Computer-Aided Drug Design: Lead Discovery and Optimization

Mingli Xiang[*,1], Yu Cao[1], Wenjie Fan[1,2], Lijuan Chen[*,1] and Yirong Mo[3]

[1]*State Key Laboratory of Biotherapy and Emergency Department, West China Hospital, West China Medical School, Sichuan University, Chengdu 610041, China*

[2]*Chemical Engineering College, Sichuan University, Chengdu 610041, China*

[3]*Department of Chemistry, Western Michigan University, Kalamazoo, Michigan 49008, USA*

Abstract: Over the past decade, there have been remarkable advances in the area of computer-aided drug design (CADD), which has been applied at almost all stages in the drug discovery pipeline. The generation of initial lead compounds and the subsequent optimization aimed at improving potency and pharmacological properties are the core activities among all. The development in these aspects over the past years will be the focus of this review.

Keywords: Binding affinity prediction, computer-aided drug design, *de novo* design, database searching, lead optimization.

INTRODUCTION

It is a life-saving yet painstaking job for pharmaceutical industry to develop a new drug from the viewpoints of time, capital investment and risks being borne. It is now well documented that the typical drug development cycle from concept to market takes about 14 years [1]. The time-consuming process of developing drug is accompanied by a huge capital investment. According to industrial estimates, the amount of investments to bring out a new drug is ranging from 0.8 to 1.7 billion of USD [2]. Although the industry almost quadrupled its investment in new drug development from 1993 to 2007 alone, the output counted in terms of new medical entities reaching the market is not positively proportional to the investment due to low efficiency of drug development and high failure rate [3]. The cost of clinical trials is prohibitively high, as it approximately accounts for 55% of the total investment in a new drug development [4]. Many drug candidates fail at later stages after going through clinical trials and taking significant amount of resources [3]. Consequently, new technologies are needed to shorten the drug discovery cycle, reduce the expenses and avoid risk of failures particularly at later stages.

There are viable approaches to reach those goals. It is believed [5] that drug development could benefit from the advances in genomics, proteomics, and structure genomics. Efficient technologies such as combinatorial chemistry and high throughput screening have been developed with great expectations [5]. Computer-aided drug design (CADD) is also a very effective way to economize the drug development process [4-7]. However, although the achievement of Human Genome Project and advances in proteomics and structure genomics have resulted in the identification of a large number of human proteins as drug targets [8] and the hope for finding means of saving time and lowering cost in the further clarification of novel therapeutic targets is largely anchored on genomic and proteomic tools [9], those developments have not reduced the cost so far [4]. Using combinatorial chemistry methods, libraries with huge number of chemical compounds for lead discovery and optimization can be easily generated in a short time [10, 11]. The applications of combinatorial chemistry have resulted in an increasing demand for high throughput screening which can screen huge libraries of compounds in short periods of time and has become an essential technique in the lead identification phase. As being expected, these efficient technologies have provided a growing number of drug leads [12]. Despite all efforts having been made to greatly increase the number of drug targets and leads, the revolutionary era of drug design is yet to come [7]. Instead, such technologies have caused increasing expenses without correspondingly augmenting the number of successfully launched new drugs [3, 13]. In contrast to this, CADD is the very one of the effective techniques to reduce costs and speed up drug discovery [3, 5-7, 13-19], even though its role should not be overestimated.

The significant advantages of CADD are obvious. Virtual screening (VS) methods which are designed for searching large libraries of compounds *in silico* usually give a much higher hit rate [3, 6, 11, 20, 21] than the traditional high throughput screening (HTS) and the hits from VS appear more drug-like than the ones from HTS [3]. "Computer mouse" can even simulate the immune system of a non-obese diabetic mouse and is far easier to manipulate than its flesh-and-blood counterpart [3]. By the aid of CADD, the cost of drug development could be reduced by up to 50% [14]. There have been many success stories in the use of CADD in the development of new drugs [3, 6, 13-15, 22, 23]. Actually, as shown in Fig. (**1**), CADD tools can now be used in almost every area of drug development, from target identification to lead discovery, from lead optimization to preclinical or clinical trials [3, 11, 14, 21]. On one hand, lead generation and optimization are crucially important steps [24, 25] among all of the phases of drug discovery process. Without lead compounds, drug discovery projects could not be initiated and continued. On the other hand, each area is a vast discipline in itself. It is impossible to inspect exhaustively all areas of CADD applications or every aspects

*Address correspondence to these authors at the # 1 Keyuan Road 4, Gaopeng Street, High Technological Development Zone, Chengdu 610041, P. R. China; Tel: +86-28-85164063; Fax: +86-28-85164060;
E-mail: xiang_mingli@scu.edu.cn (M. Xiang); Tel: +86-28-85164063;
Fax: +86-28-85164060; E-mail: lijuan17@hotmail.com (L. Chen)

1875-5402/12 $58.00+.00

Fig. (1). Drug development pipeline and stages [21]. The yellow color shows the stages in which computer-aided drug design (CADD) tools are involved (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

of an area in one review article. Therefore in this review our discussion will be limited to some parts of lead generation and optimization, the core activities among all stages in the drug development pipeline.

LEAD DISCOVERY

There are numerous ways to generate lead compounds which could be classified into many categories according to different strategies. However, these ways can also conceptually be divided into two categories: de novo design and database searching. The difference between them is that de novo design generates novel molecules with desired pharmacological properties in the binding site of a target protein, whereas database searching identifies existing molecules from libraries of chemical compounds. Therefore, our attention is paid to both de novo design and database searching for lead discovery.

Structure of Target Protein

The knowledge of the three-dimensional structure of a target protein is of great importance. The purpose of de novo design is to generate novel pharmaceutically active agents that match the binding pattern of a particular biological target. The spatial and physicochemical complementarities of the binding site of the target and the agents is an important feature of a good match [10, 26]. The well-defined three-dimensional (3-D) structures of biological macromolecules can be of significant help in the identification of potential drug target and structure-based de novo design. Without them, the structure-based approach cannot be exerted [27].

Usually, there are two ways to obtain the necessary 3-D structures of the specific macromolecules. The prominent way is to access databases. Protein Data Bank (PDB) is such a database where there are a large number of experimentally determined crystal structures of biological macromolecules with good resolutions. It is the primary and reliable source of information about 3-D structures of protein targets and is thought to be the single worldwide archive of structural data of biological macromolecules [28]. As of March 2011, the number of macromolecular structures deposited in PDB reaches 72104. Although this number is increasing rapidly, it is not rare to find that the existing experimentally solved protein structures can not completely meet the need of drug discovery. The main factors are that there exists a huge gap between the number of experimentally determined protein structures and that of protein sequences, and that many proteins can not be crystallized for X-ray diffraction. In the absence of coordinate data for the target of interest, the alternative approach that can be taken is to predict the 3-D structure of the target from its amino acid sequence by employing homology (or comparative) modeling.

