Acta Pharmacologica Sinica (2012) 33: 1131–1140
© 2012 CPS and SIMM All rights reserved 1671-4083/12 $32.00
www.nature.com/aps

Review

Computational drug discovery

Si-sheng OU-YANG, Jun-yan LU, Xiang-qian KONG, Zhong-jie LIANG, Cheng LUO, Hualiang JIANG*

Drug Discovery and Design Center, State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

Computational drug discovery is an effective strategy for accelerating and economizing drug discovery and development process. Because of the dramatic increase in the availability of biological macromolecule and small molecule information, the applicability of computational drug discovery has been extended and broadly applied to nearly every stage in the drug discovery and development workflow, including target identification and validation, lead discovery and optimization and preclinical tests. Over the past decades, computational drug discovery methods such as molecular docking, pharmacophore modeling and mapping, *de novo* design, molecular similarity calculation and sequence-based virtual screening have been greatly improved. In this review, we present an overview of these important computational methods, platforms and successful applications in this field.

Keywords: computational drug discovery; target identification; lead discovery

Acta Pharmacologica Sinica (2012) 33: 1131–1140; doi: 10.1038/aps.2012.109; published online 27 Aug 2012

Introduction

The process of novel drug discovery and development is generally recognized to be time-consuming, risky and costly. The typical drug discovery and development cycle, from concept to market, takes approximately 14 years[1], and the cost ranges from 0.8 to 1.0 billion USD[2]. Rapid developments in combinatorial chemistry and high-throughput screening technologies have provided an environment to expedite the drug discovery process by enabling huge libraries of compounds to be screened and synthesized in short time[3, 4]. Although the investment in new drug development has grown significantly in the past decades, the output is not positively proportional to the investment because of the low efficiency and high failure rate in drug discovery[5]. Consequently, various approaches have been developed to shorten the research cycle and reduce the expense and risk of failure for drug discovery. Computer-aided drug design (CADD) is one of the most effective methods for reaching these goals.

CADD is a widely used term that represents computational tools and sources for the storage, management, analysis and modeling of compounds. It covers many aspects of drug discovery, including computer programs for designing compounds, tools for systematically assessing potential lead candidates and the development of digital repositories for studying chemical interactions[6]. In the post-genomic era, benefiting from the dramatic increase in biomacromolecule and small molecule information, computational tools can be applied to most aspects of the drug discovery and development process, from target identification and validation to lead discovery and optimization; the tools can even be applied to preclinical trials[5, 7–9], which greatly alters the pipeline for drug discovery and development. Figure 1 shows a flowchart for the tasks that computational approaches have been applied to and the computational methods used at each stage. The use of computational tools could reduce the cost of drug development by up to 50%[10].

The commonly used computational drug discovery approaches can be categorized into structure-based drug design (SBDD), ligand-based drug design (LBDD) and sequence-based approaches. SBDD methods, such as molecular docking and *de novo* drug design, rely on the knowledge of the structure of the target macromolecule, which are mainly obtained from crystal structures, NMR data and homology models[11]. In the absence of three-dimensional (3D) structures of potential targets, LBDD tools, including quantitative structure-activity relationship (QSAR), pharmacophore modeling, molecular field analysis and 2D or 3D similarity assessment, can provide crucial insights into the nature of the interactions between drug targets and ligands, which allows predictive models that are suitable for lead discovery and optimization to be constructed[12]. In recent years, to deal with situations that neither the target structure nor the ligand information is

* To whom correspondence should be addressed.
E-mail hljiang@mail.shcnc.ac.cn

Received 2012-06-17 Accepted 2012-07-08

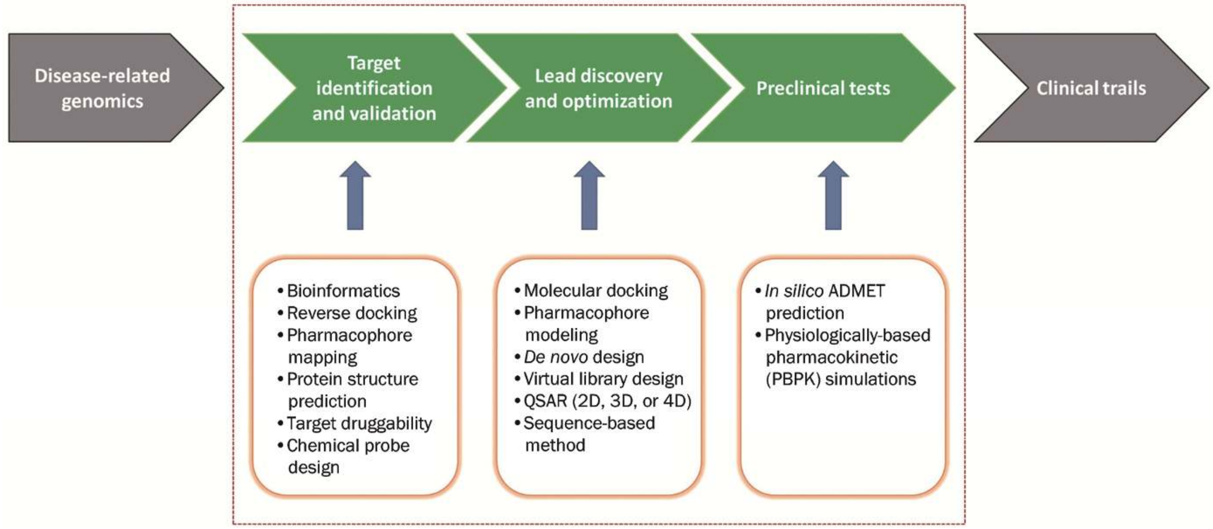

Figure 1. Multiple computational drug discovery approaches that have been applied in various stages of the drug discovery and development pipeline, including target identification and validation, lead discovery and optimization, and preclinical tests.

available, sequence-based approaches that use bioinformatic methods to analyze and compare multiple sequences have been developed to identify potential targets from scratch and to conduct lead discovery[13, 14]. Currently, all single methods are unable to fulfill the practical needs of drug discovery and development. Therefore, combinational and hierarchical strategies that employ multiple computational approaches have been frequently and successfully used.

