From Drug Target to Leads -

Sketching A Physicochemical Pathway for Lead Molecule Design *In Silico*

S. A. Shaikh, T. Jain, G. Sandhu, N. Latha and B. Jayaram*,

Department of Chemistry &

Supercomputing Facility for Bioinformatics & Computational Biology,

Indian Institute of Technology, Hauz Khas, New Delhi-110016, India

Email: bjayaram@chemistry.iitd.ac.in; Website: http://www.scfbio-iitd.res.in/

Tel: 91-11-26591505, 91-11-26596786. Fax: 91-11-26582037

Accepted for publication in Current Pharmaceutical Design, 2007

Abstract

The discovery of new pharmaceuticals via computer modeling is one of the key challenges in modern medicine. The advent of global networks of genomic, proteomic and metabolomic endeavors is ushering in an increasing number of novel and clinically important targets for screening. Computational methods are anticipated to play a pivotal role in exploiting the structural and functional information to understand specific molecular recognition events of the target macromolecule with candidate hits leading ultimately to the design of improved leads for the target. In this review, we sketch a system independent, comprehensive physicochemical pathway for lead molecule design focusing on the emerging *in silico* trends and techniques. We survey strategies for the generation of candidate molecules, docking them with the target and ranking them based on binding affinities. We present a molecular level treatment for distinguishing affinity from specificity of a ligand for a given target. We also discuss the significant aspects of drug absorption, distribution, metabolism, excretion and toxicity (ADMET) and highlight improved protocols required for higher quality and throughput of *in silico* methods employed at early stages of discovery. We present a realization of the various stages in the pathway proposed with select examples from the literature and from our own research to demonstrate the way in which an iterative process of computer design and validation can aid in developing potent leads. The review thus summarizes recent advances and presents a viewpoint on improvements envisioned in the years to come for automated computer aided lead molecule discovery.

Keywords: Computational drug discovery, *In silico* drug design, Binding affinity, Binding specificity, ADMET

Table of Contents

From Drug Target to Leads -

Sketching A Physicochemical Pathway for Lead Molecule Design *In Silico*

1. Introduction

Drug discovery and development is a cost and time intensive process involving many considerations in molecular design, synthesis, testing and evaluation of drug effects ranging from local interactions at the molecular/cellular level to global effects on the organism and population. Only 20% of drug discovery projects are reported to lead to a clinical candidate and only 10% of the compounds that enter clinical development achieve registration. An analysis of the reasons for this apparently low success reveals that poor pharmacokinetics, toxicity and lack of efficacy are the major factors responsible for failures [1]. Issues like target specificity and affinity, drug delivery, toxicity, side effects etc. must be dealt with in parallel for improving the success rates. It is now well documented that the number of years to bring out a drug from conception to market is approximately 8-10 years, costing on an average US $1.2 billion to $1.4 billion and above per drug [2]. The involvement of genomics [3], proteomics [4], bioinformatics [5] and efficient technologies like, combinatorial chemistry [6], high throughput screening (HTS) [7], virtual screening [7], *in vitro, in silico* ADMET screening [8], *de novo* and structure-based [9] drug design serves to expedite as well as economize the modern day drug discovery process.

Structure based computational drug design methods mainly focus on the design of molecules for a target site with known three dimensional structure followed by a determination of their affinity for the target, based on which a set of hits are obtained [10-12]. The process of structure-based drug design is an iterative one and often proceeds through multiple cycles before an optimized lead goes into clinical trials [13,14]. High throughput screening is the physical screening of large libraries of chemical compounds against a biological target and is still the dominant technique in drug discovery. Virtual screening forms an alternative approach

and uses computer-based methods to screen large chemical libraries targeted towards a biological receptor [15,16] and this task is facilitated significantly by the advent of high performance computing environments, data management software and internet to offer the advantage of delivering new drug candidates more quickly and at lower costs [17-19].

The major roles of computation in drug discovery [20] are; (1) virtual screening and *de novo* design [9,21], (2) evaluation of drug-likeness [22-24] and (3) advanced methods for determining protein-ligand binding [25]. This review summarizes the current computational strategies for rational drug design based on atomic models to generate candidate molecules, to identify good binders/inhibitors for the target with high affinity and specificity and attempts to sketch a pathway for what is conceivable beyond binding to arrive at a lead molecule based on a molecular/structural view of target-drug interactions in a cellular milieu. The plausible steps involved in a molecular level design and development of drug molecules with desired affinity and ADMET profiles are also discussed. In contrast to this, QSAR related computational strategies, which tend to be case specific have been more successful in the prediction of drug efficacy, its metabolism and possible toxic effects [26,27]. QSAR strategies take a more systemic view by building empirical cause to effect relationships - the atomic perspective remains inherent and hidden. In this review, we focus on the development of physicochemical atomic models for lead molecule generation.

We consider the *in silico* drug discovery process as comprising mainly three stages (Fig. 1). Stage I includes identification of a therapeutic drug target and building a heterogeneous small molecule library to be tested against it. This is followed by the development of a virtual screening protocol initialized by either docking of the small molecules from the library or building these structures in the active site employing *de novo* design methods. The next step is binding affinity prediction / scoring and optimization of the set of molecules until the desired binding affinity is achieved. Following this, molecular simulations can be performed in a select

few cases to obtain a more realistic appreciation of binding affinity and its dependence on solvent, salt and dynamics. This way, a set of molecules with desired affinity are selected as better binders to the target and termed as 'hits'. In Stage II, these selected hits are checked for specificity by docking at binding sites of other known drug targets. The hits that score best for only the target of interest and poorly for all other targets are selected as specific binding molecules. In Stage III, these selected hits are subjected to detailed *in silico* ADMET profiling studies and those molecules that pass these studies are termed as 'leads'.