The goal of homology modeling is to construct a sounding structure for the target protein from its sequence with an accuracy that is comparable to the best results achieved experimentally. Certainly, the prerequisite of this modeling is that the sequence should be similar to one or a few proteins with known structures. Based on the fact that protein structures are more evolutionarily conserved than their amino acid sequences [29], homology modeling can predict the 3-D structure of a target sequence with reasonable accuracy provided that there is a significant similarity between the sequence of the target protein and that of the template whose 3-D structure has been experimentally determined. It is true that protein structure prediction is a very challenging task. The extremely large space of possible protein structures and the lack of suitable force fields are the bottlenecks especially for proteins of larger sizes [30]. Still, encouraging progress has been observed. Protein Structure Prediction Center has been organizing a bi-annual CASP (Critical Assessment of Protein Structure Prediction) experiments since 1994 which has greatly stimulated the advance of protein modeling. It has been estimated that the predicted 3-D model of a target sequence is highly reliable when there is more than 50% sequence similarity [31]. Some of the best CASP-certified performing methods are implemented as fully automated servers (http://predictioncenter.org/). Together with popular comparative modeling servers SwissModel and MODELLER, these automated servers are available for public use.

Among all modeling methods to build tertiary structures of proteins, homology modeling has been believed to be the most accurate technique [32]. One of the most noteworthy examples of successful protein structure prediction with homology modeling was accomplished by Wissner A. and his colleagues [33]. When they began ambitiously to design inhibitors for epidermal growth-factor receptor (EGFR) kinase in 1992, the structure information about the protein did not exist. They had to construct the homology model of the kinase domain of EGFR on the basis of the cAMP-dependent protein kinase X-ray crystal structure. They further improved their homology model of EGFR kinase by the comparison with the X-ray structures of hematopoietic cell kinase and fibroblast growth factor receptor-1 kinase [34]. In the process of developing 4-anilino-3-cyanoquinoline series of inhibitors, they realized that the N3 atom of 4-anilinoquinazoline inhibitors was important for potency. On the basis of docking the inhibitors to their modeled EGFR kinase, they proposed a binding mode that the N1 atom of the quinazoline forms a hydrogen bond with the hinge region Met-769 and reasoned that the N3 atom must have a specific interaction with the enzyme by acting as a hydrogen bond acceptor. Subsequently a hypothesis that the N3 atom might interact with the enzyme through a water-bridge [33, 34] was made. The importance of the N3 atom for potency was confirmed by displacing it with a carbon

atom having an attached cyano group. This substitution has a similar overall charge distribution that is present in a quinazoline hydrogen-bonded to a water molecule. Consequently, the EGFR kinase inhibitory activities of the 3-cyanoquinoline analogs of quinazoline derivatives were similar to those of the 4-anilinoquinazoline inhibitors [35]. Miraculously, as shown in Fig. (**2**), both the proposed binding mode and the conceived concept of water-bridged interaction between the N3 atom of 4-anilinoquinazoline analogues and the EGFR kinase were verified by the later crystal structure of quinazoline-kinase complex [36].

Structure-Based *De Novo* Design

In general, de novo ligand design means ligands to be designed are novel and not in exisiting compound libraries. It is also an effective way to circumvent intellectual property constraints. In principle, it is possible by algorithmic