The efficiency, accuracy and speed of these computational methods largely depend on several technical aspects, including conformation generation and sampling, scoring functions, optimization algorithms, and molecular similarity calculations[7, 11, 15]. In this paper, we focus on these topics and the widely used computational tools in the fields of target identification and lead discovery and address some of the most recent methodologies, platforms and applications.

Methodologies and platforms

Some remarkable methodologies and platforms focused on computational drug discovery and development have been developed and constructed. In this section, several methodologies and platforms that involve target identification, docking-based virtual screening, conformation sampling, scoring functions, molecular similarity calculation, virtual library design and sequence-based drug design are summarized. These aspects are intimately linked, and improvements in any aspect could benefit the others (Figure 2).

Target identification

As the first stage in the drug discovery pipeline, the identification of drug targets from large quantities of candidate macromolecules is both important and challenging[16]. The current major tools for target identification are genomic and proteomic approaches, which are laborious and time-consuming[17]. Therefore, to complement the experimental methods, computational tools and platforms, including reverse docking and pharmacophore mapping, have been developed.

TarFisDock is a web server that identifies drug targets using a reverse docking strategy to seek all possible binding proteins for a given small molecule[18]. The development of TarFisDock was based on the widely used docking program, DOCK (version 4.0)[19, 20]. This platform consists of a front-end web interface written in PHP and HTML with MySQL as database system. DOCK is used as a back-end tool for reverse docking. The advantage of TarFisDock is obvious; it could be a valuable tool for identifying potential targets for a compound with known biological activity, a newly isolated natural product or an existing drug whose pharmacological mechanism is unclear. In addition, this platform is also able to find potential targets that could be responsible for the toxicity and side effects of a drug, which could allow for the prediction of the off-target effects of a drug candidate. Indeed, studies have shown that off-target effects have been largely responsible for the high attrition rate in drug development[21]. Furthermore, TarFisDock could provide valuable information for constructing drug target networks in order to study the drug-target interaction in a more systematic way. The reliability of this methodology has been tested on vitamin E and 4H-tamoxifen by identifying their putative binding proteins. The results indicated that TarFisDock could predict 50% of the reported corresponding targets. However, this method still has certain limitations: (1) the protein entries are not sufficient to cover all the protein information of disease related genomes; (2) the flexibility of the proteins is not considered during the docking procedure; and (3) the scoring function, which was intended to evaluate small molecules, may not be accurate enough for evaluating reverse docking[18].

A web-accessible potential drug target database (PDTD) was constructed for TarFisDock. This database currently contains more than 1100 protein entries with 3D structures

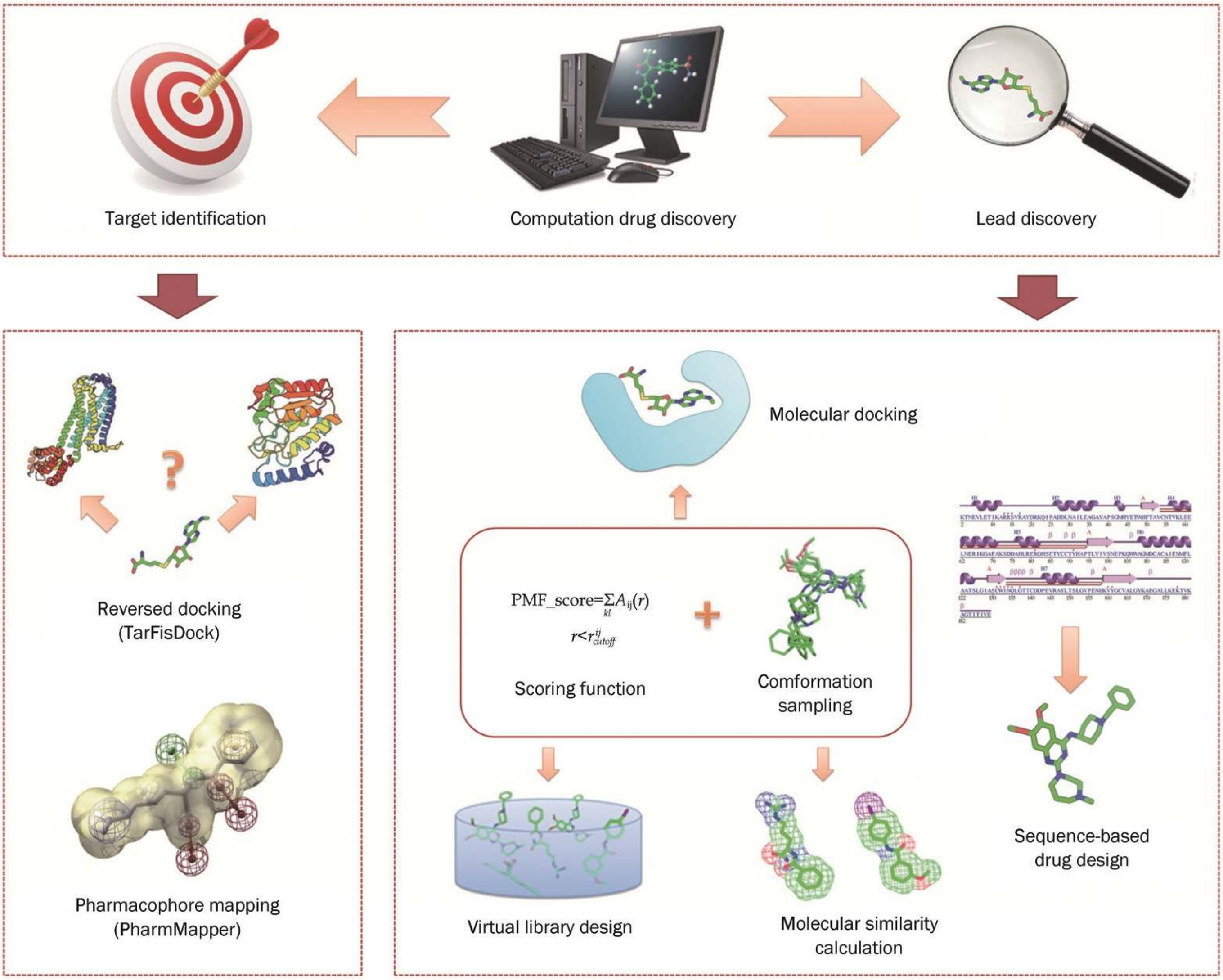

Figure 2. Important methodologies and platforms in the computational drug discovery field introduced and discussed in this article, with a focus on target identification and lead discovery fields.

obtained from the Protein Data Bank. The general information for these proteins was extracted from the literature and several online databases, such as TTD[22], DrugBank[23], and Thomson Pharma. This database contains diverse information on more than 830 potential drug targets, and each drug target has structures in both the PDB and MOL2 formats. Information on related diseases, biological functions and associated signaling pathways has also been collected. All of the targets were classified according to their function and their related diseases. PDTD has a keyword search function for parameters such as the PDB ID, the target name and the disease name[24]. As a comprehensive and unique repository of drug targets, it could be used for *in silico* drug target identification, virtual screening, and the discovery of secondary effects for existing drugs.