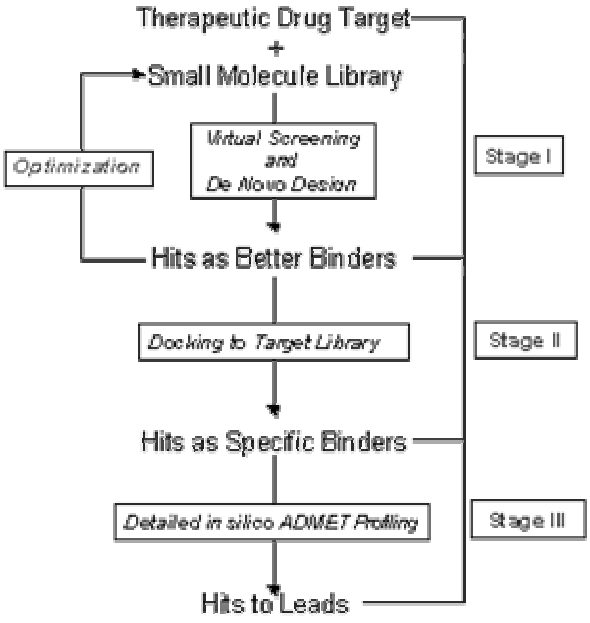

Figure 1. A flowchart outlining a plausible generalized structure-based *in silico* drug discovery strategy.

2. *In silico* identification of hits as better binders: Stage - I

Creating molecules with suitable drug-like properties for a specific target has been a cherished goal of medicinal chemists. Principles of molecular recognition have not advanced much beyond the conventional steric and electrostatic complementarities and hydrophobicity - the relative weightings often beating intuition - thus thwarting automated design of novel inhibitors and therapeutic agents based on a reliable set of rules, even when the three dimensional structure of the drug target is known. The alternative is an energy-based approach,

which conceals the principles but captures the overall thermodynamics of binding nevertheless [28]. Computational structure-based design, spurred by rapid advances in biomolecular target structure determination and computational resources as well as reliable atomic level energy functions, is now gaining ground as a means of generating new pharmaceuticals [29-32]. A computational strategy for identification of hits on the basis of binding affinities is illustrated in Fig. 2 and described in this section.

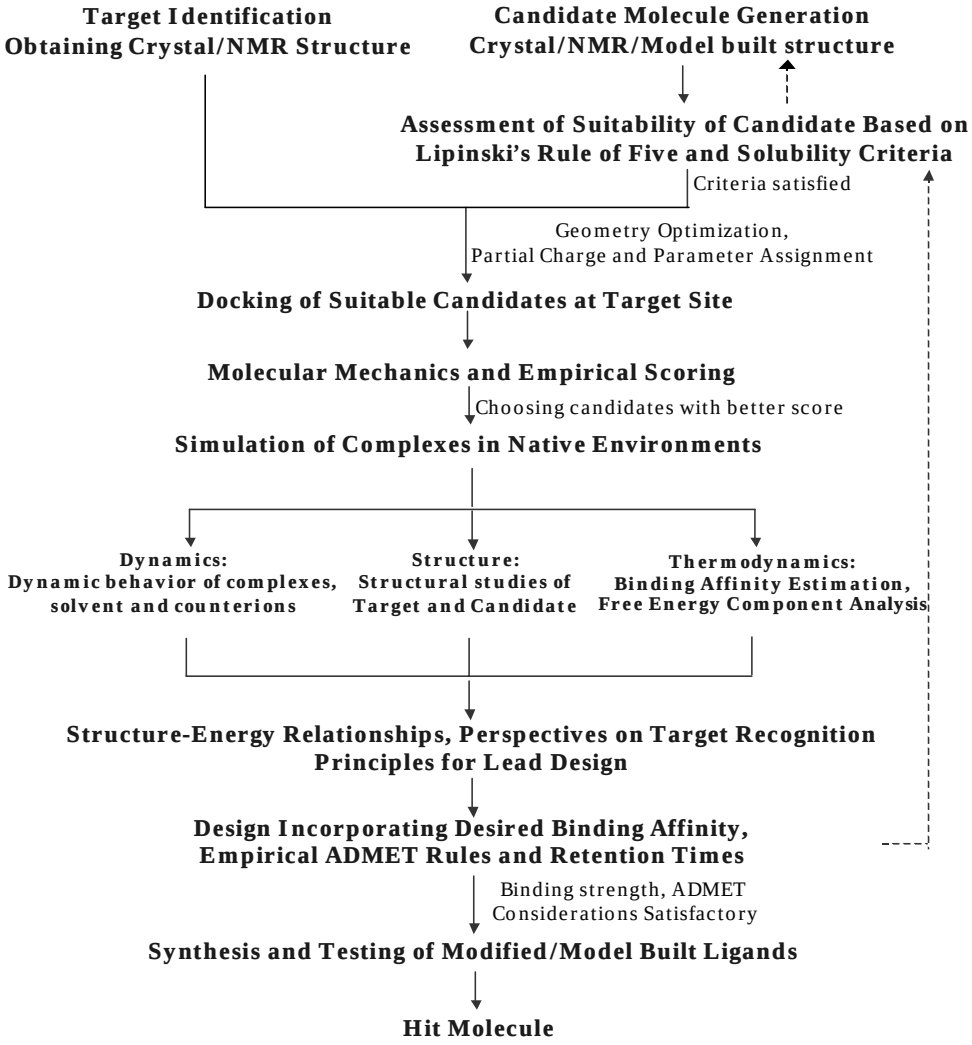

Figure 2. A computational strategy and considerations for obtaining lead-like molecules *in silico* (Stage I)

2.1. Target discovery/selection: Three-dimensional structures of drug targets

Pharmaceutical agents generally exert their therapeutic effect by binding to and regulating the activity of a particular protein or nucleic acid called the drug target. Knowledge

of target characteristics, such as protein / nucleic acid sequence features, structural properties, proteomic profiles, pathway affiliation and roles, and tissue-distribution patterns, is useful for a molecular dissection of the mechanism of action of drugs and for predicting features to guide target discovery and drug design [33,34]. Target discovery/selection is a decision which focuses on finding an agent with a particular biological action that is anticipated to have therapeutic utility and is influenced by a complex balance of scientific, medical and strategic considerations [35,36]. Two crucial questions are answered in deciding whether to accept or reject a new research target. Firstly, what is the probable risk and the likely financial return of the target? Secondly, will the project provide the industry with the right drug, for the right market niche, at the right time and the right place? [37-39].

