently explore the chemical space.

2.2. De novo drug design

De novo design is a complementary strategy for inhibitor discovery. By using the structural features present within the protein only, new inhibitor designs can be built-up sequentially according to the requirements of the targeted binding site [14]. Therefore, *de novo* design is an important technique to use in parallel with vHTS in a particular hit identification campaign, as a good *de novo* design program will examine structural spaces larger by many orders of magnitude than that of most virtual libraries currently used for this purpose (Figure 1&2). Contrary to VSs used to mine in-house and commercial collections, *de novo* design can create molecules that do not exist in known compound databases [47].

De novo design methods are automated computational procedures that build molecules by using atoms or fragments with the aim that the resulting molecular structures would fit specified property constraints. They allow the exploration of the theoretically available chemical space, a space larger than that can be enumerated by synthesis or even by computer [48]. *De novo* design attempts to use the unliganded structure of the protein to generate novel chemical structure that can bind. There are varying algorithms, most of which depend on identifying initial putative sites of interaction that are grown into complete ligands. The *de novo* molecular design computer program SPROUT, and the vHTS program eHiTS have both been applied to a number of therapeutically attractive enzyme targets and have rapidly identified inhibitors in the micromolar range or better [49].

2.3. Fragment based discovery

Fragment based discovery is based on the premise that most ligands that bind strongly to a protein active site can be considered as a number of smaller fragments or functionalities. Fragments are identified by screening a relatively small library of molecule (400-20,000) by X-ray crystallography and NMR spectroscopy. These structures of the fragment binding to the protein can be used to design new ligands by adding functionality to the fragments or by incorporating features of the fragment onto existing ligands [22].

2.4. Shape and electrostatics in virtual screening and lead hopping

A popular approach to ligand-based VS is based on searching molecules with similar shape to that of known actives, as such molecules will fit the target's binding site and will be likely to bind the target. There are a number of prospective applications of this class of techniques in the literature [45]. Similar approaches may be utilized for two different purposes: ROCS for shape-based virtual screening and lead-hopping (http://www.eyesopen.com/rocs) and BROOD for lead-hopping and bioisostere identification using shape and electrostatics (http://www.eyesopen.com/brood).

ROCS is a powerful fast shape virtual screening tool, based on the idea that molecules have similar shape if their volumes overlay well and any volume mismatch is a measure of dissimilarity. It can rapidly identify potentially active compounds with a similar shape to a known lead compound and uses a smooth Gaussian function to represent the molecular volume, so it is possible to minimize to the best global match [50]. With ROCS it is feasible to search in shape space, using shape in VS and the visual query editor vROCS. The vROCS editor can be used to generate queries for VS or lead-hopping experiments. Using vROCS it is possible to merge multiple molecules into a single query, edit molecules in a way that separates the structure of a molecule from the idea of a query, and validate the queries that class members generate in retrospective virtual screens. With robust statistical methods this can be applied to VS experiments in order to compare generated queries and enabling to choose the best query for prospective experiments [51].

BROOD is a software application designed to help project teams in drug discovery to explore chemical and property space around their hit or lead molecule. BROOD fragment searching has multiple applications, including lead-hopping, side-chain enumeration, patent breaking, fragment merging, property manipulation, and patent protection by SAR expansion. BROOD generates analogs of the lead by replacing selected fragments in the molecule with fragments that have similar shape and electrostatics with selectively modified molecular properties. BROOD's Graphical User Interface (GUI) may be used to generate an isostere replacement query and to run that query against a large pre-generated database of fragments. BROOD fragment searching has multiple applications, including lead-hopping, side-chain enumeration, patent breaking, fragment merging, property manipulation, and patent protection by SAR expansion [52].

2.5. Structure-focused pharmacophores

VS using 3D pharmacophores has been established as an important and commonly used technique for VS in recent years. Pharmacophore-based drug design process includes pharmacophore modeling and validation; pharmacophore-based VS, virtual hits profiling and lead identification. Strategies and proven methodologies for pharmacophore modeling include common feature and 3D QSAR based pharmacophore generation as well as structure-based pharmacophore development. Pharmacophore related CADD techniques can be classified into two categories: (1) structure-based pharmacophore approaches; and (2) ligand-based pharmacophore approaches [53]. The phramacophore-based approach can be accomplished with the employment of software tool such as LigandScout [54] (http://www.inteligand.com/ligandscout/), which allows for rapid and transparent development of high-quality 3D pharmacophores and provides intuitive pharmacophore overlay and interpolation workflows based on a robust and fast chemical-feature-based alignment algorithm. The underlying methods are based on several years of experience in pharmacophore creation [55]. The full-featured 3D graphical user interface makes the exploration of the active site and pharmacophore creation within the protein-ligand complex user-friendly and transparent. Binding site analysis, pharmacophore-based alignment and the creation of shared feature models are designed to make the drug discovery workflow more efficient. Model validation and predictivity assessment will include Receiver Operating Characteristic (ROC) curves and enrichment analyses [56].

2.6. Databases and compound libraries

It is vital that we access new compounds as quickly as possible. A "compound library" is a collection of compounds, just as we use 'library' for a collection of books. New compounds may also be made "in-house" but nowadays specialist chemical companies are often contracted to simply make new chemical entities (NCEs) for big pharmaceutical companies. VS is used for selecting potentially active compounds from databases of compounds available either in-house or from a vendor. Because VS is not accurate enough to identify only active compounds as hits, it is less risky to screen databases with existing compounds rather than synthesize a new library. Nevertheless, virtual libraries that can be synthesized through combinatorial chemistry and/or rapid analoging can easily be generated using *in silico* methods. These libraries are more often generated for lead optimization and synthesis prioritization [57,58].

Among all possible libraries, natural product collections arguably represent the highest degree of chemical diversity. Natural products are often extolled as sources of drug leads, however, frequently occurring natural product motifs are seldom found in drugs. Therefore, although diversity is critical, it is highly desirable to design a focused virtual library that contains synthesizable and drug-like compounds rather than one that maximally samples diversity space [42].

The applications of databases include *in silico* drug target discovery, drug design, drug docking or screening, drug metabolism prediction, drug interaction prediction, and general pharmaceutical education. Among the most frequently used databases in drug discovery process are: DrugBank (http://www.drugbank.ca/) is a freely available web-enable unique bioinformatics/chemoinformatics resource that combines detailed drug data with comprehensive drug target and drug-action information. It is fully searchable, and its applications include *in silico* drug target discovery, drug design, drug docking or screening, drug metabolism prediction, drug interaction prediction and general pharmaceutical education [59]; The Pharmacogenomics Knowledge Base (PharmGKB; http://www.pharmgkb.org/) is a database that includes literature annotations, primary data sets, PK and PD pathways, and expert-generated summaries of PK/PD relationships between drugs, diseases/phenotypes and genes

thus defining drug associated genes [60]; Protein Interaction Network Analysis (PINA; http://cbg.garvan.unsw.edu.au/pina/) is an integrated platform of protein-protein interaction (PPI) data extracted from six public databases: IntAct, MINT, BioGRID, DIP, HPRD and MIPS MPact. PINA includes self-interactions, the interactions by prediction methods and the interactions between human proteins and proteins from other species [61]; Two commercial databases, WOMBAT (version 2009) and WOMBAT-PK (version 2008) [62]; Comprehensive Medicinal Chemistry (CMC) [63]; UniProt ID mapping tool (http://www.uniprot.org/) and Entrez GENE info file (ftp://ftp.ncbi.nih.gov/gene/DATA/GENE_INFO/); The PubChem bioassay of the the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/pcassay); ZINC (http://zinc.docking.org/choose.shtml); ACD and Cambridge Crystallographic Database [64]; ChemBridge compounds (http://www.chembridge.com/); National Library of Medicine's ChemIDplus (http://chem.sis.nlm.nih.gov/chemidplus/); Chemical Abstracts Service (CAS; https://www.cas.org/) and Organic Synthesis databases (http://www.orgsyn.org/); JCHEM.INFO, free database on organic and inorganic chemicals and compound physical data; ChemSpider (http://www.chemspider.com/) with free access to > 10 million chemical structures, physical property data and systematic identifiers; Maybridge Screening collection consists of over 56,000 individually designed organic compounds, produced by innovative synthetic techniques ((http://www.maybridge.com/); MACCS-II Drug Data Report (MDDR) [65]; World Drug Index (WDI by ThomsonReuters.com), a computerized database of about 50000 drugs.