Another important issue in target identification is finding the best interaction mode between the potential target candidates and the small molecule probes. In addition to the reverse docking method, pharmacophore modeling and mapping can be used to identify the optimal interaction mode. A pharmacophore model is the spatial arrangement of features essential for a molecule to interact with a specific target receptor. PharmMapper is the first web-based tool to use a ‘reverse’ pharmacophore mapping approach to predict potential drug targets against any given small molecule[25]. However, the PharmMapper server requires a sufficient number of available pharmacophore models that describe the binding modes of known ligands at the binding sites. Thus, a large, in-house database of pharmacophore models annotated with their target information was constructed (PharmTargetDB). The target protein structures in complex with small molecules were carefully extracted from the DrugBank[26], BindingDB[27], PDBBind[28], and PDTD[24] databases, and over 7000 pharmacophore models (covering information for over 1500 drug targets) based on the complex structures were generated. A sequential combination of triangle hashing (TriHash) and genetic algorithm (GA) optimization was adopted to identify the pharmacophore that best fit the task. Benefiting from the highly efficient and robust triangle hash mapping method, PharmMapper is computationally efficient and has the ability to carry out high throughput screens. The algorithm is highly automated, and the interface is user friendly. For experienced users, optional parameters controlling speed and accuracy and the candidate targets subset can be freely customized. The major limitation of the program is that the pharmacophore database only includes drug targets that have PDB structures with a co-crystallized ligand. However, PharmTargetDB is updated periodically as the number of structures deposited in PDB grows[25].

Docking-based virtual screening

Virtual screening based on molecular docking has become one of the most widely used methods of SBDD. The primary criteria for any docking method are docking accuracy, scoring accuracy, and computational efficiency, which are all strongly influenced by the conformational searching method[29, 30]. Molecular docking is a typical optimization problem; therefore, it is difficult to obtain the global optimum solution. Most conformational optimization methods in docking programs can only deal with a single objective, such as the binding energy, shape complementarity, or chemical complementarity. This type of method is not effective for solving real-world problems, which normally involve multiple objectives[31]. Therefore, an optimization algorithm that comprises several objectives and results in more reasonable and robust binding modes between ligands and macromolecules is urgently needed.

A newly developed docking methodology, GAsDock, uses an entropy-based multi-population GA to optimize the binding poses between small molecules and macromolecule receptors[32]. Information entropy was employed in the GA for optimization, and contracted space was used as the convergence criterion, ensuring that GAsDock can converge rapidly and steadily. A validation test docking known inhibitors into the binding pockets of thymidine kinase (TK) and HIV-1 reverse RT indicated that GAsDock is more accurate than other docking programs, such as GOLD[33], FlexX[33], DOCK[33], Surflex[30], and Glide[29]. To increase the accuracy and speed of the process, an improved adaptive genetic algorithm has been developed that supports a flexible docking method. Some advanced techniques, such as multi-population genetic strategy, entropy-based searching technique with self-adaption and quasi-exact penalty, were introduced into this algorithm. A new iteration scheme was also employed in conjunction with these techniques to speed up the optimization and convergence processes, making this method significantly faster than the old method[34]. In addition, two sets of multi-objective optimization (MO) methods, denoted MOSFOM (Multi-Objective Scoring Function Optimization Methodology), that simultaneously consider both the energy score and the contact score were developed. MOSFOM primarily emphasizes a new strategy to obtain the most reasonable binding conformation and increase the hit rates rather than to accurately predicting the binding free energy[31].

Conformation sampling

One of the imperative aspects of drug design and development is to perceive the bioactive conformations of the small molecules that determine the physical and biological properties of the molecules. Many of the drug discovery methods, such as molecular docking, pharmacophore construction and matching, 3D database searching, 3D-QSAR, and molecular similarity analysis, involve a conformational sampling procedure to generate conformations of small molecules in the binding pocket and a scoring phase to rank these conformations. A practical conformation ensemble should guarantee that the conformers are energy reasonable and span the conformational space in an appropriate amount of time. Other sophisticated criteria, such as pharmacophore and binding pocket mapping, have also been implemented to sample the conformers, making the conformation generation process a multi-objective optimization process[35].

A highly efficient conformational generation method named Cyndi, which is based on the multi-objective evolution algorithm (MOEA), has been developed. Using multiple objectives to control energy accessibility as well as geometric diversity, Cyndi is capable of searching the conformational space in nearly constant time and of sampling the *Pareto* frontier at which both the energy and diversity features are favored. The conformers are encoded into GA individuals with information on the dihedral torsions of the rotatable bonds; the VDW and the torsional energy terms are two distinctive objectives for separating the generated conformers in energy space using the Tripos force field[36]. Cyndi ensures that the generated conformation ensemble simultaneously meets multiple criteria, such as low energy and geometric diversity, instead of concentrating on just one criteria[35]. Recently, Cyndi was updated to incorporate the MMFF94 force field to more rationally assess the conformational energy. A comparison between Cyndi and MacroModel integrated in Maestro V7.5 (Schrodinger Inc), focusing on the balance between the sampling depth of the conformational space and the conformational costs with respect to the algorithm method used has been performed. MacroModel was shown to have comparable performance to Cyndi in terms of retrieving the bioactive conformations, while Cyndi performed better at discovering bioactive conformations in the shortest amount of time with regard to the efficiency of the conformation sampling[37].