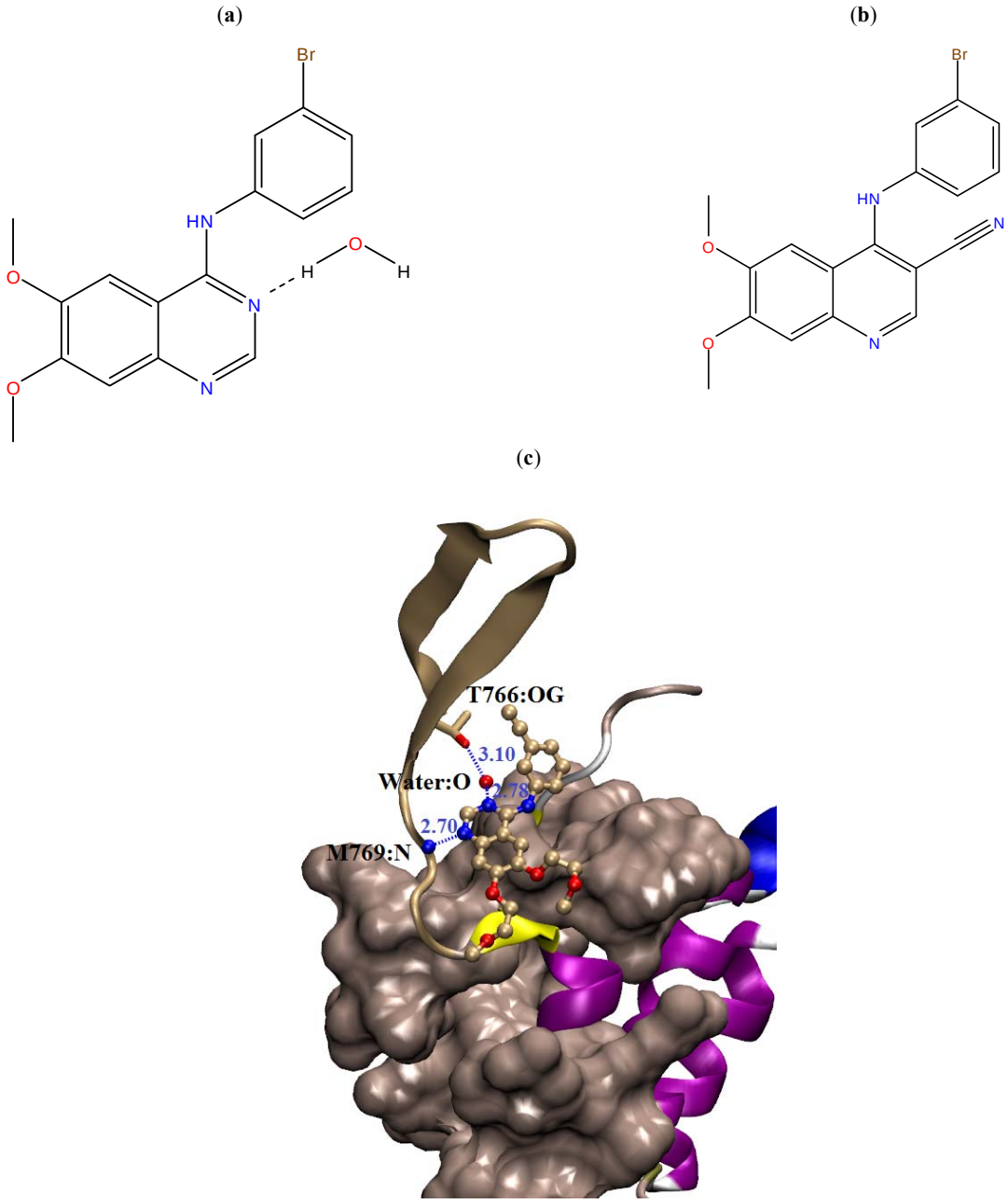

Fig. (2). The verified binding mode for the quinazoline at the APT site in EGFR and the conceived concept of water-bridged interaction between the N3 atom of 4-anilinoquinazoline inhibitor and the EGFR kinase(PDB entry 1M17) [36]. (**a**) 4-anilinoquinazoline inhibitor with speculated water bridge concept and (**b**) its 3-cyanoquinoline derivative [33]. (**c**) Crystal structure of EGFR complexed with a 4-anilinoquinazoline ligand [37] wherein the predicted water-mediated hydrogen bond between N3 of the quinazoline core and the side chain of Thr766 was confirmed. Dashed lines highlight hydrogen bonds and water-bridged interaction between quinazoline core and EGFR kinase. Ligand is shown in CPK representation (atom coloring: gold, carbon; red, oxygen; blue, nitrogen). This figure was generated with the program VMD [38] (for interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper).

connection of molecular fragments to construct a molecule possessing the desired structural and electrostatic properties complementary with the target protein if a high resolution structure of target protein is available [15, 39]. However, the essential issues faced inevitably by de novo design are how to identify the binding site of the target protein, how to assemble possible molecules, how to evaluate their quality, and how to deal with the huge sampling space of novel structures produced in the process of construction [18, 39-43].

Identification of ligand binding site on a protein is a fundamental step for drug discovery. The site could be simply determined if the 3-D structures of one or more protein–inhibitor complexes are known. Otherwise, the possible binding site of the target protein has to be explored by computational methods [44-49].

Once the binding site is identified, de novo ligand design can proceed. There are a variety of specialized programs [18, 24, 50] toward the goal of automated ligand design in binding sites. For the details of the different methods, reader is referred to literatures [18, 50]. The approaches of de novo design can be traditionally divided into two classes [6, 50] according to the strategies for assembling ligands. One class is to construct ligands in a piecewise fashion by positioning and connecting building blocks in binding site. The fundamental building blocks are either single atoms or molecular fragments. The major advantage of atom-based approaches over fragment-based approaches is that almost any structure within rules and constrains could be generated. However, this class of approaches suffers from the computational efficiency as it searches the huge chemical space. Methods using fragment-based building blocks, on the other hand, are computationally more efficient than atom-based approaches. The fragment-based strategies are applicable to the significantly reduced search space. Most probably, the merit of atom-based ways is balanced by their in-native demerit [27]. Many of the early de novo design methods were atom-based [18, 27, 50]. But now fragment-based strategies are getting popular. Even some earlier atom-based approaches were later expended to incorporate fragment building blocks [27, 50].

The second class of de novo ligand design is to grow ligands by placing a seed within the site and growing the seed iteratively in a stepwise fashion [6, 18, 50]. The position of the seed, which could be selected by user [6], program, or determined by a pre-docked seed [18], or identified by computational techniques [51], is one of the interaction sites (hydrogen bonds, electrostatic and hydrophobic interactions) within the ligand binding site. Structure is then grown gradually from the initial seed by appending fragments. It is the way of growing that it tries to provide suitable interactions between the growing structure and the binding site at each step and often results in useful structures. Certainly growing strategy has its associated weakness. It usually leads to molecules that are difficult to synthesize [43]. In addition, this approach may run into difficulties [18] when the binding site is separated by a large gap.

Needless to say, it is a central task to assess the quality of identified lead compounds. That is generally represented by the binding affinity on the computational side. Both lead discovery and optimization stages involve this essential task. Therefore, it is inappropriate to classify it into any stage. The binding affinity issue thus will be addressed in a subsection in the following part of the paper.

The potentially huge number of structures which would be considered in the process of de novo ligand design has been a big challenge for de novo lead generation. There are a large number of ways to link fragments or grow seed. That results in a tremendous number of theoretically possible topologies. Even for a single topology, there are a variety of conformations. It is estimated that the search space for de novo design compromises about 10^{60}-10^{100} compounds [13, 18, 27, 52]. De novo design has to tackle this issue of combinatorial explosion and it is crucial to meaningfully reduce the vast search space. Limiting the scope of the structure space to one particular class of ligands could efficiently reduce the search space [39]. By adapting fragment-based approaches, the search space could be significantly reduced [27] to 10^{20}-10^{30}. However, it is still a formidable problem to select suitable drug candidates from such a space which contains astronomically large number of structures.

Combinatorial search techniques are believed to be the effective way to deal with the combinatorial problem by reducing the size of search space and by sampling the space efficiently [18]. They are commonly used in de novo design [18, 27, 50, 52]. Because the combinatorial explosion problems encountered by de novo design are known to be the NP-hard [18], they can not be solved with provably optimal solution and good run time [27]. Nevertheless, combinatorial search approaches do not seek to find the best possible solution by giving up either one or both of the goals. Instead, they try to find 'good' solution in reasonable computational time [18, 52]. Thus the search space could be narrowed and explored efficiently. Simulated annealing [53, 54], genetic algorithms [43, 55, 56], molecular dynamics [16, 57], evolutionary-based [56, 58] and particle swarm algorithms [52] are the well known combinatorial strategies which are commonly used in de novo drug design [50].

Noteworthy, efforts to reduce search space by employing quantum mechanical (QM) methods have been made [59]. The main idea of these efforts is to map the discrete chemical structures onto a continuous hyper-surface so as to reduce the space by a systematic optimization of parameters introduced in the mapping procedure [59]. Although there is still a long way to apply QM methods to de novo design, sufficient attention should be paid to the significant impact of QM methods on CADD due to the ever growing computing power and development of efficient algorithms.

Apart from those essential issues discussed above, all de novo design methods have to taken into account the synthetic feasibility of the constructed structures. In many cases, early de novo design programs did not consider this issue [24, 27, 50, 58, 60] and were prone to produce synthetically difficult compounds. The first effort aimed at dealing with the synthetic accessibility issue was made by Gillet and co-workers in 1995 [24]. In their work, synthetic accessibility was calculated based on the availability of suitable starting materials and structural complexity, and molecules obtained from a de novo design program were ranked according to the estimation of synthetic feasibility. At

the beginning of 21st century, Honma and co-workers [24] constructed a supporting system for chemical structure selection by the aid of database searching. The system could automatically pick out an essential core substructure of each output obtained from de novo design program, and automatically conduct several queries for searching compound databases and reaction databases. Synthetic accessibility was assessed using the results of a database search with different kinds of queries. Even though the issue of synthetic inaccessibility was recognized as early as in 1990s and some work had been done to overcome it, the crucial importance of estimating it was not appreciated until recently. At the American Chemical Society Spring Meeting in 2006, Gasteiger [61] organized a symposium on "de novo Design and Synthetic Accessibility". During the symposium, remarkable results by tackling the problem of synthetic accessibility were reported. Meanwhile, it was also realized that the solution of this kind of problem requires knowledge from different fields and the collaboration of scientists with different backgrounds. As such, in 2009, Gillet and co-workers [60] developed a knowledge-based approach to the de novo design of synthetically feasible molecules. The method is based on reaction vectors which are derived automatically from a database of reactions. Vectors are used to describe the structural changes that take place at the reaction center. This approach has been validated by reproducing known synthetic routes and applied in different drug design scenarios. In addition, a more perceptual strategy to ensure the synthetic accessibility of de novo designed molecules is to specify particular fragment library used to build the molecules and fragment attachment point [43]. Fragments derived from splitting known drug structures by common synthetic reactions [55, 61], for example, could constitute a special set. Although current techniques for overcoming the problem of synthetic inaccessibility still have a number of limitations, they highlight the opportunities for future work.