Current drug therapy rests on about 218 targets which are classified into eight biochemical classes consisting of enzymes, receptors, nuclear receptors, nucleic acids (DNA, RNA and ribosomes), ion channels, antibody targets, transporters and unknown/miscellaneous targets (Fig. 3). There are approximately 6000 drugs currently on the market for these drug targets. Three-dimensional structures for only 130 of these targets are available [40-43].

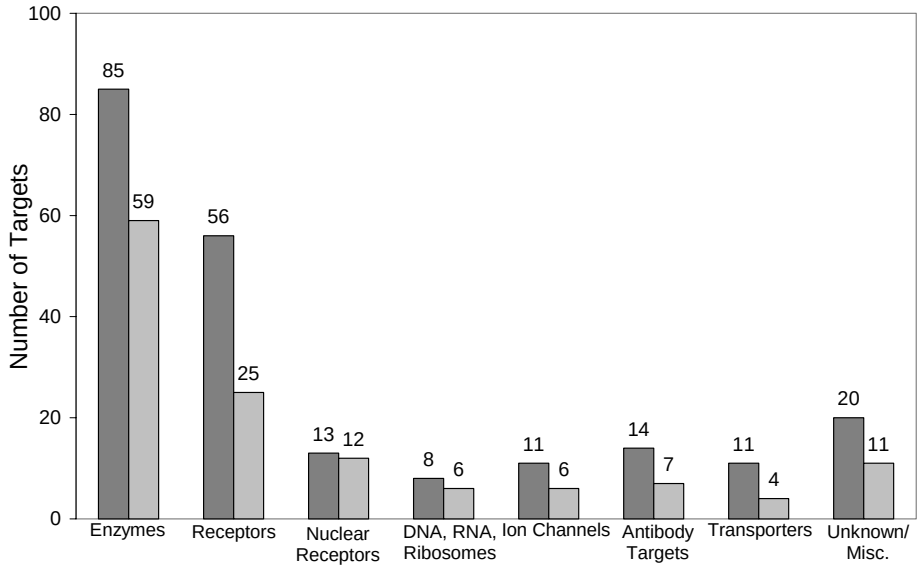

Figure 3: The eight therapeutic drug target classes [40]. The dark bar indicates the total number of known targets for that class. The light bar indicates the number of 3D structures available in PDB for that class. (Source: DrugBank database [41] and Protein Data Bank [42])

After the drug targets are identified partly by computational tools and mostly by experimentation and thorough validation, obtaining the three dimensional structure of the target and identification of the drug binding sites are necessary for proceeding with the computational design of inhibitors/activators for the target. Conventionally, this is achieved through experimental means such as X-Ray or NMR methods [44,45]. It is also possible to adopt the rapidly evolving computational routes such as bioinformatics tools or *ab initio* structure prediction methods [46-49]. In this context, computational strategies for a reliable structure prediction of DNA, RNA and proteins - particularly the membrane bound proteins - are highly relevant and in demand.

2.2. Small molecule generation

Lead-like molecules serve as a starting point to demonstrate the desired biological activity on a validated molecular target.

2.2.1. Current strategies

a. Manual/ Fragment-based Approach: Candidate molecules may be generated manually using simple drawing/building tools available in commercial/ free software or in an automated manner via *in silico* combinatorial methods involving fragments or templates derived from databases. This technique forms the basis of *de novo* design [50-54].

b. Small molecule libraries: Candidate molecules may be retrieved from databases of small molecules for further screening. Target based virtual screening strategies require such small molecule libraries which are utilized in docking [55,56].

A number of small molecule/ drug databases [57-59] have become available for culling structures serially or randomly or through a query system for testing their binding affinity with the target. An ideal small molecular database for this purpose should contain molecules with properties that are uniformly distributed over the ranges considered as appropriate for drugs, to

ensure sufficient sampling of lead-like molecules for any target. Also, if the aim is sampling of all molecules in the database, then in order to keep the process expeditious, the number of molecules in the database should be restricted.

2.2.2 Development of a non-redundant small molecule database. A lead-like molecule database should reflect diversity in chemical and structural properties and contains one or more molecules with suitable affinity to any target and appropriate bioavailability facilitating further chemical elaboration. Working towards this goal, we are developing a non-redundant database of small molecules (NRDBSM) giving special consideration to physicochemical properties and Lipinski's rule of five [60], which determine the solubility, permeability and transport characteristics across membranes. Some of these are molecular weight, number of hydrogen bond donors and acceptors, log P and molar refractivity [61]. The NRDBSM database is aimed specifically at high throughput screening of candidates and their further optimization into successful lead-like molecules hence fixed limits for selected properties have been employed as filters to assemble the database. These precincts have been chosen based on the ranges within which most small molecule databases hold a high percentage of lead-like molecules [62-64]. NRDBSM currently holds close to 17000 molecules with simple structures, low molecular weight, less number of rings and rotatable bonds, low hydrophobicity such that after screening, optimization and consequent increase in molecular complexity, they would show a drift towards 'drug-like' property space [24]. The database is prepared deliberately to avoid biases of normal distribution of these properties. Fig. 4 illustrates the distribution / frequency plots of some properties of interest for absorption and distribution of these small molecules comprising the database. The distribution plots uniformly span partition coefficient logP in -1.0 to 6.0 range, molar refractivity from 40 to 130, molecular weight from 150 to 480, number of hydrogen bond donors from 0 to 3 and hydrogen bond acceptors from 2 to 9. The NRDBSM besides facilitating

focused searches in larger databases once a hit is identified should also help in finding a small number of hits for further optimization [65].