Scoring function

The scoring function is an essential component in virtual screening. One major scoring method is the knowledge-based scoring method, which typically extracts structural information from experimentally determined protein-ligand complexes and employs the Bolztmann law to transform the atom pair preferences into distance-dependent pairwise potentials[38–41]. The potential of mean force (PMF) scoring function can convert structural information into free energy without any knowledge of the binding affinities and is therefore expected to be more applicable. This method implicitly balances many opposing contributions to binding, such as solvation effects, conformational entropy and interaction enthalpy[40]. Several remarkable methodologies focused on these fields are introduced below.

A kinase family-specific PMF scoring function named kinase-PMF was developed with a kinase data set of 872 complexes from the PDB database to assess the binding of ATP-competitive kinase inhibitors[42]. This scoring function inherits the functional form and atom type of PMF04[43]. Compared to eight other commonly used scoring methods, kinase-PMF had the highest success rate in identifying not only positive compounds from decoys but also crystal conformations. Thus, this method could allow researchers to screen and optimize hit compounds in kinase inhibitor development[42].

An improved PMF scoring function named KScore, which is based on several diverse training sets and a newly defined atom-typing scheme using 23 redefined ligand atom types, 17 protein atom types and 28 newly introduced atom types for nucleic acids, has been developed. In comparison with the existing PMF potentials, such as PMF99 and PMF04, the pairwise potentials for different atom types used in KScore have been significantly improved, particularly in the field of reflecting experimental phenomena, including the interaction distances and the strengths of hydrogen bonding, electrostatic interactions, VDW interactions, cation-π interactions and aromatic stacking. KScore is a powerful tool for distinguishing strong binders from a series of compounds and can be applied to large-scale virtual screening. In addition, further improvements should be possible by modifying the atom-typing scheme and diverse training sets[44]. KScore has been integrated into the previously mentioned molecular docking program GAsDock[32].

On the basis of the concept and formalism of PMF and a novel iteration method, a knowledge-based scoring function named IPMF was developed. This scoring function integrates additional experimental binding affinity information into the knowledge base as complementary data to the generally used structural information. The employed iteration method is to extract the 3D structural information and the binding affinity information in order to yield an "enriched" knowledge-based model. The performance of IPMF was evaluated by scoring a diverse set of 219 protein-ligand complexes and comparing the results to seven commonly used scoring functions. As a result, the IPMF score performs best in the activity prediction test. In addition, when re-ranking binding poses, IPMF also demonstrated marginal improvements over the other evaluated knowledge-based scoring functions. These results suggest that the additional binding affinity information can be used not only for developing scoring functions but also for improving their ability to predict binding affinities. The IPMF approach provides a well-defined scheme to introduce binding information into typical statistical potentials, which may be applicable to other knowledge-based scoring functions[45].

Molecular similarity methods

As the cornerstone of structure-activity relationship (SAR) and structural clustering analysis, molecular similarity is a pivotal concept in LBDD. Similarity-based virtual screening and candidate ranking are considered to be one of the most powerful tools in medicinal chemistry[46, 47] and have been successfully applied in a number of cases. Similarity searching programs can generally be categorized into 2D and 3D similarity according to whether 3D conformation information is considered. 2D similarity methods are efficient for quickly profiling neighboring compounds. However, it may to some extent provide different hits for the same queries as different 2D similarity definitions target different aspects of the information. This method also tends to discover close structural analogues instead of novel scaffold hits[48]. However, 3D similarity methods typically consider multiple aspects of the 3D conformation, including pharmacophores, molecular shapes, and molecular fields. 3D methods can be conveniently used to accomplish scaffold hopping to identify novel compounds.

Based on the pharmacophore matching approach, which was used as the engine of the previously mentioned PharmMapper Server[25], a method named SHAFTS (SHApe-FeaTure Similarity) has been developed for rapid 3D molecular similarity calculation. This method adopts hybrid similarity metrics of molecular shape and colored (or labeled) chemistry groups annotated by pharmacophore features for 3D calculation and ranking in order to integrate the strength of both pharmacophore matching and volumetric similarity approaches. The triplet hashing method is used to enumerate fast molecular alignment poses. The hybrid similarity consists of shape-densities overlaps and pharmacophore feature fit values and is used to score and rank alignment modes. SHAFTS achieved superior performance in terms of both overall and early stage enrichments of known actives and chemotypes compared to other ligand-based methods[48]. SHAFTS has been integrated into ChemMapper Server (unpublished result).

Spherical harmonic (SH) is a set of orthogonal spherical functions that can easily represent the shape of a closed curve surface, such as a molecular surface. SH expansion theory has been successfully applied in virtual screening, protein-ligand recognition, binding pocket modeling, molecular fragment similarity, and so forth. SHeMS is a novel molecular shape similarity comparison method derived from SH expansion. In this method, the SH expansion coefficients are weighted to calculate similarity, leading to a distinct contribution of overall and detailed features to the final score. In addition, the reference set for optimization can be configured by the user, which allows for system-specific and customized comparisons. A retrospective VS experiment on the directory of useful decoys (DUD) database and principal component analysis (PCA) reveals that SHeMS provides dramatically improved performance over the original SH (OSH) and ultra-fast shape recognition (USR) methods[49].

Virtual library construction

De novo drug design aims to chemically fill the binding sites of target macromolecules. One of the critical challenges of this process is to select fragment sets that have the best potential to be parts of new drug leads for a given target. Virtual library construction including focused library, targeted library and primary screening library has been suggested as one way to overcome this challenge[50]. Another challenge is to set up proper criteria for product judgement. To solve this problem, drug-likeness and structural diversity have been introduced into library design to reduce the size and increase the screening efficiency of the constructed libraries.

Focused libraries concentrate on one particular target and are built on the basis of a lead compound or pharmacophore, while targeted libraries are designed to seek drug leads against specific targets[14]. A new efficient approach that adopts the advantages of both focused and targeted libraries and integrates technologies from docking-based virtual screening and drug-like analysis was established to build, optimize and assess focused libraries. A software package named LD1.0 was successfully developed using the new approach[51]. Building blocks are selected from given fragment databases to create a series of virtual libraries. The virtual libraries are then optimized by library-based GA and evaluated on the basis of specified criteria such as docking energy, molecular diversity and drug-likeness. GA retains libraries with higher scores and creates new libraries to form the next generation of focused libraries. Once the termination condition is satisfied, GA optimization ends[51].