Database Searching

An alternative to de novo design is the database searching. Lead compounds could be identified by screening database of known molecules using the 3-D structure of a target protein with known binding site. Searching large commercial and in-house libraries is now an essential approach for structure-based lead generation [23] and is expected to play a more important role in future drug discovery efforts [62].

DOCK program is considered [63, 64] the earliest virtual screen program through which the first application of receptor binding site information in database searching was done [65]. Comparative investigation from Domain and co-workers [66] showed that the hit rate from database screening against protein tyrosine phosphatase-1B was surprisingly higher than the rate from HTS which has been the dominant method to identify novel leads for drug development. This exciting result suggested that database searching would be competitive with HTS as a lead generation approach [66]. DOCK is a geometry-based searching algorithm [63]. In the process of searching, ligands are superimposed onto a negative image of the binding site which is represented by spheres generated by SPHGEN [67]. Another example of geometry-based algorithm is FRED (Fast Rigid Exhaustive Docking) [68]. The shape of the protein binding site is effectively determined by Gaussian docking functions [68]. Ligands that do not have sufficient shape complementarity to the binding site are rejected. FRED is regarded as one of the best database screening methods [69] and is free to academic users.

DOCK and FRED use geometric features that are detected based on the shape of the molecule, whereas others select chemical features, such as hydrogen bond donor and/or acceptor atoms [63]. Chemical points in the binding site serve as docking center. FlexX algorithm and GOLD algorithm [70] are classical examples [63] of using chemical features of the acceptor binding site to screen databases.

Pharmacophore-based database searching is another class of strategy which is applied to the identification of lead compounds in the absence of information about the receptor structure. The term pharmacophore refers to the 3D arrangement of the essential features responsible for a drug's biological activity [64, 71]. The obvious advantage of pharmacophore-based methods is their ability to suggest a diverse set of lead compounds which potentially possess a desired biological activity but have totally different chemical scaffolds [71]. Furthermore, the features in a pharmacophore model are highly transparent for medicinal chemists, and models are intuitively understandable [72]. In this strategy, a pharmcophore model is generally derived from a series of active compounds [10, 71, 73] at first. Then, the model is employed as a search query to retrieve potential leads from compound libraries [10, 64, 74].

In the early days, pharmacophore was only used for ligand-base approaches when the receptor structure was not readily available [64]. Presently, 3D structures of proteins and of receptor-ligand complexes are being determined more rapidly than ever. The availability of such structures provides exceptional opportunities for developing detailed and accurate receptor-based pharmacophore models [71, 75, 76]. In addition, a pharmacophore model can also be used in novo design [71, 77, 78] to develop totally novel drugs that satisfy the pharmacophore requirements and for lead optimization [71, 74].

Many successful applications of pharmacophore approaches in computational drug discovery have been reported in recent years [62, 64, 73-75, 79-83] and a few major computational programs for pharmacophore identification have been available [71, 72].

We would like to note that pharmacophore methods are becoming extremely important ones in facilitating drug discovery, but they will not be addressed here, as there have been several excellent review articles [62, 64, 72, 74, 78, 81, 84, 85] specifically on the topic of pharmacophores.

LEAD OPTIMIZATION

Lead optimization is a subsequent stage after lead discovery. After being identified, lead compounds that exhibit less potency would be optimized so that their drug-like properties could be enhanced. In this stage, attention is paid to improving the drug-like properties of a lead compound by making small modification to the lead

structure [60] according to the principle that any incremental change in the chemical structure produces incremental (positive or negative) changes in bioactivity [15]. The properties include the bioactivity and ADMET (absorption, distribution, metabolism, excretion, and toxicity) properties [60].

Lead optimization is often a step-wise process in which synthesis of a series of analogues and test of their biological activities are accompanied by computational analysis with structure-activity relationship (SAR) information. The SAR information could provide suggestions to modify particular areas of the lead molecule [12]. Here our attention is only focused on computational approaches to improve biological activity. At first, a computational parameter to estimate the quality of lead compounds is needed. That is the binding affinity.

Binding Affinity Prediction

It is assumed that the biological activity of a compound is closely related to the affinity of the compound to macromolecular receptor [10]. Hence, on the computational side, the key for lead optimization is to predict accurately the receptor-ligand binding affinities [23]. Factors, such as electrostatic interactions between the ligand and the receptor, contributions from solvation and desolvation, the spatial complementarity of both binding partners, enthalpic and entropic contributions resulted from changes in the number of degrees of freedom and conformational changes of ligand and receptor experienced upon complex formation, are the essential components [10] for determining the affinity. While it is hoped that the method to predict affinities is rapid and reliable, the most accurate ones are the most rigorous yet the most time-consuming. It seems that there exists no consonance, at least at present, between the reliability and efficiency. Therefore, many approaches to evaluate affinities have been approximated for adapting different requirements.

Efficient Approaches

Scoring functions are one type of approaches that can estimate binding affinity efficiently and often used in the lead generation stage. Scoring functions can be broadly grouped into three categories [5, 27, 62, 86, 87]: force-field-based, empiric-based, and knowledge-based. Force field scoring functions predict the binding affinity of a protein-ligand complex by typically adding up individual contributions from non-bond interactions such as van de Waals and electrostatic interactions [86, 88, 89]. The non-bond energy is extremely sensitive to inter-atomic distances. This feature requires the flexibility and adaptability of protein upon ligand binding to be predicted accurately. But current technology can not satisfy this requirement [62]. The absence of solvation and entropic terms is also the weak side of traditional force field scoring functions [86].