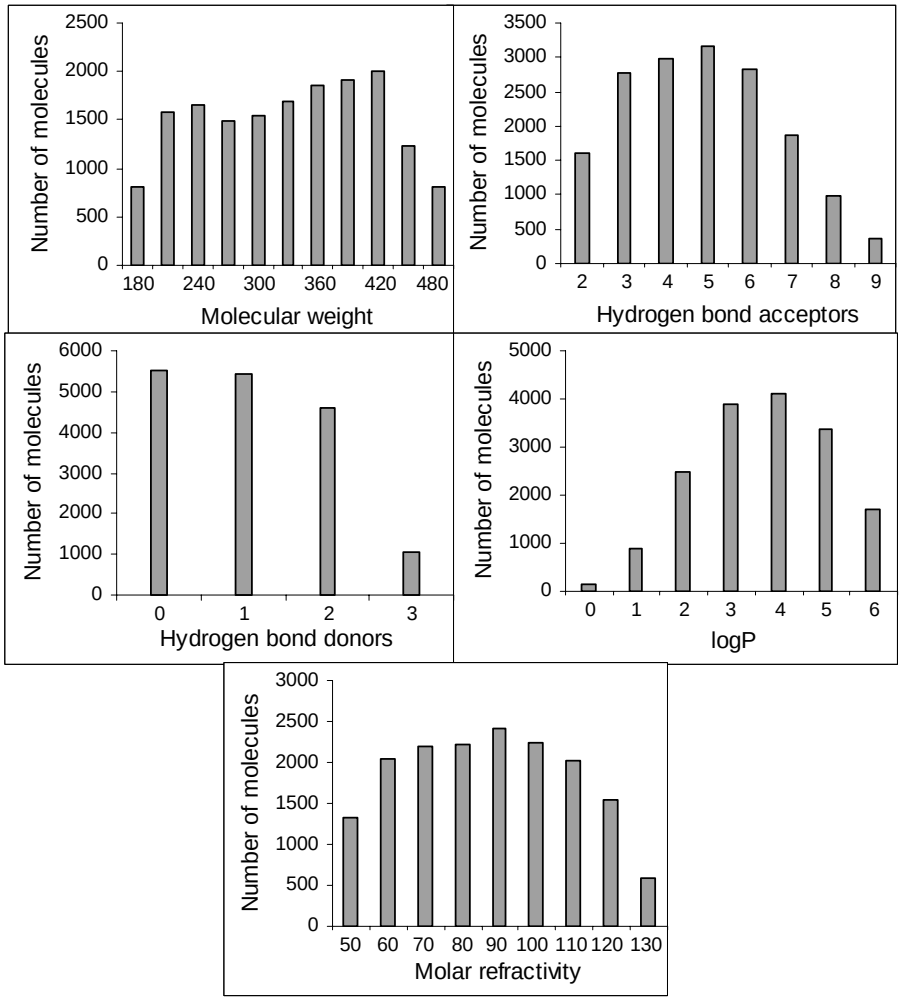

Figure 4. Distribution of molecules according to physicochemical properties in the non-redundant database of small molecules (NRDBSM accessible at: website: www.scfbio-iitd.res.in/drugdesign/software/nrdbsm/)

To filter out probable candidates, apart from strategies like restricted exploration of isomeric structures, selection based on similarity to bioactive compounds [65], one may also introduce a pre-processor embedding active-site information in terms of functional groups required and desired distances between the substituents on potential candidates, volume and shape of the candidates etc. essentially imposing the condition that the candidate be a complementary negative image of the active site. The predominant consideration in most lead-design protocols is activity or ingrained active-site information such that the molecules generated bind well in the active site. Ease of synthesis is also a crucial issue and the intuition

of an organic chemist has to be converted into a computational filter in the *in silico* combinatorial approach.

2.3. Preparation of target and small molecule for energy based processing

Once the set of candidate molecules satisfying the required criteria is obtained, the target and all the candidates are prepared for further computational analysis, energy and force calculations in particular. Current generation molecular mechanical (force field based) methods are extensively validated on biomolecular systems having the advantage over *ab initio* or semi-empirical quantum mechanical methods in being expeditious, and are preferred for modeling and simulation of biomolecular complexes [66,67]. Many force fields are now available for biomolecules, containing pre-calculated partial atomic charges and parameters for proteins and nucleic acids, obviating the need for parameterization for these [68,69]. AMBER [70,71], CHARMM [72], GROMOS [73], OPLS [74] are some of the currently popular force fields developed for simulating biological macromolecules like proteins, nucleic-acids, lipids, carbohydrates and protein-ligand systems. For small molecules, however, rules of transferability are less reliable thus necessitating a derivation of partial atomic charges and geometries using rigorous quantum mechanical methods or fast approximate methods employing semi-empirical calculations, followed by a biomolecular force field compatible parameter assignment appropriate for small molecules [75]. Given the huge dimensionality of chemical space, generating a limited set of appropriate parameters for a wide range of compounds is not a trivial problem. Several force fields like MMFF [76, CVFF [77], CHARMm [78], CFF [79], COMPASS [80], MM2/MM3/MM4 series [81], UFF [82], GAFF [83] among others, have been designed to reproduce internal geometries, vibrations and conformational energies of small molecules. Force fields for metal ions have also been designed [84]. Combination of GAFF with AMBER is one prescription which offers a useful molecular mechanics tool for rational

drug design and other areas where protein-ligand or DNA-ligand simulations are employed. The virtual molecules and the target thus prepared proceed to the next step.