Sequence-based drug design

The 3D structures of most proteins have not previously been determined, and many of the proteins do not even have a known ligand. In this situation, neither structure-based methods nor ligand-based methods can be employed to conduct drug discovery and development research. Therefore, a method to predict ligand-protein interactions (LPIs) in the absence of 3D or ligand information is urgently needed. Recently, a sequence-based drug design model for LPI was constructed solely on the basis of the primary sequence of proteins and the structural features of small molecules using the support vector machine (SVM) approach[13]. This model was trained using 15 000 LPIs between 626 proteins and over 10 000 active compounds collected from the Binding Database[52]. In the validation test of this model, nine novel active compounds against four pharmacologically important targets were found using only the sequence of the target. This is the first example of a successful sequence-based drug design campaign[13].

Applications

The newly developed computational drug discovery approaches have been successfully applied in several cases, which suggests that these methods may further emphasize the role of computational drug discovery in the drug R&D workflow.

Application of computational methods to target identification

The combinational strategy of the reverse docking tools TarFisDock and the PDTD database have been successfully used to identify the targets for several bioactive compounds whose *in vivo* targets are unknown. Colonization of the human stomach by the pathogenic bacterium *Helicobacter pylori* is a major cause of gastrointestinal illnesses. However, because of the lack of mature protein targets, discovering anti-*H pylori* agents is a daunting task. Using the active natural product discovered by anti-*H pylori* screening as a probe, potential binding proteins were screened from PDTD using the reverse docking tool TarFisDock. A subsequent homology search indicated that among the 15 candidates discovered by reverse docking, only diaminopimelate decarboxylase (DC) and peptide deformylase (PDF) had homologous proteins in the *H pylori* genome. Enzymatic assays demonstrated that the natural product and one of its analogs are potent inhibitors against *H pylori* PDF (*Hp*PDF), with IC_{50} values of 10.8 and 1.25 μmol/L, respectively. The X-ray crystal structures of *apo*-*Hp*PDF and inhibitor-*Hp*PDF complexes were determined, demonstrating at the atomic level that *Hp*PDF is a potential target for screening new anti-*H pylori* agents[53].

A natural component of ginger, [6]-gingerol, has been reported to exhibit anti-inflammatory and antioxidant properties and exert substantial anticarcinogenic and antimutagenic activities[54]. Despite its potential efficacy in cancer, the mechanism by which it exerts its chemopreventive effects was elusive. By using TarFisDock, [6]-gingerol was docked to each target in PDTD to identify its potential *in vivo* targets. The top 2% of protein hits from the ranked list were considered to be potential target candidates. Subsequent experimental data revealed that [6]-gingerol can effectively suppress tumor growth in nude mice by inhibiting leukotriene A_4 hydrolase (LTA_4H). These findings indicated a crucial role for LTA_4H in cancer and supported the anticancer role of [6]-gingerol in targeting LTA_4H to prevent colorectal cancer[55].

Sphingosine-1-phosphate (S1P) is a sphingolipid metabolite that regulates many cellular and physiological processes, including cell growth, survival, movement, angiogenesis, vascular maturation, immunity and lymphocyte trafficking[56–58]. Although S1P could exert its biological function by binding to five S1P receptors on the cytomembrane, considerable evidence has suggested that S1P has direct intracellular targets. Using an *in silico* target identification approach, S1P was discovered to specifically bind to the histone deacetylases HDAC1 and HDAC2 to regulate histone acetylation[59]. S1P was also found to be a missing cofactor for the E3 ubiquitin ligase TRAF2[60]. These achievements illustrate the pivotal role of S1P in the "inflammation-cancer" chain-related TNFα signaling pathway and in the regulation of gene expression and transcription.

Applications of computational methods in lead discovery

RhoA, one of the most characterized member of the Rho GTPase family, is essential for multiple cellular processes, including cytoskeletal rearrangement, gene expression, membrane trafficking as well as cell adhesion, migration, differentiation, proliferation and apoptosis[61–63]. This protein is a promising target for treating cardiovascular diseases. Using a docking-based virtual screening strategy in conjunction with chemical synthesis and bioassays, a series of first-in-class small molecular RhoA inhibitors were discovered from the SPECS database. A hierarchical docking strategy was adopted: DOCK4.0[19] was used for the initial screening, and the standard DOCK score was used to rank the resulting list; the top 3000 candidates were further docked and ranked by their new scores with Glide in standard precision (SP) mode[29, 64]. In the end, eight compounds showed high RhoA inhibition activities, and two of them showed significant inhibitory effects against PE-induced contraction in thoracic aorta artery rings[65].

Insulin-like growth factor-1 receptor (IGF-1R), a receptor tyrosine kinase, plays a pivotal role in signaling pathways involved in cell growth, proliferation and apoptosis[66]. IGF-1R has been shown to be overexpressed in many human cancers, which suggests it might be a promising target for cancer therapy[67]. Pharmacophore-based virtual screening combined with molecular docking was applied hierarchically to discover IGF-1R inhibitors. Beginning with the complex crystal structure of IGF-1R and its inhibitor, pyridine-2-one, the key interactions between the protein and the ligand at the ATP-binding site were used to construct a pharmacophore model. The SPECS database was screened using this model. The top ranked hits were then docked to the ATP-binding site using Glide[29, 64]. This strategy led to the identification of a series of novel thiazolidine-2,4-dione analogues as potential IGF-1R inhibitors; the molecules demonstrate favorable inhibitory potency and selectivity against IGF-1R over insulin resistance (IR)[68].

A prospective application of the LBDD program SHAFTS is the discovery of novel inhibitors for p90 ribosomal S6 protein kinase 2 (RSK2). Overexpression and aberrant activation of RSK2 have been linked to many human diseases, such as breast cancer, prostate cancer, and human head and neck squamous cell carcinoma[69]. Using the putative 3D conformations of two weakly binding RSK2 inhibitors with moderate activity as the query templates, 16 compounds with IC_{50} lower than 20 μmol/L, which would be missed by conventional 2D methods, were identified via chemotype switching directed by the SHAFTS calculation. The most potent hits show low micromolar inhibitory activities specifically for RSK2, and one compound also exhibits potent anti-migration activity in MDA-MB-231 tumor cells[70].