From the above databases, drug related information can be extracted and all the data may be organized into the open source MySQL relational database management system to facilitate cross-database searching (Structured Query Language; by Oracle, USA; http://www.mysql.com/). Each data set can be managed in the MySQL database as tables that store specific information, whereas keys (e.g., DrugBank_ID, GeneBank ID and PharmGKB_ID) may be extensively used for relational linking. The combinatorial data query and data processing can be implemented in PHP on drug-target interactome (DTome) server and the system interface can be implemented using HTML and Javascript.

2.7. Structure based molecular docking

Once molecules have been designed for improved drug activity they must be tested. The first round of testing is in the computer. The most heavily used tool for this testing is molecular docking which is a virtual screening method that uses geometrical matching to dock small molecules to the target site 3D structure followed by the analysis of binding feasibility by consideration of chemical complementary and molecular interaction energies. *De novo* design of small molecules is clearly becoming a valuable and integral part of the drug discovery, and the results on successful generation of leads from structure-based drug design (SBDD) and medicinal chemistry efforts are emerging [66].

Docking is an automated computer algorithm that determines how a compound will bind in the active site of a protein. Docking programs capture all of the energy contributions that are used in molecular mechanics and some have the ability to allow for solvation, induced fit, entropy corrections and maybe even inhibitors being covalently bound to the active site. The docking results show the optimal molecular position in the active site, and report a quantitative binding energy. High throughput screening docking and scoring techniques can be applied to computationally screening a database of hundreds of thousands or even millions of compounds against a number of target proteins. Computational methods that predict the 3D structure of a protein ligand complex are often referred to as "molecular docking approaches" and have become a crucial component of many drug discovery programs, from hit identification to lead optimization and beyond, and approaches such as ligand or structure-based virtual screening techniques are widely used (Figures 3-6) [67,68].

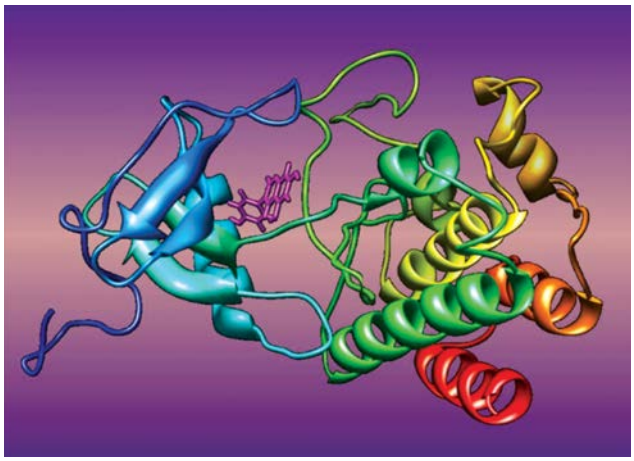

Figure 3. A simulation of *in silico* molecular docking of genistein, in Bcr-Abl fusion protein tyrosine kinase. The complex is illustrated with genistein bound in proper binding cavity of the enzyme. Helices are illustrated as cartoon colored by chain bow (the final structure was ray-traced) (Authors' unpublished data).

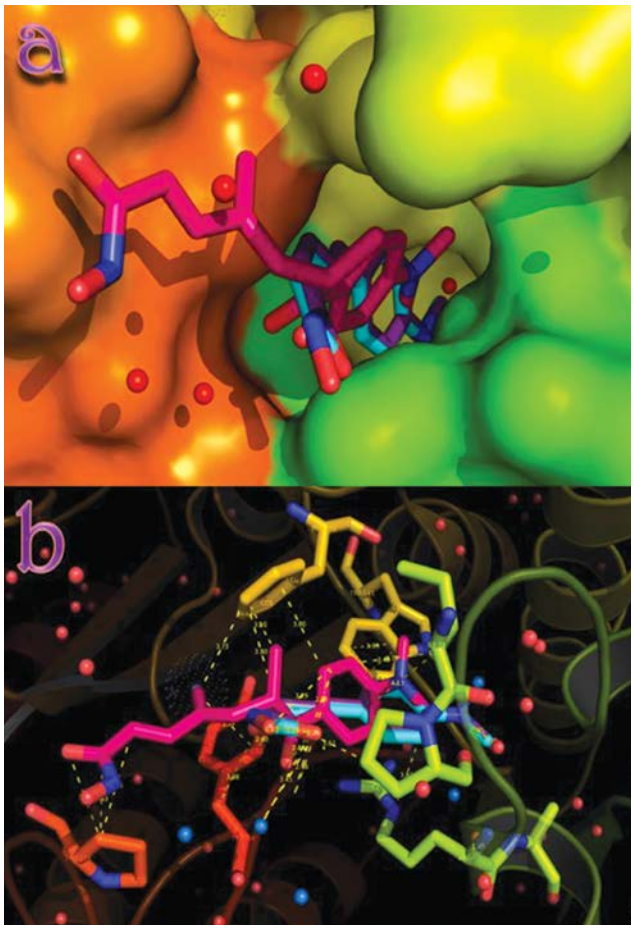

Figure 4. Molecular docking of novel hydroxamic acid compounds 5ii and 5iv on target 1t64 (HDAC8, histone deacetylase 8) with its inhibitor TSA. Ligand binding cavity of the target protein is represented as a solid surface colored by chain bow, while superimposed bound molecules are shown in stick representation and colored by atom type: 5ii, cyan; 5iv, purple; TSA, hot-pink) **(a).** Binding interactions (yellow dashed lines) between 5ii, 5iv and TSA molecules (coloring same as in **(a)**) and the amino acid residues of the binding site of HDAC protein (helices are illustrated as semi-transparent cartoon colored by chain) **(b).** Both images were ray-traced (Authors' unpublished data).

2.7.1. The docking problem

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. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules using for example scoring functions. Docking is used to predict the binding orientation of small molecule drug candidates to their protein targets in order to predict the affinity and activity of the small molecule. The search for the global minimum or the complete set of low energy minima on the free energy surface when two molecules come in contact is commonly referred to as the "docking problem" [69]. Any useful molecular docking program must be computationally efficient in determining the most favorable binding mode, sufficiently sensitive in its scoring function to discriminate between alternate binding modes and the correct mode, and robust enough to allow various ligand-receptor systems to be studied. The associations between biologically relevant molecules such as proteins, nucleic acids, carbohydrates, and lipids play a central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (e.g., agonism vs antagonism). Therefore, docking may predict both the strength and type of signal produced.

Molecular docking can be thought of as an optimization problem of "lock-and-key" or "best-fit" or "hand-glove", where one is interested in finding the correct relative orientation of the "key" (ligand) which will open up the "lock" (protein). During the course of the process, the ligand and the protein adjust their conformation to achieve an overall "best-fit" and this kind of conformational adjustments resulting in the overall binding is referred to as "induced-fit" [70]. The aim of molecular docking is to achieve an optimized conformation for both the protein and ligand and relative orientation between protein and ligand such that the free energy of the overall system is minimized. This includes determining the orientation of the compound, its conformational geometry, and the scoring. By using the knowledge gained from the docking study, fewer compounds need be synthesized and assayed, and a higher percentage of compounds assayed are found to be active. Hence docking plays an important role in the rational design of drugs [71].