Empirical scoring functions assume that binding affinity can be approximated by a sum of several individual unrelated terms [5, 10, 89]. Terms employed by different empirical functions include hydrogen bonding, hydrophobic contacts, rotor terms, desolvation etc. The weighting of terms is approximated by fitting a regression model to a test set of protein-ligand complexes with known binding affinity. The performance of empirical scoring functions is highly dependent on the test set which should be large enough with several hundred entries of diverse type. In addition, the values of binding affinities of the complexes in the set are required to range over several orders of magnitude [90]. Knowledge-based scoring functions are grounded on a statistical analysis of large sets of protein-ligand complex structures [18, 91]. Results obtained by knowledge-based scoring functions are superior to those obtained by force-fields-based approaches [10], and this type of functions have become popular during the past few years [18]. Knowledge-based scoring functions are based on the hypothesis that a sufficiently large data sample can serve to derive rules and general principles inherently stored in this knowledge base [10]. Thus, The size and quality of the knowledge base for deriving the statistical rules and principles have great impact on the accuracy of the scoring functions [62].

Even though, many successful applications could be found for every category of scoring functions to endorse their merits, the whole situation is not satisfying at this moment. In general, however, the correlation between docking scores and experimental binding affinities is unsatisfactory [87].

Accurate Approaches

Many demerits associated with scoring functions (without considering solvation effect and entropy contribution, for instance) could be avoided by using molecular dynamics (MD) simulation technique [7]. The effect of explicit solvent molecules on protein structure could be well studied by employing MD. It is also possible with MD to obtain different thermodynamic parameters, including interaction energies and entropies. Therefore, MD has become one of the most versatile and widely applied computational techniques for investigating biological macromolecules [7]. Among numerous MD-based techniques for computing binding affinities, free energy perturbation (FEP) [92] method and thermodynamic integration (TI) [92] method are believed [7, 14, 87] to be the most rigorous and accurate approaches. But they are also so computationally expensive that they are not widely applied.

Molecular mechanics/Poisson-Boltzmann surface area (MM/PBSA) [93] method and linear interaction energy method (LIE) [94] are representatives of simplified MD approaches. After a number of simple approximations have been invoked, MM/PBSA and LIE could provide relatively good binding affinity value at a moderate computational cost [7, 14]. MM/PBSA method combines molecular mechanics (MM) and continuum solvent approaches to approximate binding affinities. MM/PBSA does have its drawbacks. One of them is that it is difficult for MM/PBSA to predict the entropic contribution to the binding affinity. LIE is a semi-empirical MD approach. It assumes that the binding affinity can be extracted from simulations of the free and bound states of the ligand. However, LIE also has its own drawbacks [7]: two MD simulations are needed. One is for complex, and the other is for free ligand in water. For each particular complex, the empirically determined constant may need to be modified. These approximate procedures have not been proven to be accurate enough to guide lead optimization [23].

Quantum mechanics/molecular mechanics (QM/MM) hybrid method is a newly emerging way to approximate binding affinity. The fundamental idea of QM/MM approaches is to combine the strength of both QM (accuracy) and MM (efficiency) methods to generate a powerful tool for the study of biological systems [59, 95-97]. Usually a biological system is partitioned into two different regions: the inner region that is treated quantum-mechanically and the outer region that is described by a force field. Inner and outer regions are often referred as QM region and MM region, respectively. For a protein-ligand complex, the QM region would naturally be the ligand, and the MM region would be the protein. As for the case of a biological system in which ligand covalently linked to protein, special precautions need be taken [96] at the boundary between the inner and outer region. The treatment of boundary atoms has been well discussed in literature [98].

Originally, QM/MM approaches were the method of choice for modeling reactions in biological systems [96, 99-101]. Now, QM/MM approaches gradually emerge to be a major tool to predict the biological activity of particular inhibitor [102]. It was found that there is a strong correlation between the computed average interaction energies and biological activity of inhibitors [97, 102]. Although it seems impractical to use the QM/MM interaction energy as a scoring function in database screening, the QM/MM interaction energy could be used to improve the identified lead compounds with a reasonable computational cost.

Optimization of Lead

As mentioned above, the purpose of lead optimization is to improve drug-like properties of lead compounds. Quantitative structure-activity relationship (QSAR) models play a vital role in lead optimization especially in the absence of information about the receptor structure. The goal of building QSAR model is to increase efficiency and lower attrition. Classical 2D-QSAR methods can only study datasets of structurally similar ligands [10]. But in 3D-QSAR methods, physicochemical and structural parameters of molecules are used as descriptors to express biological properties (e.g. affinity or selectivity). The prerequisite of building 3D-QSAR model is that bioactive conformations of all ligands have to be considered to be aligned with each other [10]. Comparative molecular field analysis (CoMFA) method is one of the typical 3D-QSAR methods. It employs both interactive graphics and statistical techniques for correlating shapes of ligands with their biological activities [103]. By learning from the information provided by a ligand-training set, CoMFA could yield surprisingly good affinity predictions [104]. But QSAR model are less interpretable, and it is not easy to work out what modification is needed to make a ligand more active [105]. As discussed before, pharmacophore model could be constructed based on the compounds with known bioactivity [106]. pharmacophore model is another kind of way which could be used for optimizing a series of known ligands and making them more drug-like [71, 74].

Compared with traditional 3D-QSAR model and pharmacophore model, a totally different model was employed by W.L. Jorgensen and co-workers [23] for optimizing lead compounds. The difference could be described at least from three aspects. The first one is to build the model of protein-ligand complex. They simulated complexes in the presence of hundreds or thousands of explicit water molecules using Monte Carlo (MC) statistical mechanics [107]. The second aspect is that MC/FEP calculations were used to evaluate the quality of optimized leads. Their work demonstrated that there was a good correlation between the MC/FEP results and the experimental bioactivities [23]. The final one is the FEP-guided way of modifying lead structure. In the process of lead optimization, several structural modifying strategies [23], heterocycle scan, small group scan, and small group and linker refinement, for example, were adopted. With these strategies, initial leads with activities at low-micromolar concentrations have been optimized rapidly to low-nanomolar inhibitors [23]. The important thing is that this optimization method is effective and more interpretable even though it is still a computationally expensive way.

CONCLUSION

Computer aided drug design (CADD) has been involved in almost the whole pipeline of drug development. Over the past years, striking success in the aspects of development and applications of novel approaches has been made. This consequently facilitates both lead discovery and lead optimization which are the core activities among drug development stages. Especially, structure-based de novo drug design has greatly benefitted the continuous progress in protein structure predictions. The light of group consciousness of synthetic accessibility of lead compounds and organized efforts to overcome it has also signaled the bright future of de novo design. In addition, free energy perturbation guided lead optimization has been a huge success by quickly and largely improving bioactivity of lead compounds. Quantum mechanics/molecular mechanics hybrid method has shown a balance point between accuracy and efficiency in evaluating the quality of lead compounds. Although overall computer aided drug design is still in its infancy, it can be expected from so many impressive progresses achieved during the past years that computer aided drug design will play a much more important role in the near future.

REFERENCES

[1] Myers, S.; Baker, A. Drug discovery—an operating model for a new era. *Nat. Biotechnol.,* **2001**, *19*, 727-730.