In another study, a series of novel small molecule inhibitors of cyclophilin A (CypA) were identified using a *de novo* drug design approach. CypA plays an essential role in many biological processes, including enhancing the rate of protein folding/unfolding[71, 72], inhibiting the serine/threonine phosphatase activity of calcineurin[73, 74], facilitating viral replication and infection[75, 76], and inducing neuroprotective/neurotrophic effects[77, 78]. In addition, CypA has been reported to be overexpressed exclusively in cancer cells, particularly in solid tumors, suggesting that CypA is an important regulator of carcinogenesis[79]. The identification of potent, structurally novel small molecule CypA inhibitors is urgently needed, as the most currently available CypA inhibitors are primarily natural products and peptide analogs that may face pharmacokinetic problems. Using the fragment structures of previously discovered CypA inhibitors[80] as building blocks, a focused combinatorial library containing 255 molecules was designed using the LD1.0[51] program. By applying a docking-based virtual screening strategy that targets the binding pocket of CypA, 16 compounds were selected for synthesis and bioassay. According to the experimental results, these compounds all showed high CypA inhibitory activities. The binding affinity and inhibitory activity of the most potent compound among the identified novel CypA inhibitors are approximately 10 times more potent than the best previously known inhibitor[81].

Outlook

Great progress has been made in methodology development and the application of computational drug discovery, resulting in a paradigm change in both industry and academics. Taking advantage of computational methods, potent hits can be obtained in a matter of weeks[82]. Searching for new chemical entities has led to the construction of high quality datasets and libraries that can be optimized for either molecular diversity or similarity. In addition, distributed computing has become more popular in large-scale virtual screening, in part because of increasingly powerful technology[6].

Although it is apparent that computational drug discovery methods have great potential, one should not rely on computational techniques in a black box manner and should beware of the Garbage In-Garbage Out (GIGO) phenomenon. The *in silico* components in research must still be coupled with experiment resources, and computational discovery tools are not substitutions for the more important *in cerebro* component[9, 83, 84]. In the future, in addition to increasing the accuracy and effectiveness of existing technologies, the most important tendency in computational drug discovery field will be the integration of computational chemistry and biology together with chemoinformatics and bioinformatics, which will result in a new field known as pharmacoinformatics[14, 85]. Inspired by the completion of the human genome and numerous pathogen genomes, great efforts will be made to understand the role of gene products in order to exploit their functions, which could be of great help for discovering new drug targets[86]. Computational methods involving target identification will become more attention-getting[87, 88], and designed small molecules will also be extensively used as probes for functional research[89, 90].

Acknowledgements

We gratefully acknowledge financial support from the State Key Program of Basic Research of China grant (2009CB918502), the National Natural Science Foundation of China grants (21021063, 20972174, and 91029704), the Chinese Academy of Sciences grant (XDA01040305) and the National High

Technology Research and Development Program of China (2012AA020302).

References

1 Myers S, Baker A. Drug discovery - an operating model for a new era. Nat Biotechnol 2001; 19: 727–30.

2 Moses H 3rd, Dorsey ER, Matheson DH, Thier SO. Financial anatomy of biomedical research. JAMA 2005; 294: 1333–42.

3 Lahana R. How many leads from HTS? Drug Discov Today 1999; 4: 447–8.

4 Lobanov V. Using artificial neural networks to drive virtual screening of combinatorial libraries. Drug Discov Today Biosilico 2004; 2: 149–56.

5 Shekhar C. *In silico* pharmacology: computer-aided methods could transform drug development. Chem Biol 2008; 15: 413–4.

6 Song CM, Lim SJ, Tong JC. Recent advances in computer-aided drug design. Brief Bioinform 2009; 10: 579–91.

7 Jorgensen WL. The many roles of computation in drug discovery. Science 2004; 303: 1813–8.

8 Xiang M, Cao Y, Fan W, Chen L, Mo Y. Computer-aided drug design: lead discovery and optimization. Comb Chem High Throughput Screen 2012; 15: 328–37.

9 Zhang S. Computer-aided drug discovery and development. Methods Mol Biol 2011; 716: 23–38.

10 Tan JJ, Cong XJ, Hu LM, Wang CX, Jia L, Liang XJ. Therapeutic strategies underpinning the development of novel techniques for the treatment of HIV infection. Drug Discov Today 2010; 15: 186–97.

11 Chen L, Morrow JK, Tran HT, Phatak SS, Du-Cuny L, Zhang S. From laptop to benchtop to bedside: structure-based drug design on protein targets. Curr Pharm Des 2012; 18: 1217–39.

12 Acharya C, Coop A, Polli JE, Mackerell AD Jr. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. Curr Comput Aided Drug Des 2011; 7: 10–22.

13 Wang F, Liu DX, Wang HY, Luo C, Zheng MY, Liu H, *et al*. Computational screening for active compounds targeting protein sequences: methodology and experimental validation. J Chem Inf Model 2011; 51: 2821–8.

14 Tang Y, Zhu WL, Chen KX, Jiang HL. New technologies in computer-aided drug design: Toward target identification and new chemical entity discovery. Drug Discov Today Technol 2006; 3: 307–13.

15 Jorgensen WL. Efficient drug lead discovery and optimization. Acc Chem Res 2009; 42: 724–33.

16 Hajduk PJ, Huth JR, Tse C. Predicting protein druggability. Drug Discov Today 2005; 10: 1675–82.

17 Huan CM, Elmets CA, Tan DC, Li F, Yusuf N. Proteomics reveals that proteins expressed during the early stage of *Bacillus anthracis* infection are potential targets for the development of vaccines and drugs. Genomics Proteomics Bioinformatics 2004; 2: 143–51.

18 Li HL, Gao ZT, Kang L, Zhang HL, Yang K, Yu KQ, *et al*. TarFisDock: a web server for identifying drug targets with docking approach. Nucleic Acids Res 2006; 34: W219–24.

19 Ewing TJ, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des 2001; 15: 411–28.

20 Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982; 161: 269–88.

21 Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, *et al*. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat Rev Drug Discov 2010; 9: 203–14.

22 Chen X, Ji ZL, Chen YZ. TTD: Therapeutic Target Database. Nucleic Acids Res 2002; 30: 412–5.

23 Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, *et al*. DrugBank: a comprehensive resource for *in silico* drug discovery and exploration. Nucleic Acids Res 2006; 34: D668–72.