2.7.2. Small molecules docked to a protein

However, the identification of molecular features being responsible for specific biological recognition, or the prediction of compound modifications that improve potency, are complex issues that are often difficult to understand and - even more so - to simulate on a computer. In view of these challenges, docking is generally devised as a multi-step process in which each step introduces one or more additional degrees of complexity. The process begins with the application of docking algorithms that pose small molecules in the active site [72]. This is challenging, as even relatively simple organic molecules can contain many conformational degrees of freedom. Sampling these degrees of freedom must be performed with sufficient accuracy to identify

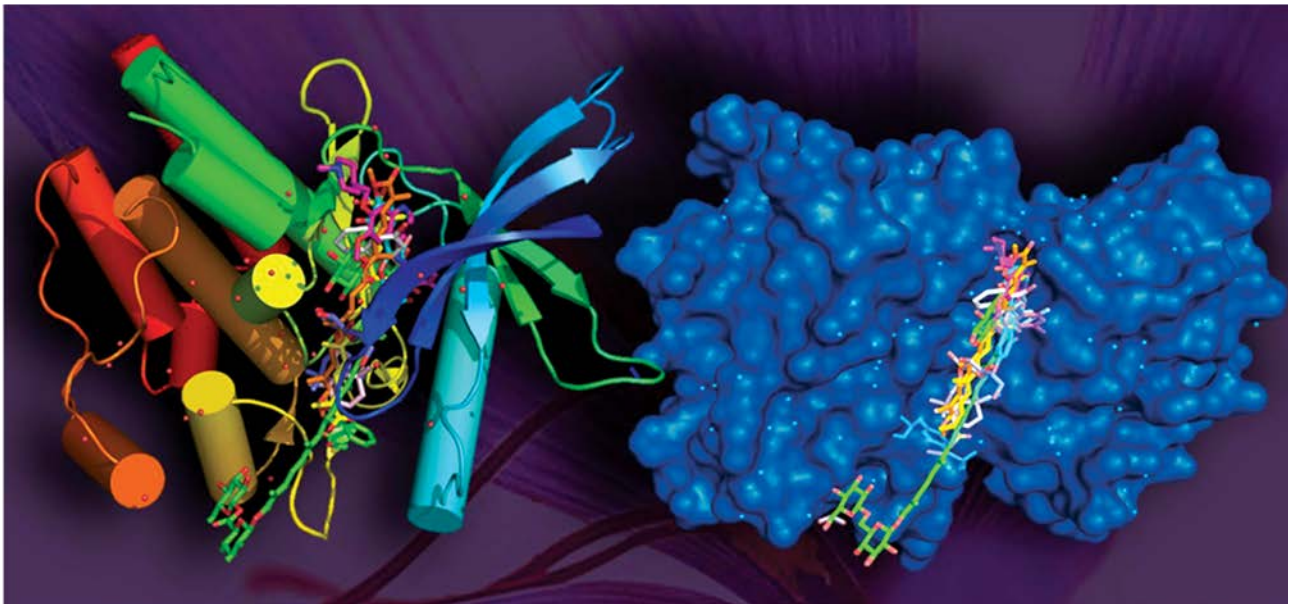

Figure 5. Ret tyrosine kinase protein target (PDB ID 2ivu) bound with Crocus Sativus L. constituents, crocin (CRC, green C atoms), crocetin (CRT, yellow C atoms), dimethylcrocetin (DMCRT, orange C atoms), picrocrocin (PCRC, blue C atoms), safranal (SFR, light pink C atoms) and the flavonoid compounds quercetin (QUR, cyan C atoms) and flavone (FLV, white C atoms), after docking simulation into the ligand binding pocket of ret TK target protein. In the protein (either depicted as cartoon or semi-transparent surface colored in chain bow or chain, respectively), is already bound the Ret TK inhibitor ZD6474 (magenta C atoms). Both images were ray-traced (Authors' unpublished data).

the conformation that best matches the receptor structure, and must be fast enough to permit the evaluation of thousands of compounds in a given docking run. Algorithms are complemented by scoring functions that are designed to predict the biological activity through the evaluation of interactions between compounds and potential targets. Pre-selected conformers are often further evaluated using more complex scoring schemes with more detailed treatment of electrostatic and van der Waals interactions, and inclusion of at least some solvation or entropic effects [68]. It should also be noted that ligand-binding events are driven by a combination of enthalpic and entropic effects, and that either entropy or enthalpy can dominate specific interactions. This often presents a conceptual problem for contemporary scoring functions (discussed below), because most of them are much more focused on capturing energetic than entropic effects.

In addition to problems associated with scoring of compound conformations, other complications exist that make it challenging to accurately predict binding conformations and compound activity. These include, among others, limited resolution of crystallographic targets, inherent flexibility, induced fit or other conformational changes that occur on binding, and the participation of water molecules in protein-ligand interactions. Without doubt, the docking process is scientifically complex.

2.7.3 Solving the "docking problem". Shape complementarity and Simulation

There have been many approaches to the solution of the "docking problem" over the last years. For example, some methods rely on complex molecular dynamics simulations while others use less costly graph matching approaches. There is generally a compromise between speed and accuracy, with some methods giving much more information and insight into the nature of the protein/ligand interactions and other methods optimized for speed of docking thousands of putative ligands. One approach uses a matching technique that describes the protein and the ligand as complementary surfaces [73]. The second approach simulates the actual docking process in which the ligand-protein pair-wise interaction energies are calculated [74]. Both approaches have significant advantages as well as some limitations.

Geometric matching/ shape complementarity methods describe the protein and ligand as a set of features that make them dockable [43]. The receptor's molecular surface is described in terms of its solvent-accessible surface area and the ligand's molecular surface is described in terms of its matching surface description. The complementarity between the two surfaces amounts to the shape matching description that may help finding the complementary pose of docking the target and the ligand molecules. Other complementary

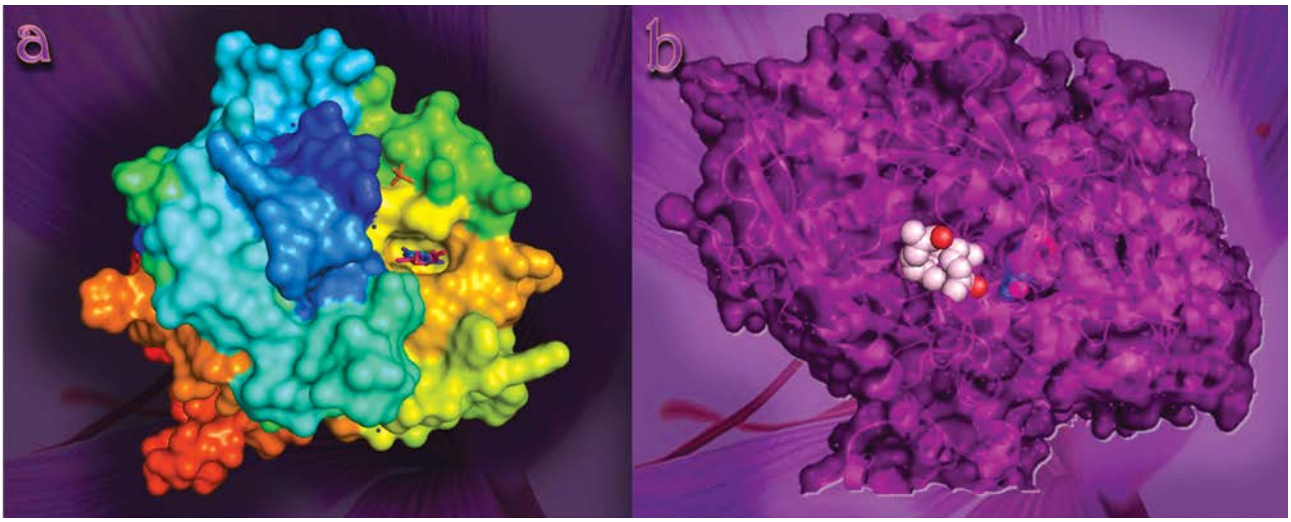

Figure 6. Molecular docking pose of safranal (SFR) (rendered in hotpink C atoms) on human Coagulation Factor IXa in complex with PBZ (P-amino benzamidine) & TBU (tertiary-butyl alcohol) (PDB ID Nr: 1RFN) (blue and red C atoms, respectively) **(a),** crocetin (CRT) (rendered in white C atom spheres) docked into Glycogen Phosphorylase B (PDB ID 1GPB) already bound with its inhibitor PLP999 (rendered in blue C atom spheres) (target protein is illustrated as solid **(a)** or semitransparent **(b)** surface along with cartoon representation and colored by chainbow and chain, respectively) (Authors' unpublished data).

approaches are to describe the hydrophobic features of the protein using turns in the main-chain atoms or to use a Fourier shape descriptor technique [75]. Shape complementarity methods can quickly scan through several thousand ligands in a matter of seconds and actually figure out whether they can bind at the protein's active site, and are usually scalable to even protein-protein interactions. They are also much more amenable to pharmacophore-based approaches, since they use geometric descriptions of the ligands to find optimal binding. Whereas the shape-complementarity-based approaches are typically fast and robust, they cannot usually model the movements or dynamic changes in the ligand/protein conformations accurately.