[2] Adams, C. P.; Brantner, V. V. The real cost of drug development. *Drug Dev.*, **2006**, *25*, 23-24.

[3] Shekhar, C. *In Silico* Pharmacology: Computer-aided methods could transform drug development. *Chem. Biol.,* **2008**, *15*, 413-414.

[4] Meunier, B. Hybrid molecules with a dual mode of action: dream or reality? *Acc. Chem. Res.,* **2008**, *41* (1), 69-77.

[5] Shaikh, S. A.; Jain, T.; Sandhu, G.; Latha, N.; Jayaram, B. From drug target to leads--sketching a physicochemical pathway for lead molecule design *in silico*. *Curr. Pharm. Des.,* **2007**, *13* (34), 3454-3470.

[6] Jorgensen, W. L. The many roles of computation in drug discovery. *Science,* **2004**, *303*, 1813-1818.

[7] Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.,* **2006**, *26* (5), 531-568.

[8] Pang, Y. P. *In silico* drug discovery: solving the "target-rich and lead-poor" imbalance using the genome-to-drug-lead paradigm. *Clin. Pharmacol. Ther.,* **2007**, *81* (1), 30-34.

[9] Wang, S.; Sim, T. B.; Kim, Y.-S.; Chang, Y.-T. Tools for target identification and validation. *Curr. Opin. Chem. Biol.,* **2004**, *8*, 371-377.

[10] Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem. Int. Ed.,* **2002**, *41* (15), 2644-2676.

[11] Liu, R.; Hsieh, C. Y.; Lam, K. S. New approaches in identifying drugs to inactivate oncogene products. *Semin. Cancer Biol.,* **2004**, *14* (1), 13-21.

[12] Reddy, M. R.; Erion, M. D. Computer-aided drug design strategies used in the discovery of fructose 1, 6-bisphosphatase inhibitors. *Curr. Pharm. Des.,* **2005**, *11* (3), 283-294.

[13] Song, C. M.; Lim, S. J.; Tong, J. C. Recent advances in computer-aided drug design. *Brief Bioinform.,* **2009**, *10* (5), 579-591.

[14] Tan, J. J.; Cong, X. J.; Hu, L. M.; Wang, C. X.; Jia, L.; Liang, X. J. Therapeutic strategies underpinning the development of novel techniques for the treatment of HIV infection. *Drug Discov. Today,* **2010**, *15* (5-6), 186-197.

[15] Ooms, F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr. Med. Chem.,* **2000**, *7* (2), 141-158.

[16] Amaro, R. E.; Li, W. W. Emerging methods for ensemble-based virtual screening. *Curr. Top. Med. Chem.,* **2010**, *10*, 3-13.

[17] Rao, B. G. Recent developments in the design of specific matrix metalloproteinase inhibitors aided by structural and computational studies. *Curr. Pharm. Des.,* **2005**, *11* (3), 295-322.

[18] Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.,* **2005**, *4* (8), 649-663.

[19] Burkhard, K.; Smith, S.; Deshmukh, R.; MacKerell, A. D., Jr.; Shapiro, P. Development of extracellular signal-regulated kinase inhibitors. *Curr. Top. Med. Chem.,* **2009**, *9* (8), 678-689.

[20] Zhong, S.; Macias, A. T.; MacKerell, A. D., Jr. Computational identification of inhibitors of protein-protein interactions. *Curr. Top. Med. Chem.,* **2007**, *7* (1), 63-82.

[21] Tang, Y.; Zhu, W.; Chen, K.; Jiang, H. New technologies in computer-aided drug design: toward target identification and new chemical entity discovery. *Drug Discov. Today:Technologies,* **2006**, *3* (3), 307-313.

[22] Von, I. M.; Wu, W.; Kok, G. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature,* **1993**, *363*, 418-423.

[23] Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.,* **2009**, *42* (6), 724-733.

[24] Honma, T. Recent advances in de novo design strategy for practical lead identification. *Med. Res. Rev.,* **2003**, *23* (5), 606-632.

[25] Keseru, G. M.; Makara, G. M. Hit discovery and hit-to-lead approaches. *Drug Discov. Today,* **2006**, *11* (15-16), 741-748.

[26] Lengauer, T.; Zimmer, R. Protein structure prediction methods for drug design. *Brief Bioinform.,* **2000**, *1* (3), 275-288.

[27] Schuller, A.; Suhartono, M.; Fechner, U.; Tanrikulu, Y.; Breitung, S.; Scheffer, U.; Gobel, M. W.; Schneider, G. The concept of template-based de novo design from drug-derived molecular fragments and its application to TAR RNA. *J. Comput. Aided Mol. Des.,* **2008**, *22* (2), 59-68.

[28] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Res.,* **2000**, *28* (1), 235-242.

[29] Chothia, C.; Lesk, A. M. The relation between the divergence of sequence and structure in proteins. *EMBO J.,* **1986**, *5* (4), 823-826.

[30] Zhang, Y. Progress and challenges in protein structure prediction. *Curr. Opin. Struc. Biol.,* **2008**, *18* (3), 342-348.

[31] Baker, D.; Sali, A. Protein structure prediction and structural genomics. *Science,* **2001**, *294* (5540), 93-96.

[32] Kiefer, F.; Arnold, K.; Kunzli, M.; Bordoli, L.; Schwede, T. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res.,* **2009**, 37, D387-D392.

[33] Wissner, A.; Mansour, T. S. The development of HKI-272 and related compounds for the treatment of cancer. *Arch. Pharm. Chem. Life Sci.,* **2008**, *341*, 465-477.

[34] Wissner, A.; Berger, D.; Boschelli, D.; Floyd Jr, M.; Greenberger, L.; Gruber, B.; Johnson, B.; Mamuya, N.; Nilakantan, R.; Reich, M. 4-Anilino-6, 7-dialkoxyquinoline-3-carbonitrile inhibitors of epidermal growth factor receptor kinase and their bioisosteric relationship to the 4-anilino-6, 7-dialkoxyquinazoline inhibitors. *J. Med. Chem.,* **2000**, *43* (17), 3244-3256.

[35] Wissner, A.; Brawner Floyd, M.; Rabindran, S.; Nilakantan, R.; Greenberger, L.; Shen, R.; Wang, Y.; Tsou, H. Syntheses and EGFR and HER-2 kinase inhibitory activities of 4-anilinoquinoline-3-carbonitriles: analogues of three important 4-anilinoquinazolines currently undergoing clinical evaluation as therapeutic antitumor agents. *Bioorg. Med. Chem. Lett.,* **2002**, *12* (20), 2893-2897.

[36] Stamos, J.; Sliwkowski, M. X.; Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.,* **2002**, *277*, 46265-46272.

[37] Michalczyk, A.; Kluter, S.; Rode, H. B.; Simard, J. R.; Grutter, C.; Rabiller, M.; Rauh, D. Structural insights into how irreversible inhibitors can overcome drug resistance in EGFR. *Bioorg. Med. Chem.,* **2008**, *16*, 3482-3488.