24 Gao ZT, Li HL, Zhang HL, Liu XF, Kang L, Luo XM, *et al*. PDTD: a web-accessible protein database for drug target identification. BMC Bioinformatics 2008; 9: 104.

25 Liu XF, Ouyang SS, Yu BA, Liu YB, Huang K, Gong JY, *et al*. PharmMapper server: a web server for potential drug target identification using pharmacophore mapping approach. Nucleic Acids Res 2010; 38: W609–14.

26 Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, *et al*. DrugBank: a knowledgebase for drugs, drug actions and drug targets. Nucleic Acids Res 2008; 36: D901–6.

27 Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 2007; 35: D198–201.

28 Wang R, Fang X, Lu Y, Wang S. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. J Med Chem 2004; 47: 2977–80.

29 Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, *et al*. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem 2004; 47: 1739–49.

30 Jain AN. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. J Med Chem 2003; 46: 499–511.

31 Li HL, Zhang HL, Zheng MY, Luo J, Kang L, Liu XF, *et al*. An effective docking strategy for virtual screening based on multi-objective optimization algorithm. BMC Bioinformatics 2009; 10: 58.

32 Li HL, Li CL, Gui CS, Luo XM, Chen KX, Shen JH, *et al*. GAsDock: a new approach for rapid flexible docking based on an improved multi-population genetic algorithm. Bioorg Med Chem Lett 2004; 14: 4671–6.

33 Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem 2000; 43: 4759–67.

34 Kang L, Li HL, Jiang HL, Wang XC. An improved adaptive genetic algorithm for protein-ligand docking. J Comput Aided Mol Des 2009; 23: 1–12.

35 Liu XF, Bai F, Ouyang SS, Wang XC, Li HL, Jiang HL. Cyndi: a multi-objective evolution algorithm based method for bioactive molecular conformational generation. BMC Bioinformatics 2009; 10: 101.

36 Shoichet BK, Leach AR, Kuntz ID. Ligand solvation in molecular docking. Proteins 1999; 34: 4–16.

37 Bai F, Liu XF, Li JB, Zhang HY, Jiang HL, Wang XC, *et al*. Bioactive conformational generation of small molecules: a comparative analysis between force-field and multiple empirical criteria based methods. BMC Bioinformatics 2010; 11: 545.

38 Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 2000; 295: 337–56.

39 Mitchell JBO, Laskowski RA, Alex A, Thornton JM. BLEEP - Potential of mean force describing protein-ligand interactions: I. Generating potential. J Comput Chem 1999; 20: 1165–76.

40 Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 1999; 42: 791–804.

41 Sippl MJ. Boltzmann's principle, knowledge-based mean fields and

protein folding. An approach to the computational determination of protein structures. J Comput Aided Mol Des 1993; 7: 473–501.

42 Xue MZ, Zheng MY, Xiong B, Li YL, Jiang HL, Shen JK. Knowledge-based scoring functions in drug design. 1. Developing a target-specific method for kinase-ligand interactions. J Chem Inf Model 2010; 50: 1378–86.

43 Muegge I. PMF scoring revisited. J Med Chem 2006; 49: 5895–902.

44 Zhao XY, Liu XF, Wang YY, Chen Z, Kang L, Zhang HL, *et al*. An improved PMF scoring function for universally predicting the interactions of a ligand with protein, DNA, and RNA. J Chem Inf Model 2008; 48: 1438–47.

45 Shen QC, Xiong B, Zheng MY, Luo XM, Luo C, Liu XA, *et al*. Knowledge-based scoring functions in drug design: 2. Can the knowledge base be enriched? J Chem Inf Model 2011; 51: 386–97.

46 Muchmore SW, Edmunds JJ, Stewart KD, Hajduk PJ. Cheminformatic tools for medicinal chemists. J Med Chem 2010; 53: 4830–41.

47 Maldonado AG, Doucet JP, Petitjean M, Fan BT. Molecular similarity and diversity in chemoinformatics: from theory to applications. Mol Divers 2006; 10: 39–79.

48 Liu XF, Jiang HL, Li HL. SHAFTS: a hybrid approach for 3D molecular similarity calculation. 1. Method and assessment of virtual screening. J Chem Inf Model 2011; 51: 2372–85.

49 Cai CQ, Gong JY, Liu XF, Jiang HL, Gao DQ, Li HL. A novel, customizable and optimizable parameter method using spherical harmonics for molecular shape similarity comparisons. J Mol Model 2012; 18: 1597–610.

50 Drewry DH, Young SS. Approaches to the design of combinatorial libraries. Chemometr Intell Lab Syst 1999; 48: 1–20.

51 Chen G, Zheng SX, Luo XM, Shen JH, Zhu WL, Liu H, *et al*. Focused combinatorial library design based on structural diversity, druglikeness and binding affinity score. J Comb Chem 2005; 7: 398–406.

52 Liu TQ, Lin YM, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res 2007; 35: D198–201.

53 Cai JH, Han C, Hu TC, Zhang J, Wu DL, Wang FD, *et al*. Peptide deformylase is a potential target for anti-*Helicobacter pylori* drugs: Reverse docking, enzymatic assay, and X-ray crystallography validation. Protein Sci 2006; 15: 2071–81.

54 Surh Y. Molecular mechanisms of chemopreventive effects of selected dietary and medicinal phenolic substances. Mutat Res 1999; 428: 305–27.

55 Jeong CH, Bode AM, Pugliese A, Cho YY, Kim HG, Shim JH, *et al*. [6]-Gingerol suppresses colon cancer growth by targeting leukotriene A4 hydrolase. Cancer Res 2009; 69: 5584–91.

56 Spiegel S, Milstien S. Sphingosine-1-phosphate: An enigmatic signalling lipid. Nat Rev Mol Cell Biol 2003; 4: 397–407.

57 Chun J, Rosen H. Lysophospholipid receptors as potential drug targets in tissue transplantation and autoimmune diseases. Curr Pharm Des 2006; 12: 161–71.

58 Schwab SR, Cyster JG. Finding a way out: lymphocyte egress from lymphoid organs. Nat Immunol 2007; 8: 1295–301.

59 Hait NC, Allegood J, Maceyka M, Strub GM, Harikumar KB, Singh SK, *et al*. Regulation of histone acetylation in the nucleus by sphingosine-1-phosphate. Science 2009; 325: 1254–7.