The simulation of the docking process as such is a much more complicated process. In this approach, the protein and the ligand are separated by some physical distance, and the ligand finds its position into the protein's active site after a certain number of "moves" in its conformational space. The moves incorporate rigid body transformations such as translations and rotations, as well as internal changes to the ligand's structure including torsion angle rotations. Each of these moves in the conformation space of the ligand induces a total energetic cost of the system, and hence after every move the total energy of the system is calculated [76]. The obvious advantage of the method is that it is more amenable to incorporate ligand flexibility into its modeling whereas shape complementarity techniques have to use some ingenious methods to incorporate flexibility in ligands. and that the process is physically closer to what happens in reality, when the protein and ligand approach each other after molecular recognition. A clear disadvantage of this technique is that it takes longer time to evaluate the optimal pose of binding since they have to explore a rather large energy landscape. However, grid-based techniques as well as fast optimization methods have significantly ameliorated these problems.

2.7.4. Mechanics of docking: Searching the conformational space for docking

To perform a docking screen, the first requirement is a structure of the protein of interest. Usually the structure has been determined using a biophysical technique such as x-ray crystallography, or less often, NMR spectroscopy. This protein structure and a database of potential ligands serve as inputs to a docking program. The success of a docking program depends on two components: the search algorithm and the scoring function.

The search space in theory consists of all possible orientations and conformations of the protein paired with the ligand. However, in practice with current computational resources, it is impossible to exhaustively explore the search space and all possible distortions of conformation. Most docking programs in use account for a flexible ligand, and several attempt to model a flexible protein receptor. Each "snapshot" of the pair is referred to as a "pose". A variety of conformational

search strategies have been applied to the ligand and to the receptor. These include: (1) systematic or stochastic torsional searches about rotatable bonds; (2) molecular dynamics simulations; and (3) genetic algorithms to "evolve" new low energy conformations [77]. Posing is the process of determining whether a given conformation and orientation of a ligand fits the active site. This is usually a fuzzy procedure that returns many alternative results.

2.7.5. Ligand and receptor flexibility

Conformations of the ligand may be generated in the absence of the receptor and subsequently docked or conformations may be generated on-the-fly in the presence of the receptor binding cavity [78], or with full rotational flexibility of every dihedral angle using fragment-based docking (eHiTS) [9]. Force field energy evaluations are most often used to select energetically reasonable conformations but knowledge-based methods have also been used [79]. Dealing with receptor flexibility in docking methodologies is still a thorny issue. The main reason behind this difficulty is the large number of degrees of freedom that have to be considered in this kind of calculations. Neglecting it, however, leads to poor docking results in terms of binding pose prediction. Multiple static structures experimentally determined for the same protein in different conformations [80] or rotamer libraries of amino acid side chains that surround the binding cavity may be searched to generate alternate but energetically reasonable protein conformations [81].

2.7.6. Scoring function: Posing and ranking

Both posing and ranking involve scoring [72]. The pose score is often a rough measure of the fit of a ligand into the active site. The rank score is generally more complex and might attempt to estimate binding energies. The scoring function takes a pose as input and returns a number indicating the likelihood that the pose represents a favorable binding interaction [77]. Most scoring functions are physics-based molecular mechanics force fields that estimate the energy of the pose; a low (negative) energy indicates a stable system and thus a likely binding interaction.

An alternative approach is to derive a statistical potential for interactions from a large database of protein-ligand complexes, such as the Protein Data Bank, and evaluate the fit of the pose according to this inferred potential. Scoring functions trained with this data can dock high affinity ligands correctly, but they will also give plausible docked conformations for ligands that do not bind. This gives a large number of false positive hits (ligands predicted to bind to the protein that actually do not). One way to reduce the number of false positives is to recalculate the energy of the top scoring poses using (potentially) more accurate but computationally more intensive techniques such as Generalized Born or Poisson-Boltzmann methods [74]. One early method was developed by Böhm [82] and used the Gibbs free energy equation to develop a general-purposed empirical scoring function in order to describe the binding energy. The following "Master Equation" was derived: $\Delta G_{bind} = -RT \ln K_d$, $K_d = [Receptor][Acceptor]/[Complex]$, $\Delta G_{bind} = \Delta G_{desolvation} + \Delta G_{motion} + \Delta G_{configuration} + \Delta G_{interaction}$, where: desolvation – enthalpic penalty for removing the ligand from solvent, motion – entropic penalty for reducing the degrees of freedom when a ligand binds to its receptor, configuration – conformational strain energy required to put the ligand in its "active" conformation and interaction – enthalpic gain for "resolvating" the ligand with its receptor.

Various computational methods are used to estimate each of the components of the master equation. For example, the change in polar surface area upon ligand binding can be used to estimate the desolation energy and the number of rotatable bonds frozen upon ligand binding is proportional to the motion term. The configurational or strain energy can be estimated using molecular mechanics calculations. Finally, the interaction energy can be estimated using methods such as the change in non-polar surface, statistically derived potentials of mean force, the number of hydrogen bonds formed, etc. In practice, the components of the master equation are fit to experimental data using multiple linear regression. This can be done with a diverse training set including many types of ligands and receptors to produce a less accurate but more general ("global") model or a more restricted set of ligands and receptors to produce a more accurate but less general ("local") model [77,83]. Ranking is a more advanced process than pose scoring, that typically takes several results from an initial scoring phase and re-evaluates them [72]. This process usually attempts to estimate the free energy of binding as accurately as possible. Although the posing phase might use simple energy calculations (electrostatic and van der Waals), ranking procedures typically involve more elaborate calculations (perhaps including

properties such as entropy or explicit solvation).

2.8. Docking applications

A binding interaction between a small molecule ligand and an enzyme protein may result in activation or inhibition of the enzyme. If the protein is a receptor, ligand binding may result in agonism or antagonism. Docking the field of drug design may be applied to: (1) hit identification (see virtual screening); (2) lead optimization - to predict in where and in which relative orientation a ligand binds to a protein ("binding mode" or "pose") and (3) Bioremediation - to predict pollutants that can be degraded by enzymes [84]. Flexible ligand-search methods include: (1) Random/stochastic, implemented in the following programs: AutoDock (MC), MOE-Dock (MC, TS), GOLD (GA), PRO_LEADS (TS); (2) Systematic, implemented in: DOCK (incremental), FlexX (incremental), Glide (incremental), Hammerhead (incremental), FLOG (database); and (3) Simulation, implemented in: DOCK, Glide, MOE-Dock, AutoDock and Hammerhead. There are a lot of popular molecular docking software programs [85]. A list of established docking software programs and *de novo* design tools is depicted in Table 1.

3. Homology modeling

Often, the crystal structure of the therapeutic target is not available. In this case a 3D structure of a homologous protein will have to be determined. Homology modeling, also known as "comparative modeling of protein", refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental 3D structure of a related structurally known homologous protein (the "template") [86]. Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structure [87].