[38] Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual molecular dynamics. *J. Molec. Graphics,* **1996**, *14* (1), 33-38.

[39] Moon, J. B.; Howe, W. J. Computer design of bioactive molecules: a method for receptor-based de novo ligand design. *Proteins,* **1991**, *11* (4), 314-328.

[40] Bohm, H. J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.,* **1992**, *6* (1), 61-78.

[41] Rotstein, S. H.; Murcko, M. A. GroupBuild: a fragment-based method for de novo drug design. *J. Med. Chem.,* **1993**, *36* (12), 1700-1710.

[42] Barakat, M. T.; Dean, P. M. The atom assignment problem in automated de novo drug design. 4. Tests for site-directed fragment placement based on molecular complementarity. *J. Comput. Aided Mol. Des.,* **1995**, *9* (5), 448-456.

[43] Pegg, S. C.; Haresco, J. J.; Kuntz, I. D. A genetic algorithm for structure-based de novo design. *J. Comput. Aided Mol. Des.,* **2001**, *15* (10), 911-933.

[44] Glick, M.; Robinson, D. D.; Grant, G. H.; Richards, W. G. Identification of ligand binding sites on proteins using a multi-scale approach. *J. Am. Chem. Soc.,* **2002**, *124* (10), 2337-2344.

[45] An, J.; Totrov, M.; Abagyan, R. Comprehensive identification of "Druggable" protein ligand binding sites. *Genome Informatics,* **2004**, *15* (2), 31-41.

[46] Huang, B. MetaPocket: a meta approach to improve protein ligand binding site prediction. *OMICS,* **2009**, *13* (4), 325-330.

[47] Brylinski, M.; Skolnick, J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. USA,* **2008**, *105* (1), 129-134.

[48] Zhong, S.; MacKerell, A. D., Jr. Binding response: a descriptor for selecting ligand binding site on protein surfaces. *J. Chem. Inf. Model.,* **2007**, *47* (6), 2303-2315.

[49] Sung, E.; Kim, S.; Shin, W. Binary image representation of a ligand binding site: its application to efficient sampling of a conformational ensemble. *BMC Bioinformatics,* **2010**, *11*, 256-269.

[50] Loving, K.; Alberts, I.; Sherman, W. Computational approaches for fragment-based and de novo design. *Curr. Top. Med. Chem.,* **2010**, *10* (1), 14-32.

[51] Pirard, B. Insight into the structural determinants for selective inhibition of matrix metalloproteinases. *Drug Discov. Today,* **2007**, *12* (15-16), 640-646.

[52] Hartenfeller, M.; Proschak, E.; Schuller, A.; Schneider, G. Concept of combinatorial de novo design of drug-like molecules by particle swarm optimization. *Chem. Biol. Drug Des.,* **2008**, *72*, 16-26.

[53] Todorov, N. P.; Dean, P. M. Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J. Comput. Aided Mol. Des.,* **1997**, *11* (2), 175-192.

[54] Stahl, M.; Todorov, N. P.; James, T.; Mauser, H.; Boehm, H.-J.; Dean, P. M. A validation study on the practical use of automated de novo design. *J. Comput. Aided Mol. Des.,* **2002**, *16* (7), 459-478.

[55] Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S. LEA3D: a computer-aided ligand design for structure-based drug design. *J. Med. Chem.,* **2005**, *48* (7), 2457-2468.

[56] Dey, F.; Caflisch, A. Fragment-based de novo ligand design by multiobjective evolutionary optimization. *J. Chem. Inf. Model.,* **2008**, *48* (3), 679-690.

[57] Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D. Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.,* **2006**, *16* (3), 663-667.

[58] Liu, Q.; Masek, B.; Smith, K.; Smith, J. Tagged fragment method for evolutionary structure-based de novo lead generation and optimization. *J. Med. Chem.,* **2007**, *50* (22), 5392-5402.

[59] Zhou, T.; Huang, D.; Caflisch, A. Quantum mechanical methods for drug design. *Curr. Top. Med. Chem.,* **2010**, *10*, 33-45.

[60] Patel, H.; Bodkin, M. J.; Chen, B.; Gillet, V. J. Knowledge-based approach to de novo design using reaction vectors. *J. Chem. Inf. Model.,* **2009**, *49* (5), 1163-1184.

[61] Gasteiger, J. De novo design and synthetic accessibility. *J. Comput. Aided Mol. Des.,* **2007**, *21*, 307-309.

[62] Sun, H. Pharmacophore-based virtual screening. *Curr. Med. Chem.,* **2008**, *15* (10), 1018-1024.

[63] Schneidman-Duhovny, D.; Nussinov, R.; Wolfson, H. Predicting molecular interactions *in silico*: II. Protein-protein and protein-drug docking. *Curr. Med. Chem.,* **2004**, *11* (1), 91-107.

[64] Guner, O. F. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr. Top. Med. Chem.,* **2002**, *2* (12), 1321-1332.

[65] Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science,* **1992**, *257*, 1078-1082.

[66] Doman, T. N.; McGovern, S. L.; Witherbee, B. J.; Kasten, T. P.; Kurumbail, R.; Stallings, W. C.; Connolly, D. T.; Shoichet, B. K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J. Med. Chem.,* **2002**, *45* (11), 2213-2221.

[67] Kuntz, I.; Blaney, J.; Oatley, S.; Langridge, R.; Ferrin, T. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.,* **1982**, *161* (2), 269-288.

[68] McGann, M.; Almond, H.; Nicholls, A.; Grant, J.; Brown, F. Gaussian docking functions. *Biopolymers,* **2003**, *68* (1), 76-90.

[69] McGaughey, G.; Sheridan, R.; Bayly, C.; Culberson, J.; Kreatsoulas, C.; Lindsley, S.; Maiorov, V.; Truchon, J.; Cornell, W. Comparison of topological, shape, and docking methods in virtual screening. *J. Chem. Inf. Model.,* **2007**, *47* (4), 1504-1519.

[70] Jones, G.; Willet, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.,* **1997**, *267* (3), 724-748.

[71] Dror, O.; Shulman-Peleg, A.; Nussinov, R.; Wolfson, H. J. Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. *Curr. Med. Chem.,* **2004**, *11* (1), 71-90.

[72] Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discov. Today,* **2008**, *13* (1-2), 23-29.

[73] Chen, G. S.; Chang, C. S.; Kan, W. M.; Chang, C. L.; Wang, K. C.; Chern, J. W. Novel lead generation through hypothetical pharmacophore three-dimensional database searching: discovery of isoflavonoids as nonsteroidal inhibitors of rat 5 alpha-reductase. *J. Med. Chem.,* **2001**, *44* (23), 3759-3763.