2.4. Docking

The most common computer aided drug design strategy is molecular docking and scoring [85,86]. Docking involves positioning ligands optimally within the target binding site and scoring them for potential activity. Molecular docking is often used in virtual screening methods, whereby large virtual libraries of compounds are reduced in size to a manageable subset, which if successful, includes molecules with high binding affinity to the target receptor [87,88]. Theoretical prediction of the correct placement of ligands at the binding site is a major challenge and is typically attempted using various docking protocols employing search algorithms such as Monte Carlo, genetic algorithms, molecular dynamics, fragment based approach, point complementarity, distance geometry, tabu searches, systematic searches and multiple methods [86,89]. For the target under study, an appropriate docking strategy must be chosen based on its efficiency in cases where the (i) structure of a reference complex is already known, (ii) the active site is known but the structure of a reference complex is not known, (iii) the structure of the target is known but no information on the active site and finally, (iv) the structure of the target is also not known but a pharmacophore model could be built based on known bioactive compounds for the target and/or sequence similarity with other proteins whose structures are known. Protein flexibility is fundamental to understanding the ways in which drugs exert biological effects, their binding site location, binding orientation, binding kinetics, metabolism and transport [90-94]. Some of the most popular rigid and flexible docking approaches are; Prodock [95], ICM [96], MCDOCK [97], DockVision [98], QXP [99], AutoDock [100,101], GOLD [102], DIVALI [103], DOCK [104], FlexX [105], LUDI [106], SLIDE [107], FTDOCK [108] among others which have been proposed for structure-based drug

design. The concepts, applications, success and limitations of various docking protocols have been reviewed by many authors in great detail [109-110].

2.5. Binding affinity prediction / Scoring

A physicochemically rigorous and rapid computational method for binding affinity prediction or scoring will have widespread application in structure-based drug design, virtual screening and *de novo* design protocols. In spite of several recent developments in this area, accurate prediction of binding affinities using computational methods based on an atomic level description of the energetic components of binding, thus transferable across a wide range of targets, has proved to be a major challenge [111]. Computational approaches which utilize the receptor structure information for estimating binding affinities can be pooled into five major classes with respect to their methodological background [25,112,113] - (A) Molecular simulation based approaches, (B) Empirical / force field / additivity based approaches, (C) Knowledge based approaches, (D) Quantum mechanics based approaches and (E) Hybrid approaches.

2.5.1. Binding affinity calculations via scoring functions. The success of docking molecules into a target site and designing lead-like molecules ultimately depends on the accuracy of the scoring function in capturing the correct configuration in the docked structure and in ranking accurately the compounds based on estimates of their relative binding affinities. Some requirements for a good scoring function are: accuracy in structure and affinity prediction, efficiency in virtual screening and speed. Scoring functions are classified into three categories: knowledge based, force field based and empirical [114, 115]. Force field based scoring functions typically account for non-bonded interactions viz. van der Waals (Lennard Jones) and electrostatic (Coulombic) interactions [78, 101, 116-118]. Empirical scoring functions employ a set of terms contributing to the binding energy, which are computed and trained against

experimental data to determine their relative weights. The resulting equation with parameterized terms is verified on a test set and then applied to the systems under study [119-127]. Many terms have been employed by different empirical functions, such as hydrogen bonding, hydrophobic contacts, rotor terms, desolvation etc.. Knowledge based methods are developed via statistical analyses of a large database of protein-ligand structures, where the frequency of occurrence of properties such as interatomic contacts, pairwise potentials etc. are determined across the data set and adopted for scoring [114, 128-132]. The major advantage of such scoring functions is that they are computationally swift. Empirical and knowledge based methods, however, do not guarantee extensions to other classes of molecules that differ from the data set on which the function is parameterized/trained. Comparative evaluations of different docking programs in combination with various scoring functions for their applications in virtual screening have been carried out [133-136] and results show that many of the popular scoring functions are able to select correctly docked from misdocked structures, but correlation with experimental binding affinities still remains a major limiting factor in computational drug discovery [137].

Once a candidate ligand is designed and docked, its interaction/binding energy with the target (protein/nucleic acid) is calculated and compared with that for other proposed compounds and existing ligands, thus allowing assignment of a 'score' to the molecule and facilitating automated selection of ligands with desired binding affinity (Fig. 5).

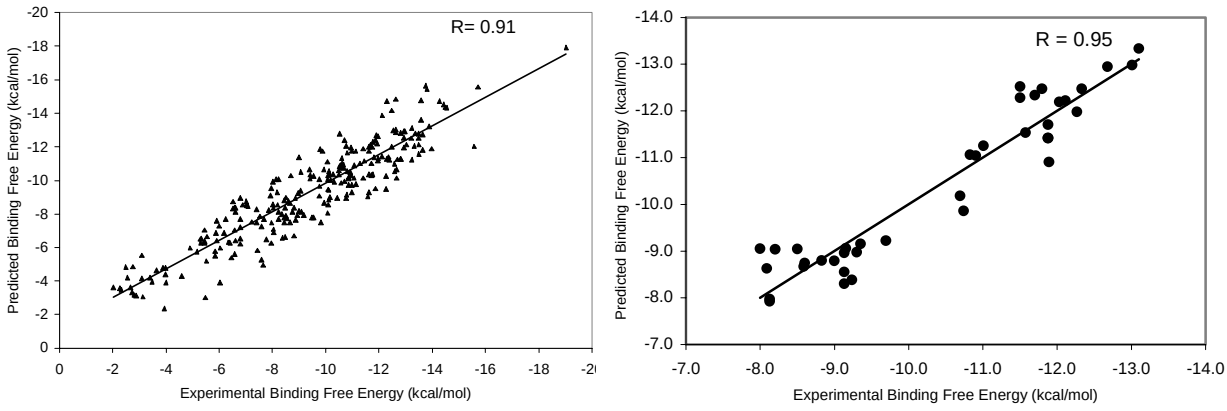

Figure 5. Current accuracies with atomic level energy based scoring functions - Correlation between predicted binding energies and experimentally determined standard free energies of binding (A) Data computed for 251 protein-ligand complexes comprising 60 unique targets using the BAPPL server (available at http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp) (B) Data computed for 39 DNA-ligand complexes comprising 6 unique base sequences using the PreDDICTA server (available at http://www.scfbio-iitd.res.in/ preddicta).