4. Combinatorial chemistry

With the increasing characterization of the 3D structures of receptors and enzymes, the design of molecules that interact with these biological targets appeared to be an intellectual approach

Table 1. Molecular docking software and de novo design tools useful in rational drug design and development approaches

Molecular docking tools		
ADAM	GLIDE	MultiDock
AutoDOCK	GOLD	QXP
Bielefeld Protein Docking	GRAMM	PatchDock
BioCaChe	GREEN	Prodock
CDOCKER	GRID	PRO LEADS
CombiDOCK	Hammerhead	RDOCK
DIVALI	Hex	RosettaDOCK
DockVision	ICM	RPScore
DOCK	KORDO	SEED
DOT	LUDI	SDOCKER
eHiTS	MCDOCK	SLIDE
FTDock	Molecular Docking Server	Validate
FlexX	Molegro	WISDOM
FLOG	MolFit	Virtual Docker
GEMDOCK	MPI Protein Docking	ZDOCK
De novo design tools		
LUDI	DLD/MCSS	HOOK
BUILDER	Genstar	Legend
SMOG	Group-Build	MCDNLG
CONCEPTS	Grow	SPROUT

using modern computer technology. Even with the advent of a plethora of such widely accepted and practiced methodologies and superb statistical methods, traditional synthetic techniques and time-tested biological screening procedures, invariably prove to be very expensive and sometimes turn out to be non-productive in nature in the end. The aim of combinatorial chemistry ("CombiChem") is the generation of large numbers of compounds very quickly [88]. Interestingly, it overwhelmingly makes use of comprehensive and extended libraries of chemical functional moieties which specifically interact either with a "base molecule" or with a "parent molecule" in a highly systematic small quantum of well-defined purely synthetic stepwise procedures.

With CombiChem it has become possible and cost-effectively to use solid- or solution-phase syntheses with different chemistries and scaffolds to produce libraries tailor-made for finding a new lead or optimizing a lead directed at almost any class of target to find newer "prototype drug molecules" and to refine and optimize the QSAR [89]. New tools, such as molecular docking algorithms [64,72], mapping of protein binding sites by NMR [90] and homology modeling of proteins [88], allow an unprecedented level of rational design to guide the synthesis of prospective drugs. Salient features of CombiChem include the chemical diversity of products originating with simple "same" clean, reproducible reactions to yield thousands of molecules; one makes use of the solid-state synthetic techniques to allow the desired growth of drug molecules upon polymer support; and "robotics" is employed profusely to cut down the "effective cost of synthesis" drastically. Two kinds of approaches are used in CombiChem, rational designing and reverse designing. Rational designing is the most "popular technique" and exploits the lock-and-key model with respect to the ligand-receptor docking. However, there are limitations (e.g. conformational flexibilities for both ligand and receptor), problems with active conformers (water and salt effects) and with binding high-energy conformers [36]. Reverse designing essentially involves the grouping together and searching of functionally and structurally identical chemical entities by making use of common and biologically effective motif, termed as pharmacophore [91], which is specifically found either in the corporate or commercial database. Importantly, at every articulated step carried out meticulously in the intricate discovery phenomenon one has to heavily depend upon the manipulative skills related to CADD, which provides an advantage of data-processing into vital and relevant information for future analysis in drug design.

4.1. Combinatorial library design

Combinatorial library design is an important application of CombiChem that can exploit automation and robotics to enable the rapid production of large numbers of compounds. Libraries are synthesized for both lead identification and lead optimization purposes [92]. The resultant libraries consist of products formed by combining "reactants" (reagents, monomers) with each other or with a "scaffold" (template,core). While many larger pharmaceutical companies maintain libraries of several hundred thousand natural and synthetic compounds, combinatorial chemistry can generate libraries with millions of compounds with low cost and relative ease. Parallel synthesis, in which multiple analogs are synthesized at a time efficiently using reactants and automation/robotics, is now a standard part of the drug discovery process.

5. Pharmacophores

A pharmacophore was first defined by Paul Ehrlich in 1909 as "a molecular framework that carries (phoros) the essential features responsible for a drug's (pharmacon's) biological activity" [93]. A modern definition of a pharmacophore is the specific 3D arrangement of functional groups (steric and electronic features) within a molecular framework that are necessary to bind to a macromolecule and/or an enzyme active site (with optimal supramolecular interactions) in order to trigger or inhibit its biological response [55,94]. This approach to lead generation has proved highly successful [95]. In particular it is possible to identify active compounds that contain a different core structure from that of the compounds used to generate the model (lead-hopping). Pharmacophores help to create a molecular descriptor for similarity-and diversity-related tasks. The diversity-related descriptor identifies in a systematic way all the potential pharmacophores that a molecule can exhibit. Triplet (three-point) and quartet (four-point) pharmacophore representations have been extensively used (in addition to two-point/2D approaches), with a variety of features sampled at each point and inter-feature distances considered in a discrete set of ranges ("bins") [96].

5.1. Pharmacophore models

The pharmacophoric assumption led to a prob-

lem statement that logically is composed of two processes. First is the determination, by chemical modification and biological testing, of the relative importance of different functional groups in the drug to receptor recognition. This can give some indication of the nature of the functional groups in the receptor that are responsible for binding of the set of drugs. Second, a hypothesis is proposed concerning correspondence, either between functional groups (pharmacophore) in different congeneric series of the drug or between recognition sites points postulated to exist within the receptor (binding-site model).

Once a pharmacophore is established, the medicinal chemist has a host of 3D database search tools to retrieve novel compounds that fit the pharmacophore model [97]. The designation of a pharmacophore is the first essential step towards understanding the interaction between a receptor and a ligand. The "recognition of receptor" could be determined by carrying out various chemical structural modifications vis-a-vis biological screening of the different functional moieties present in the drug molecule(s). However, such investigative exercises would throw ample light with respect to the very fundamental nature of the functional moieties strategically positioned in the receptor which are exclusively responsible for affording the right type of linkage to the set of drugs. Secondly, a proposed hypothesis is formed of close similarity either between the pharmacophore (i.e., functional moieties) in various structural analogues of the drug molecule (i.e., congeneric series), or between the suggested recognition site points duly postulated to exist very much within the receptor i.e., the binding-site model.

New and extended methods of QSAR analysis have evolved to translate pharmacophore information into QSAR models that, in turn can be used as virtual high-throughput screens for activity profiling of a library [98]. Moreover, a successful application of fingerprinting approach was previously employed to generate 10,549 three-point pharmacophores by enumerating several distance ranges and pharmacophoric features. Subsequently, the fingerprint was used as a descriptor for developing a QSAR model using partial least squares [99].

This ability of a pharmacophore model to find new classes of inhibitors when one class is known is called "scaffold hopping". Once a few active compounds have been found, it is very easy to run sub-structure searches of large compound databases. Substructure searches find any compound that contains the exact pattern of atoms and bond given in the search query. For example, a substructure search can find all compounds containing an identical fused ring system. Substructure searches are certainly a useful starting point but they fail to find very similar compounds, such as analogs that have five-member rings replaced by six-member rings.