[74] Chou, K. C.; Wei, D. Q.; Du, Q. S.; Sirois, S.; Zhong, W. Z. Progress in computational approach to drug development against SARS. *Curr. Med. Chem.,* **2006**, *13* (27), 3263-3270.

[75] Schuster, D.; Nashev, L. G.; Kirchmair, J.; Laggner, C.; Wolber, G.; Langer, T.; Odermatt, A. Discovery of nonsteroidal 17beta-hydroxysteroid dehydrogenase 1 inhibitors by pharmacophore-based screening of virtual compound libraries. *J. Med. Chem.,* **2008**, *51* (14), 4188-4199.

[76] Moro, S.; Deflorian, F.; Bacilieri, M.; Spalluto, G. Novel strategies for the design of new potent and selective human A3 receptor antagonists: an update. *Curr. Med. Chem.,* **2006**, *13* (6), 639-645.

[77] Huang, Q.; Li, L. L.; Yang, S. Y. PhDD: a new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility. *J. Mol. Graph Model.,* **2010**, *28* (8), 775-787.

[78] Yang, S., Pharmacophore modeling and applications in drug discovery: challenges and recent advances. *Drug Discov. Today,* **2010**, *15* (11-12), 444-450.

[79] Michaux, C.; de Leval, X.; Julemont, F.; Dogne, J. M.; Pirotte, B.; Durant, F. Structure-based pharmacophore of COX-2 selective inhibitors and identification of original lead compounds from 3D database searching method. *Eur. J. Med. Chem.,* **2006**, *41* (12), 1446-1455.

[80] Seifert, M.; Wolf, K.; Vitt, D. Virtual high-throughput in silico screening. *Biosilico,* **2003**, *1* (4), 143-149.

[81] Oprea, T. I.; Matter, H. Integrating virtual screening in lead discovery. *Curr. Opin. Chem. Biol.,* **2004**, *8* (4), 349-358.

[82] Chiang, Y. K.; Kuo, C. C.; Wu, Y. S.; Chen, C. T.; Coumar, M. S.; Wu, J. S.; Hsieh, H. P.; Chang, C. Y.; Jseng, H. Y.; Wu, M. H.; Leou, J. S.; Song, J. S.; Chang, J. Y.; Lyu, P. C.; Chao, Y. S.; Wu, S. Y. Generation of ligand-based pharmacophore model and virtual screening for identification of novel tubulin inhibitors with potent anticancer activity. *J. Med. Chem.,* **2009**, *52* (14), 4221-4233.

[83] Murumkar, P. R.; Zambre, V. P.; Yadav, M. R. Development of predictive pharmacophore model for in silico screening, and 3D QSAR CoMFA and CoMSIA studies for lead optimization, for designing of potent tumor necrosis factor alpha converting enzyme inhibitors. *J. Comput. Aided Mol. Des.,* **2010**, *24* (2), 143-156.

[84] Adane, L.; Bharatam, P. V. Modelling and informatics in the analysis of P. falciparum DHFR enzyme inhibitors. *Curr. Med. Chem.,* **2008**, *15* (16), 1552-1569.

[85] 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* (1), 37-49.

[86] Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein–ligand docking: Current status and future challenges. *Proteins,* **2006**, *65*, 15-26.

[87] Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlais, R. L. Docking: successes and challenges. *Curr. Pharm. Des.,* **2005**, *11* (3), 323-333.

[88] Pang, Y. P.; Perola, E.; Xu, K.; Prendergast, F. G. EUDOC: a computer program for identification of drug interaction sites in macromolecules and drug leads from chemical databases. *J. Comput. Chem.,***2001**, *22*, 1750-1771.

[89] Ewing, T. J.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0:search strategies for automated molecular docking of flexible molecule database. *J. Comput. Aided Mol. Des.,* **2001**, *15*, 411-428.

[90] McConkey, B. J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand-protein docking. *Curr. Science,* **2002**, *83* (7), 845-856.

[91] Leach, A.; Shoichet, B.; Peishoff, C. Prediction of protein- ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem.,* **2006**, *49* (20), 5851-5855.

[92] Kollman, P. A. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.,* **1993**, *93*, 2395-2417.

[93] Swanson, J. M.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.,* **2004**, *86*, 67-74.

[94] Aquist, J.; Marelius, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen.,* **2001**, *4*, 613-626.

[95] Spiegela, K.; Magistrato, A. Modeling anticancer drug–DNA interactions *via* mixed QM/MM molecular dynamics simulations. *Org. Biomol. Chem.,* **2006**, *4*, 2507-2517.

[96] Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.,* **2009**, *48*, 1198-1229.

[97] Menikarachchi, L. C.; Gascón, J. A. QM/MM approaches in medicinal chemistry research. *Curr. Top. Med. Chem.,* **2010**, *10*, 46-54.

[98] Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A,* **1998**, *102*, 4714-4721.

[99] Proust-De Martin, F.; Dumas, R.; Field, M. A hybrid-potential free-energy study of the isomerization step of the acetohydroxy acid isomeroreductase reaction. *J. Am. Chem. Soc.,* **2000**, *122* (32), 7688-7697.

[100] Gleeson, M. P.; Hillier, I. H.; Burton, N. A. Theoretical analysis of peptidyl a-ketoheterocyclic inhibitors of human neutrophil elastase: Insight into the mechanism of inhibition and the application of QM/MM calculations in structure-based drug design. *Org. Biomol. Chem.,* **2004**, *2*, 2275-2280.

[101] Friesner, R. A. Combined quantum and molecular mechanics (QM/MM). *Drug Disc. Today,* **2004**, *1* (3), 253-260.

[102] Alzate-Morales, J. H.; Contreras, R.; Soriano, A.; Tunon, I.; Silla, E. A computational study of the protein-ligand interactions in CDK2 inhibitors: Using quantum mechanics/molecular mechanics interaction energy as a predictor of the biological activity. *Biophys. J.,* **2007**, *92* (2), 430-439.

[103] Cramer, R.D. III.; Wold, S. B. Comparative molecular field analysis(COMFA). U.S. Patent 5,307,287, April 26, 1994.

[104] Moro, S.; Deflorian, F.; Bacilieri, M.; Spalluto, G. Novel strategies for the design of new potent and selective human A_3 receptor antagonists: An update. *Curr. Med. Chem.,* **2006**, *13*, 639-645.

[105] Lewis, R. A. A general method for exploiting QSAR models in lead optimization. *J. Med. Chem.,* **2005**, *48* (5), 1638-1648.

[106] Chou, K.-C.; Wei, D.-Q.; Du, Q.-S.; Sirois, S.; Zhong, W.-Z. Progress in computational approach to drug development against SARS. *Curr. Med. Chem.,* **2006**, *13*, 3263-3270.

[107] Jorgensen, W. L.; Tirado–Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. *J. Comput. Chem.,* **2005**, *26*, 1689-1700.

Received: January 5, 2011 **Revised: April 5, 2011** **Accepted: April 6, 2011**