2.5.2. Free energy based methods. The Molecular Mechanics-Generalized Born-Solvent Accessibility (MMGBSA) [138,139], Molecular Mechanics-Poisson Boltzmann-Solvent Accessibility (MMPBSA) [140-142] and the Linear Interaction Energy (LIE) [143] are methods, which elicit binding free energies from structural information and may be used as an alternative to the computationally more intensive free energy simulations. The MMGBSA/MMPBSA approaches are parameterized within the additivity approximation wherein the net free energy change is treated as a sum of a comprehensive set of individual energy components, each with a physical basis and estimated in a force field compatible manner. In the MMGBSA method, molecular mechanical terms are adopted to account for the direct van der Waals and electrostatics (between the target and the candidate molecule), the Generalized Born model for solvation electrostatics and the solvent accessibility for solvation van der Waals and hydrophobic contribution [144-146]. The MMPBSA method differs from this only in the calculation of solvation electrostatics contributions, which are determined as solutions to the Poisson-Boltzmann equation [147]. Extra terms for entropic contributions are often incorporated in these calculations. Both these methods when applied to energy minimized structures have been shown to be computationally rapid and fairly reliable for assessing the contribution of various components to the binding free energy, limited only by the semi-quantitative nature of the results obtained. The LIE [148,149] method calculates the binding affinity as a sum of two parameterized terms that reflect the binding phenomenon. The parameters are obtained from experimental data and multiplied with ensemble averaged energy terms obtained from simulations. The MMGBSA and MMPBSA methods have been effectively applied on both, single structures as well as ensembles obtained from simulations (a preferred choice), allowing the flexibility of choosing speed over accuracy or vice versa. Free energy

simulations [150,151] may be employed in cases where accuracy and theoretical rigor are of utmost importance and computational expense a minor issue. Here, techniques like free energy perturbation, thermodynamic integration etc. are often employed for binding affinity determination. All these methodologies are amenable to further systematic improvements.

2.6. Study of the dynamics of promising target - candidate molecule complexes

Target-candidate molecule complexes with high binding affinity can be further processed in a dynamic environment employing simulation strategies such as molecular dynamics. Though computationally expensive, such simulation strategies provide a route to investigating the effects of conformational flexibility, solvent and salt, and entropic factors. Simulations with explicit solvent are highly time consuming and the time scales may limit probing conformational changes induced by inhibitors or allosteric changes known to occur with activators [152-156]. Crossing conformational barriers higher than thermal energies need special treatment such as simulated annealing which have also become standard protocols [157,158]. Recently, flexible Monte Carlo simulations applied to DNA-drug systems have shown considerable promise in this regard [159].

3. *In silico* processing of identified hits as specific binders: Stage - II

Apart from the requisite binding affinity, a key consideration during drug design is specificity [9-12]. Therapeutic strategies generally require inhibitors that are highly selective for a particular target. However, the molecular features driving selectivity *in vivo* remain only little understood.

Computational tools have been and are continuing to be developed to extract molecular parameters from the large body of ligand binding data responsible for affinity discrimination toward structurally related proteins [160,161]. Traversing on the thermodynamic path, drugs

with low specificity could potentially bind to a large number of targets, which could result in high toxicity. Also, very small amounts of the drug become available to bind to the target, thus requiring higher dosage further increasing the risk of toxic side effects.

A computational strategy for addressing the issue of specificity would be to assess the binding of the candidate molecule with all potential targets in the human cell - not an impossible task in the emerging low cost, high performance computing scenario with reliable scoring functions, improved annotations and mounting structural data. This could be achieved by building a database of possible binding sites for all potential targets and docking the candidate molecules to all these targets followed by binding affinity estimates. High affinity binding to non-target sites translates to low target specificity of the candidate thus indicating potential side effects. Unsuitable candidates could be filtered out on this basis while the remainder further optimized for improved affinity and specificity.

Computational methods have come of age to generate binding affinity columns of a candidate to diverse targets and diverse candidates to the same target. An illustrative example is shown in Fig. 6 where a two-dimensional specificity matrix generated *in silico* for 14 drugs and their corresponding targets representing all currently known classes (Fig. 3) of therapeutic drug targets. Each column in the figure represents binding affinity of a drug to all the 14 targets. Each row represents the affinities of the 14 drugs to a target. If the drugs are specific to the targets, high affinities should occur only along the diagonal and all the off diagonal cells should ideally represent nonspecific (weak) binding. However, if the drugs are not highly target specific, the off-diagonal elements could represent strong binding, which could be used as an indicator for improving drug specificity as well as predicting possible toxicity and side effects. This matrix was generated based on the docking and binding affinity calculation for protein-ligand and DNA-ligand interactions using an in-house software (http://www.scfbio-iitd.res.in/software/drugdesign/bappl.jsp and http://www.scfbio-iitd.res.in/preddicta). It may be

discerned from Fig. 6 that six out of the 14 drugs studied (Drug3, Drug5, Drug6, Drug8, Drug10, Drug11, Drug12) are specific to their corresponding targets (i.e. they do not bind to any other target with a higher affinity – this is indicated by the absence of gray cells in the drug column). The other drugs bind strongly to some non-targets too and could possibly have side effects. Once such indications are obtained from computational analyses, further investigations on toxicity/side-effects can be made and the drug design/delivery process can be modified to ensure higher specificity. The caveat, however, is that such predictions are strongly dependent on the accuracy/efficiency of the docking and binding affinity prediction methods employed. The matrix nonetheless portends the methodological developments to ensue in computer aided drug design.