60 Alvarez SE, Harikumar KB, Hait NC, Allegood J, Strub GM, Kim EY, *et al*. Sphingosine-1-phosphate is a missing cofactor for the E3 ubiquitin ligase TRAF2. Nature 2010; 465: 1084–8.

61 Ridley AJ. Rho family proteins: coordinating cell responses. Trends Cell Biol 2001; 11: 471–7.

62 Sander EE, Collard JG. Rho-like GTPases: Their role in epithelial cell-cell adhesion and invasion. Eur J Cancer 1999; 35: 1302–8.

63 Wheeler AP, Ridley AJ. Why three Rho proteins? RhoA, RhoB, RhoC, and cell motility. Exp Cell Res 2004; 301: 43–9.

64 Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, *et al*. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem 2004; 47: 1750–9.

65 Deng J, Feng EG, Ma S, Zhang Y, Liu XF, Li HL, *et al*. Design and synthesis of small molecule RhoA inhibitors: a new promising therapy for cardiovascular diseases? J Med Chem 2011; 54: 4508–22.

66 Randhawa R, Cohen P. The role of the insulin-like growth factor system in prenatal growth. Mol Genet Metab 2005; 86: 84–90.

67 Khandwala HM, McCutcheon IE, Flyvbjerg A, Friend KE. The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. Endocr Rev 2000; 21: 215–44.

68 Liu XF, Xie H, Luo C, Tong LJ, Wang Y, Peng T, *et al*. Discovery and SAR of thiazolidine-2,4-dione analogues as insulin-like growth factor-1 receptor (IGF-1R) inhibitors via hierarchical virtual screening. J Med Chem 2010; 53: 2661–5.

69 Doehn U, Hauge C, Frank SR, Jensen CJ, Duda K, Nielsen JV, *et al*. RSK is a principal effector of the RAS-ERK pathway for eliciting a coordinate promotile/invasive gene program and phenotype in epithelial cells. Mol Cell 2009; 35: 511–22.

70 Lu WQ, Liu XF, Cao XW, Xue MZ, Liu KD, Zhao ZJ, *et al*. SHAFTS: a hybrid approach for 3D molecular similarity calculation. 2. Prospective case study in the discovery of diverse p90 ribosomal S6 protein kinase 2 inhibitors to suppress cell migration. J Med Chem 2011; 54: 3564–74.

71 Dornan J, Taylor P, Walkinshaw MD. Structures of immunophilins and their ligand complexes. Curr Top in Med Chem 2003; 3: 1392–409.

72 Galat A. Peptidylprolyl *cis/trans* isomerases (immunophilins): biological diversity-targets-functions. Curr Top Med Chem 2003; 3: 1315–47.

73 Liu J, Farmer JD, Lane WS, Friedman J, Weissman I, Schreiber SL. Calcineurin is a common target of cyclophilin-cyclosporin A and FKBP-FK506 complexes. Cell 1991; 66: 807–15.

74 Zuo XJ, Matsumura Y, Prehn J, Saito R, Marchevesky A, Matloff J, *et al*. Cytokine gene expression in rejecting and tolerant rat lung allograft models: analysis by RT-PCR. Transpl Immunol 1995; 3: 151–61.

75 Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP. Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. Cell 1993; 73: 1067–78.

76 Luo C, Luo HB, Zheng SX, Gui CS, Yue LD, Yu CY, *et al*. Nucleocapsid protein of SARS coronavirus tightly binds to human cyclophilin A. Biochem Biophys Res Commun 2004; 321: 557–65.

77 Curtis M, Nikolopoulos SN, Turner CE. Actopaxin is phosphorylated during mitosis and is a substrate for cyclin B1/cdc2 kinase. Biochem J 2002; 363: 233–42.

78 Dawson TM, Steiner JP, Lyons WE, Fotuhi M, Blue M, Snyder SH. The immunophilins, FK506 binding protein and cyclophilin, are discretely localized in the brain: relationship to calcineurin. Neuroscience 1994; 62: 569–80.

79 Choi KJ, Piao YJ, Lim MJ, Kim JH, Ha J, Choe W, *et al*. Overexpressed cyclophilin A in cancer cells renders resistance to hypoxia- and cisplatin-induced cell death. Cancer Res 2007; 67: 3654–62.

80 Li J, Chen J, Gui CS, Zhang L, Qin Y, Xu Q, *et al*. Discovering novel chemical inhibitors of human cyclophilin A: virtual screening, synthesis, and bioassay. Bioorg Med Chem 2006; 14: 2209–24.

81 Li J, Zhang J, Chen J, Luo XM, Zhu WL, Shen JH, *et al*. Strategy for discovering chemical inhibitors of human cyclophilin a: focused library design, virtual screening, chemical synthesis and bioassay. J Comb Chem 2006; 8: 326–37.

82 Singh S, Malik BK, Sharma DK. Molecular drug targets and structure based drug design: A holistic approach. Bioinformation 2006; 1: 314–20.

83 Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. Applying computational modeling to drug discovery and development. Drug Discov Today 2006; 11: 806–11.

84 Kubinyi H. Drug research: myths, hype and reality. Nat Rev Drug Discov 2003; 2: 665–8.

85 Schuffenhauer A, Jacoby E. Annotating and mining the ligand-target chemogenomics knowledge space. Drug Discov Today Biosilico 2004; 2: 190–200.

86 Kopec KK, Bozyczko-Coyne D, Williams M. Target identification and validation in drug discovery: the role of proteomics. Biochem Pharmacol 2005; 69: 1133–9.

87 Paul N, Kellenberger E, Bret G, Muller P, Rognan D. Recovering the true targets of specific ligands by virtual screening of the protein data bank. Proteins 2004; 54: 671–80.

88 Chen YZ, Ung CY. Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand-protein inverse docking approach. J Mol Graph Model 2001; 20: 199–218.

89 Shen JH, Xu XY, Cheng F, Liu H, Luo XM, Shen JK, *et al*. Virtual screening on natural products for discovering active compounds and target information. Curr Med Chem 2003; 10: 2327–42.

90 Stockwell BR. Exploring biology with small organic molecules. Nature 2004; 432: 846–54.