5.1.1. Components of a pharmacophore

Most often, 3D pharmacophores in terms of enzyme-ligand interaction types are described as follows: ion-ion, ion-dipole, dipole-dipole, hydrogen bonds, induced polarization, charge transfer, dispersion forces and steric repulsion. Some phar-

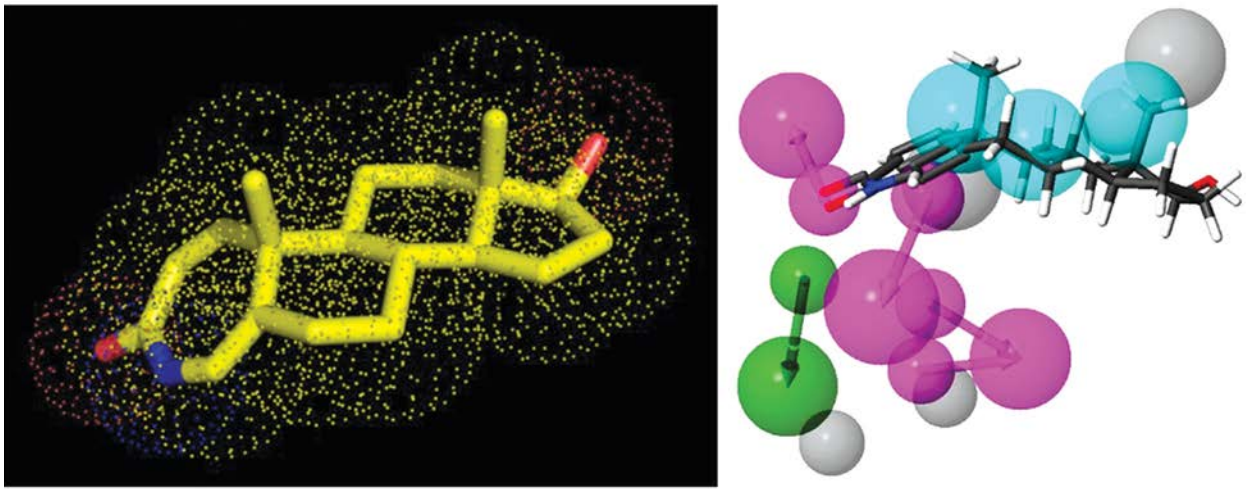

Figure 7. Pharmacophore model of an aza-steroidal compound depicted in a ray-traced stick representation with inclusion of electron density surface (left) for the target protein glutamine gamma-glutamyltransferase E (PDB ID Nr 1L9N) derived from a pharmacophore approach through searching of over 2000 human protein targets, to identify potential target candidates for the given molecule with unidentified binding targets (right). The pharmacophore map consists of three hydrophobic features (cyan spheres), one hydrogen bond donor point (green spheres and vector) and four hydrogen bond acceptor points (purple spheres and vector) (Authors' unpublished data).

macophore searches, particularly those from earlier in the development of pharmacophore searching, will search for a particular arrangement of specific atoms: for example, a carbonyl, a benzene ring, and a hydroxy group at specified distances apart [100].

Most pharmacophore descriptions in the more recently developed software packages are more general. Some features typically include hydrogen bond donors, hydrogen bond acceptors, aromatic rings (may be ring atoms, ring center, or normal to the ring), hydrophobic centers (also called neutral centers), positive charge centers, negative charge centers, acidic groups, basic groups, bulky groups engaged in steric interactions, planar atoms, CO2 centroid (i.e., ester or carboxylic acid), NCNp centroid, metal (also called a metal ligator), excluded volumes - forbidden regions, where the protein is and the ligand cannot have functional groups (Figure 7).

Some programs can impose additional constraints on the centers, such as noting that one of the above centers is only satisfied if the atom is in a chain or in a ring. Some programs can produce an IC50 estimate based on a curve fit to the pharmacophore model. These tend to be very crude estimates, sometimes leading in error by several orders of magnitude. The strength of pharmacophore searching is in identifying compounds that should be tested for activity, not in quantitative prediction [99,100].

5.2. Binding-site models

Among the many tools available, 3D alignment-dependent methods are usually slow and based on simplified representations of binding site atoms or surfaces. On the other hand, fast and efficient alignment-free methods have recently been described but suffer from a lack of interpretability [101]. There exists one major disadvantage in the "pharmacophore models" with the overlapping of the various functional moieties in perfect alignment to the pharmacophoric hypothesis. It has been established that molecules bearing such functional moieties which exhibit 3D similarity may interact with the same site and specific geometrically identical structures closely associated with one particular site are able to interact with almost equal favor and affinity leading to a number of plausible structural orientations of the identical functional moieties. Therefore, it has become almost necessary to take into consideration the converging point of approximately equal energetic configurations associated with either an H-bond donor or acceptor to circumvent the ensuing problem skillfully.

Recently, Wood and his colleagues [102] have presented KRIPO (Key Representation of Interaction in POckets): a new method for quantifying the similarities of binding site subpockets based on pharmacophore fingerprints. Interestingly, a comparison between the specific pharmacophore and the binding-site hypotheses establishes the fact that the latter is definitely more acceptable and plausible physiochemically due to the fact that the ensuing overlap of the functional moieties in the process of binding to a receptor is found to be more restrictive than anticipating that the particular site does remain more or less fixed when getting bound to different ligands.

6. Informatics

The unprecedented flood of information from genome sequences and functional genomics in one hand and combinatorial chemistry, HTS, and virtual screening on the other hand has given rise to new fields of bioinformatics and chemoinformatics, which combines elements of biology and chemistry with mathematics, statistics and computer sciences. Analyses in bioinformatics and chemoinformatics predominantly focus on several types of large datasets available such as macromolecular structures, genome sequences, 3D chemical databases and compound libraries. A tremendous number of bioinformatics tools have been created and provided to the entire scientific community a response to an increasing need in terms of sequence analyses and structure predictions. Informatics methodologies rely on a variety of computational techniques [103] including sequence and structural alignment, database design and data mining, macromolecular geometry, phylogenetic tree construction, prediction of protein structure and function, gene searching and expression data clustering, chemical-similarity clustering, diversity analysis, library design, virtual screening and QSAR.

Combinatorial chemistry and HTS primarily depend on chemoinformatics to increase their effectiveness. Recent advances in chemoinformatics include new molecular descriptors and pharmacophore mapping techniques, statistical tools and novel visualization methods. A major task of informatics in the future is to develop software tools that provide the means to store, extract, analyze, and display data in a way that chemists can easily understand and appreciate. In attempts to decipher chemical/biological information, com-

puters require the use of descriptors. Hence, hundreds of molecular descriptors have been reported in the literature, ranging from simple bulk properties to elaborate 3D formulations and complex molecular fingerprints [104,105].

7. Conclusions

Remarkable progress has been made during the past few years in almost all the areas concerning the complex field of drug design and discovery involving the application of many different fields of knowledge. Multi-target drugs have been found effective in controlling complex diseases as cancer. However, how to design multi-targeted drugs presents a great challenge.

To this end, rational strategies for future drug research and development should be proposed. An active area for the search of new anticancer therapies is concerned with the use of computational approaches based on chemoinformatics and bioinformatics. These approaches consider only small or larger series of structurally related compounds and the studies are generally realized for only one target (e.g. a protein). This is an important limitation since the possible target-proteins for a given drug may be numerous. Developing multi-target discriminant models, for the *in silico* design and virtual screening of anticancer agents against numerous proteins involved in a specific signal transduction pathway are a means to overcome the limitations.

Drug discovery in the new millennium is armed with not only new and efficient techniques for producing, purifying and screening new drug entities, but also with computing power, that was unimaginable a decade ago. Thus, with data compiled in various DrugBanks, scientists can *a priori* predict absorption and distribution properties of lead molecules *in silico*.

Technology for molecular modeling and drug design, coupled with increased availability of structural information, regarding molecular targets, signaling pathways and the completion of the human genome project, have created major new opportunities for drug discovery, in all areas of human disease including cancer. Technologies such as *in silico* virtual drug screening, identification of novel therapeutic targets, and new innovative techniques in combinatorial chemistry, combined with latest methodologies in bioinformatics, molecular modeling and docking, will allow biological information to be analyzed and managed on a very large scale, and in an extremely short time. Due to the recent advances in this area of research, modern multi-targeted drug discoveries are expected to have an impact on cancer treatment. The aim of future research should be to hit multiple targets with single molecules and we believe that, within the next five to ten years, we'll see more agents entering clinical trials that use rational multi-target approach as the basis of drug discovery.

Conflict of interests

The authors declare no confict of interests.