	Drug1	Drug2	Drug3	Drug4	Drug5	Drug6	Drug7	Drug8	Drug9	Drug10	Drug11	Drug12	Drug13	Drug14
Target1	-5.15	-3.53	-3.59	-3.78	-3.55	-3.63	-3.63	-5.14	-2.86	-1.86	-4.38	-4.4	-5.91	-5.49
Target2	-2.41	-6.67	-2.94	-2.77	-4.87	-2.62	-3.54	-3.42	-2.26	-1.72	-3.13	-4.06	-4.37	-5.02
Target3	-7.09	-6.99	-9.8	-6.93	-7.78	-7.01	-6.84	-8.65	-5.67	-3.65	-7.22	-8.54	-8.65	-9.02
Target4	-3.52	-5.06	-4.01	-6.75	-5.15	-4.29	-5.17	-6.44	-4.54	-3.26	-2.66	-5.06	-3.77	-6.37
Target5	-8.21	-7.71	-6.51	-7.45	-9.63	-8.07	-7.9	-7.11	-7.26	-7.14	-7.49	-8.06	-9.59	-8.64
Target6	-6.24	-6.69	-5.97	-7.17	-6.03	-9.73	-6.78	-8.26	-6.42	-3.76	-7.46	-8.44	-7.97	-9.12
Target7	-2.53	-3.5	-3.23	-2.84	-5.99	-3.74	-6.78	-3.63	-2.73	-3.49	-2.19	-5.13	-1.55	-4.67
Target8	-5.03	-4.51	-7.03	-4.99	-7.32	-5.83	-6.13	-8.99	-4.87	-4.1	-5.09	-7.82	-4.59	-8.2
Target9	-1.12	-1.03	-1.41	-1.4	-1.68	-1.64	-1.33	-1.95	-2.16	-0.4	-1.86	-1.34	-1.47	-2.32
Target10	-8.06	-7.42	-6.79	-7.07	-7.03	-7.53	-7.31	-6.93	-6.63	-9.38	-7.65	-7.8	-8.79	-8.37
Target11	-6.26	-6.35	-7.74	-6.79	-6.88	-6.31	-6.81	-8.61	-5.18	-3.72	-8.83	-8.06	-9.69	-9.44
Target12	-5.84	-5.92	-6.74	-6.48	-7.09	-6.05	-6.67	-8.05	-5.93	-5.44	-5.77	-10.35	-7.94	-8.81
Target13	-4.56	-3.91	-4.52	-3.77	-2.36	-3.85	-3.95	-5.41	-2.01	-1.57	-5.43	-4.93	-13.12	-5.52
Target14	-3.46	-2.61	-4.38	-3.67	-4.59	-4.89	-4.58	-4.45	-3.38	-2.6	-4.9	-3.5	-7.38	-8.88

DIAGONAL ELEMENTS (Black): Drug-Target Affinity OFF-DIAGONAL ELEMENTS (Gray, White): Drug-Non Target Affinity

Figure 6: Specificity Matrix for drugs and their targets/non-targets. Drug 1 corresponds to Target 1, Drug 2 corresponds to Target 2 and so on. Grey cells represent drug binding to non-targets with higher affinity than the original drug-target interaction, thus indicating low specificity. White cells show low affinity.
Target 1 is lymphocyte function-associated antigen LFA-1 (CD11A) (1CQP; Immune system adhesion receptor) and Drug 1 is lovastatin.Target 2 is Human Coagulation Factor (1CVW; Hormones & Factors) and Drug 2 is 5-dimethyl amino 1-naphthalene sulfonic acid (dansyl acid). Target 3 is retinol-binding protein (1FEL; Transport protein) and Drug 3 is n-(4-hydroxyphenyl)all-trans retinamide (fenretinide). Target 4 is human cardiac troponin C (1LXF; metal binding protein) and Drug 4 is 1-isobutoxy-2-pyrrolidino-3[n-benzylanilino] propane (Bepridil). Target 5 is DNA {1PRP; d(CGCGAATTCGCG)} and Drug 5 is propamidine. Target 6 is progesterone receptor (1SR7; Nuclear receptor) and Drug 6 is mometasone furoate. Target 7 is platelet receptor for fibrinogen (Integrin Alpha-11B) (1TY5; Receptor) and Drug 7 is n-(butylsulfonyl)-o-[4-(4-piperidinyl)butyl]-l-tyrosine (Tirofiban). Target 8 is human phosphodiesterase 4B (1XMU; Enzyme) and Drug 8 is 3-(cyclopropylmethoxy)-n-(3,5-dichloropyridin-4-yl)-4-(difluoromethoxy)benzamide (Roflumilast). Target 9 is Potassium Channel (2BOB; Ion Channel) and Drug 9 is tetrabutylammonium. Target 10 is {2DBE; d(CGCGAATTCGCG)} and Drug 10 is Diminazene aceturate (Berenil). Target 11 is Cyclooxygenase-2 enzyme (4COX; Enzymes) and Drug 11 is indomethacin. Target 12 is Estrogen Receptor (3ERT; Nuclear Receptors) and Drug 12 is 4-hydroxytamoxifen. Target 13 is ADP/ATP Translocase-1 (1OKC; Transport protein) and Drug 13 is carboxyatractyloside. Target 14 is Glutamate Receptor-2 (2CMO; Ion channel) and Drug 14 is 2-({[(3e)-5-{4-[(dimethylamino)(dihydroxy)-

lambda~4~-sulfanyl]phenyl}-8-methyl-2-oxo-6,7,8,9-tetrahydro-1H-pyrrolo[3,2-H]isoquinolin-3(2H)-ylidene]amino}oxy)-4-hydroxybutanoic acid.

A candidate molecule could be scanned against the entire genome / proteome in the cell if the sequence specific DNA conformation and the three dimensional structures of all proteins in the target cell are established. Even pharmacophore models can be of help in pressing docking-scoring strategy into service to ensure selectivity for the target. The number of proteins expressed in a particular cell is reported to be around 10000-20000 although the human genome can code for many more proteins. Thus, efforts need to be routed to ensure specificity for target vis-à-vis these cell specific proteins. However, the spatial and temporal issues of gene regulation/ genome expression in cells are only poorly or partly understood.