References

1. Hopfinger AJ. Computer-assisted drug design. J Med Chem 1985;28:1133-1139.
2. Xie L, Bourne PE. A robust and efficient algorithm for the shape description of protein structures and its application in predicting ligand binding sites. BMC Bioinformatics 2007;8:S9.
3. de Azevedo WF, Dias R. Computational methods for calculation of ligand-binding affinity. Curr Drug Targets 2008;9:1031-1039.
4. Blundell TL, Jhoti H, Abell C. High throughput crystallography for lead discovery in drug design. Nature Rev Drug Discov 2002;1:45-54.
5. Guido RV, Oliva G, Andricopulo AD. Modern drug discovery technologies: opportunities and challenges in lead discovery. Comb Chem High Throughput Screen 2011;14:830-839.
6. Polgár T, Keseru GM. Integration of virtual and high throughput screening in lead discovery settings. Comb Chem High Throughput Screen 2011;14:889-897.
7. Sukumar N, Das S. Current trends in virtual high throughput screening using ligand-based and structure-based methods. Comb Chem High Throughput Screen 2011;14:872-888.
8. Sun H. Pharmacophore-based virtual screening. Curr Med Chem 2008;15:1018-1024.
9. Zsoldos Z, Reid D, Simon A, Sadjad SB, Johnson AP. eHiTS: A new fast, exhaustive flexible ligand docking system. J Mol Graph Mod 2007;26:198-212.
10. Cheng F, Zhou Y, Li J, Li W, Liu G, Tang Y. Prediction of chemical-protein interactions: multitarget-QSAR versus computational chemogenomic methods. Mol Biosyst 2012;8:2373-2384.
11. Leach AR, Harren J (Eds): Structure-based Drug Discovery. 2007;Berlin: Springer.
12. Kroemer RT. Structure-based drug design: docking and scoring. Curr Protein Pept Sci 2007; 8:312-328.
13. Wang R,Gao Y, Lai L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. J Mol Mod

2000;6:498-516.

14. Schneider G, Fechner U. Computer-based de novo design of drug-like molecules. Nat Rev Drug Discov 2005;4:649-663.
15. Liang S, Meroueh SO, Wang G, Qiu C, Zhou Y. Consensus scoring for enriching near-native structures from protein-protein docking decoys. Proteins 2009;75:397-403.
16. Deng Z, Chuaqui C, Singh J. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. J Med Chem 2004;47:337-344.
17. Amari S, Aizawa M, Zhang J et al. VISCANA: visualized cluster analysis of protein-ligand interaction based on the ab initio fragment molecular orbital method for virtual ligand screening. J Chem Inf Model 2006;46:221-230.
18. De Vivo M. Bridging quantum mechanics and structure-based drug design. Front Biosci 2011;16:1619-1633.
19. Ma XH, Shi Z, Tan Z et al. In-Silico approaches to multi-target drug discovery : computer aided multi-target drug design, multi-target virtual screening. Pharm Res 2010;27:739-749.
20. Zhang X, Gibbs AC, Reynolds CH, Peters MB, Westerhoff LM. Quantum Mechanical Pairwise Decomposition Analysis of Protein Kinase B Inhibitors: Validating a New Tool for Guiding Drug Design. J Chem Inf Model 2010;5:651-661.
21. Mitra S, Bartlett SG, Newcomer ME. Identification of the Substrate Access Portal of 5-Lipoxygenase. Biochemistry 2015;54:6333-6342.
22. Zoete V, Grosdidier A, Michielin O. Docking, virtual high throughput screening and in silico fragment-based drug design. J Cell Mol Med 2009;13:238-248.
23. Mehla K, Ramana J. Structural signature of Ser83Leu and Asp87Asn mutations in DNA gyrase from enterotoxigenic Escherichia coli and impact on quinolone resistance. Gene 2016;576:28-35.
24. Wang R,Gao Y, Lai L. LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design. J Mol Mod 2000;6:498-516.
25. Filz OA, Lagunin AA, Filimonov DA, Poroikov VV. In silico fragment-based drug design using a PASS approach. SAR QSAR Environ Res 2012;23:279-296.
26. Westbrook J, Feng Z, Chen L, Yang H, Berman HM. The protein data bank and structural genomics. Nucleic Acid Res 2003;31:489-491.
27. Balani SK, Miwa GT, Gan LS, Wu JT, Lee FW. Strategy of utilizing in vitro and in vivo ADME tools for lead optimization and drug candidate selection. Curr Top Med Chem 2005;5:1033-1038.
28. Singh SS. Preclinical pharmacokinetics: an approach towards safer and efficacious drugs. Curr Drug Metab 2006;7:165-182.
29. Gombar VK, Silver IS, Zhao Z. Role of ADME characteristics in drug discovery and their in silico evaluation: in silico screening of chemicals for their metabolic stability. Curr Top Med Chem 2003; 3:1205-1225.
30. Smith DA. Design of drugs through a consideration of drug metabolism and pharmacokinetics. Eur J Drug Metab Pharm 1994;3:193-199.
31. Moriguchi I, Shuichi H, Liu Q, Nakagome I, Matsushita Y. Simple method of calculating octanol/water partition coefficient. Chem Pharm Bull 1992;40:127-130.
32. Tetko IV, Bruneau P, Mewes HW, Rohrer DC, Poda GI. Can we estimate the accuracy of ADME-Tox predictions? Drug Discov Today 2006;11:700-707.
33. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 1997;23:3-25.
34. Leeson P. Drug discovery: Chemical beauty contest. Nature 2012;481:455-456.
35. Oprea TI, Davis AM, Teague SJ, Leeson PD. Is There a Difference between Leads and Drugs? A Historical Perspective. J Chem Inf Comput Sci 2001;41:1308-1315.
36. Ghose AK, Viswanadhan VN, Wendoloski JJ. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. J Combin Chem 1999;1:55-68.
37. Bois FY. Physiologically-based modelling and prediction of drug interactions, Basic Clin Pharmacol Toxicol 2009;106:154-161.
38. Smith DA. Design of drugs through a consideration of drug metabolism and pharmacokinetics. Eur J Drug Metab Pharm 1994;3:193-199.
39. Blum LC, Reymond JL. 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13. J Am Chem Soc 2009;131:8732-8733.
40. Rester U. From virtuality to reality - Virtual screening in lead discovery and lead optimization: A medicinal chemistry perspective. Curr Opin Drug Discov Devel 2008;11:559-568.
41. Badrinarayan P, Sastry GN. Virtual high throughput screening in new lead identification. Comb Chem High Throughput Screen 2011;14:840-860.
42. Walters WP, Stahl MT, Murcko MA. Virtual screening- an overview. Drug Discov Today 1998;3:160-178.
43. Shoichet BK. Virtual screening of chemical libraries. Nature 2004;432:862-865.
44. Schneider G. Virtual screening: an endless staircase? Nat Rev Drug Discov 2010;9:273-276.
45. Ballester PJ. Ultrafast shape recognition: method and applications. Future Medicinal Chemistry 2011;3:65-78.
46. Willet P, Barnard JM, Downs GM. Chemical similarity searching. J Chem Inf Comput Sci 1998;38:983-996.
47. Hartenfeller M, Schneider G. De novo drug design. Methods Mol Biol 2011;672:299-323.
48. Triggle DJ. The chemist as astronaut: searching for biologically useful space in the chemical universe. Biochem Pharmacol 2009;78:217-223.
49. Simmons KJ, Chopra I, Fishwick CW. Structure-based discovery of antibacterial drugs. Nat Rev Microbiol

2010;8:501-510.

50. Grant JA, Gallardo MA, Pickup B. A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. J Comput Chem 1996;17:1653-1666.
51. Rush TS, Grant JA, Mosyak L, Nicholls A. A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein-Protein Interaction. J Med Chem 2005;48:1489-1495.
52. Wang L, Evers A, Monecke P, Naumann T. Ligand base lead generation – considering chemical accessibility in rescaffolding approaches via BROOD. Cheminform 2012;4: O20.
53. Gao Q, Yang L, Zhu Y. Pharmacophore based drug design approach as a practical process in drug discovery. Curr Comput Aided Drug Des 2010; 6:37-49.
54. Wolber G, Langer T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J Chem Inf Model 2005;45:160-169.
55. Wolber G, Seidel T, Bendix F, Langer T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. Drug Discov Today 2008;13:23-29.
56. Kirchmair J, Distinto S, Schuster D, Spitzer G, Langer T, Wolber G. Enhancing drug discovery through in silico screening: strategies to increase true positives retrieval rates. Curr Med Chem 2008;15:2040-2053.
57. Cramer RD, Patterson DE, Clark RD, Soltanshahi F Lawless MSJ. Virtual compound libraries: A new approach to decision making in molecular discovery research. J Chem Inf Comput Sci 1998;38:1010-1023.
58. Horvath D (Ed): Combinatorial Library Design and Evaluation. Principles, Software Tools, and Applications in Drug Discovery. Marcel Dekker, New York, 2001, pp 429-472.
59. Wishart DS, Knox C, Guo AC et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008;36:D901-906.
60. Hewett M, Oliver DE, Rubin DL et al. PharmGKB: the pharmacogenetics knowledge base. Nucleic Acids Res 2002;30:163-165.
61. Wu J, Vallenius T, Ovaska K, Westermarck J, Makela TP, Hautaniemi S. Integrated network analysis platform for protein-protein interactions. Nat Methods 2009;6:75-77.
62. Olah M, Rad R, Ostopovici L et al. WOMBAT and WOMBAT-PK: bioactive databases for lead and drug discovery. In: Schreiber SL, Kapoor TM & Wess G (Eds): Chemical Biology: From Small Molecules to Systems Biology and Drug Design. Wiley-VCH, New York, 2007, pp 760-786.
63. Zheng S, Luo X, Chen G et al. A new rapid and effective chemistry space filter in recognizing a druglike database. J Chem Inf Model 2005;45:856-862.
64. Burkhard P, Hommel U, Sanner M, Walkinshaw MD. The discovery of steroids and other FKBP inhibitors using a molecular docking program. J Mol Biol 1999;287:853-858.
65. Frimurer TM, Bywater R, Naerum L, Lauritsen LN, Brunak S. Improving the odds in discriminating "drug-like" from "non drug-like" compounds. J Chem Inf Comput Sci 2000;40:1315-1324.
66. Gozalbes R, Simon L, Froloff N, Sartori E, Monteils C, Baudelle R. Development and experimental validation of a docking strategy for the generation of kinase-targeted libraries. J Med Chem 2008;51:3124-3132.
67. Haraki KS, Sheridan RP, Venkataraghavan R, Dunn DA, McCulloch R. Looking for Pharmacophores in 3-D Databases: Does Conformational Searching Improve the Yield of Actives? Tetrahedron Comput Methodol 1990;3:565-573.
68. Gohlke H, Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int 2002;41:2644-2676.
69. Leach, AR (Ed): Molecular Modelling: Principles and Applications. (2nd Edn). Prentice, New York, 2001, p 744. Prentice Hall.
70. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol 2004;337:1161-1182.
71. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov 2004;3:935-949.
72. Dias R, Filgueira de Azevedo W. Molecular Docking Algorithms. Curr Drug Targets 2008;9:1040-1047.
73. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy evaluation. J Comput Chem 2004;13:505-524.
74. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 2004;25:265-284.
75. Kahraman A, Morris RJ, Laskowski RA, Thornton JM. Shape variation in protein binding pockets and their ligands. J Mol Biol 2007;368:283-301.
76. Oshiro CM, Kuntz ID, Knegtel RMA. Molecular docking and structure-based design. In P. von Rague Schleyer et al. (Eds.): Encyclopedia of Computational Chemistry. Wiley, New York, 1998, pp 1606-1613.
77. Wang R, Lu Y, Wang S. Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 2003;46:2287-2303.
78. Friesner RA, Banks JL, Murphy RB et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004;47:1739-1749.
79. Klebe G, Mietzner T. A fast and efficient method to generate biologically relevant conformations. J Comput Aided Mol Des 1994;8:583-606.
80. Totrov M, Abagyan R. Flexible ligand docking to multiple receptor conformations: a practical alternative. Curr Opin Struct Biol 2008;18:178-184.
81. Hartmann C, Antes I, Lengauer T. Docking and scoring with alternative side-chain conformations. Proteins 2009;74:712-726.

82. Böhm, HJ. 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 1994;8:243-256.
83. Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB. Consensus scoring for ligand/protein interactions. J Mol Graph Model 2002;20:281-295.
84. Suresh PS, Kumar A, Kumar R, Singh VP. An in silico [correction of insilico] approach to bioremediation: laccase as a case study. J Mol Graph Model 2008;26:845-849.
85. Kellenberger E, Ridrigo J, Muller P, Rognan D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. Proteins 2004;57:225-242.
86. Tomich CH, da Silva P, Carvalho I. Homology modeling and molecular interaction field studies of a-glycosidase as a guide to structure-based design of novel proposed anti-HIV inhibitors. J Comput Aided Mol Des 2005;19:83-92.
87. Kaczanowski, S, Zielenkiewicz P. Why similar protein sequences encode similar three-dimensional structures? Theor Chem Acc 2010;125:543-550.
88. Stigers KD, Soth MJ, Nowick JS. Designed molecules that fold to mimic protein secondary structure. Curr Opin Chem Biol 1999;3:714-723.
89. Rademann J, Jung G. Integrating combinatorial synthesis and bioassays. Science 2000;287:1947-1948.
90. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovery of high affinity ligands for proteins. SAR by NMR. Science 1996;274:1531-1534.
91. Schneider G, Neidhart W, Giller T, Schmid G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. Angew Chem Int Ed Engl 1999;38:2894-2896.
92. Zuegg J, Cooper MA. Drug-likeness and increased hydrophobicity of commercially available compound libraries for drug screening. Curr Top Med Chem 2012;12:1500-1513.
93. Ehrlich P. Present status of chemotherapy. Ber Dtsch Chem Ges 1909; 42:17-47.
94. Wermuth CG, Ganellin CR, Lindberg P, Mitscher LA. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997). Annu Rep Med Chem 1998;33:385-395.
95. Milne GW, Nicklaus MC, Wang S. Pharmacophores in drug design and discovery. SAR QSAR Environ Res 1998;9:23-38.
96. Good AC, Mason JS, Pickett SD. Virtual Screening for Bioactive Molecules. Methods Princ Med Chem 2000;10:131-159.
97. Ou-Yang SS, Lu JY, Kong XQ, Liang ZJ, Luo C, Jiang H. Computational drug discovery. Acta Pharmacol Sin 2012;33:1131-1140.
98. Kar S, Roy K. QSAR of phytochemicals for the design of better drugs. Expert Opin Drug Discov 2012; 7:877-902.
99. McGregor MJ, Muskal SM. Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. J Chem Inf Comp Sci 1999;39:569-574.
100. Kurogi Y, Güner OF. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. Curr Med Chem 2001;8:1035-1055.
101. Desaphy J, Azdimousa K, Kellenberger E, Rognan D. Comparison and druggability prediction of protein-ligand binding sites from pharmacophore-annotated cavity shapes. J Chem Inf Model 2012;27:2287-2299.
102. Wood DJ, Vlieg J, Wagener M, Ritschel T. Pharmacophore fingerprint-based approach to binding site subpocket similarity and its application to bioisostere replacement. J Chem Inf Model 2012;52:2031-2043.
103. Manly CJ, Louise-May S, Hammer JD. The impact of informatics and computational chemistry on synthesis and screening. Drug Discov Today 2001;6:1101-1110.
104. Xue L, Bajorath J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. Comb Chem High Throughput Screen 2000;3:363-372.
105. Xue L, Godden JW, Bajorath J. Evaluation of descriptors and mini-fingerprints for the identification of molecules with similar activity. J Chem Inf Comput Sci 2000;40:1227-1234.