4. Beyond binding affinities - Towards a molecular treatment of ADMET profiles of candidates: Stage - III

The success of a drug journey through the body is measured in the dimensions of absorption, distribution, metabolism and excretion (ADME) properties (Fig. 7). An ideal oral drug should be rapidly and completely absorbed from the alimentary canal and find its way directly and specifically to its site of action. It should not bind to, or interact with related receptors and or bind specifically to passing serum proteins. There should also be no risk that breakdown of this ideal compound gives rise to any toxic metabolites and the compound should have an appropriate half-life, passing gradually through the kidneys without harming them.

Leads discovered using virtual screening and *de novo* design methodologies need to be optimized to produce candidates with improved bioavailability and low toxicity [162]. Lead molecules are ligands that typically exhibit suboptimal target binding affinity. Studies have shown that there exists a difference between leads and drugs [63], which can be expressed as follows: Leads exhibit, on average, less molecular complexity (less molecular weight, less number of rings and rotatable bonds), are less hydrophobic (lower ClogP and LogD74) and

have lower polarizability (less calculated molar refractivity, CMR). Leads should display the following properties to be considered for further development in the drug discovery process or to be called as "drug-like" [63]: (1) relatively simple chemical features, amenable for combinatorial and medicinal chemistry optimization efforts; (2) membership to a well established SAR (structure-activity relationship) series, wherein compounds with similar structures exhibit similar target binding affinity; (3) favorable patent situation; and (4) good ADME properties. The ADME characteristics of a drug, together with its pharmacological properties are conventionally viewed as part of drug development - the process of making a molecule as effective as possible as a medicine [163]. Studies have indicated that poor pharmacokinetics and toxicity are the most important causes of high attrition-rates in drug development and it has been widely accepted that these areas should be considered as early as possible in the drug discovery process, thus improving the efficiency and cost-effectiveness of the industry [23,24].

Human ADMET predictions can be attempted at several levels [164]: (1) *In silico* or computational predictions from QSAR models to project *in vitro* or *in vivo* data, (2) Inter-species, *in vivo-in vivo* (including allometry) using data from pre-clinical species and (3) *In vitro-in vivo* using data obtained from tissue or recombinant material from human and pre-clinical species. *In silico* methods are already being harnessed to predict the probable ADMET profiles of any molecule, thus reducing the number of experimental studies required for compound selection and improving the success rate [9,165-167]. *In silico* prediction of drug-likeness at an early stage involves evaluation of various ADMET properties using computational approaches like QSAR or molecular modeling [165,168]. A number of studies have been conducted to identify properties that make a drug distinct from other chemicals [61,169,170]. Availability of large databases of drug or drug-like molecules, e.g. CMC (Comprehensive Medicinal Chemistry), MDDR (MACCS-II Drug Data Report), WDI (World

Drug Index) provide useful information about the properties of drugs. The most influential study of "Lipinski's rule-of-five" identifies several critical properties that should be considered for compounds with oral delivery as concern [171]. A deeper understanding of the relationships between important ADME parameters and molecular structure and properties is needed to develop better *in silico* models to predict ADMET properties [9]. Some of the ADME properties evaluated using *in silico* models are; intestinal permeability, aqueous solubility, human intestinal absorption, human oral bioavailability, active transport, efflux by P-glycoprotein, blood-brain barrier permeation, plasma protein binding, metabolic stability, interactions with cytochrome P450s and toxicity.

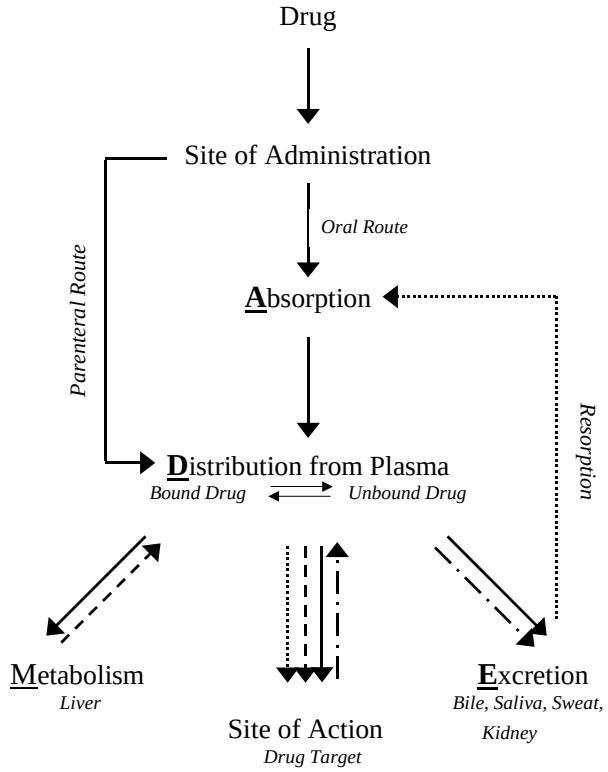

Figure 7. The distribution path of an orally administered drug molecule inside the body is depicted. Black solid arrows: Complete path of drug starting from absorption at site of administration to distribution to the various compartments in the body, like sites of metabolism, drug action and excretion. Dashed arrows: Path of the drug after metabolism. Dash-dot arrows: Path of drug after eliciting its required action on the target. Dot arrows: Path of the drug after being reabsorbed into circulation from the site of excretion.

4.1. Absorption

Drug absorption is a property of paramount importance in drug design. Oral absorption, A, also termed bioavailability, is typically measured as [172],

$$A = (D_o / D_{iv}) \times 100$$

where D_o is the drug distribution after oral administration, D_{iv} after intravenous administration.

For reasons of ease of administration and patient compliance, there is an overwhelming preference for drugs to be orally bioavailable. One of the key requirements for oral bioavailability is that a compound should be soluble in the gastric fluid and be capable of permeating the intestinal epithelium, crossing from the gut into systemic circulation. Absorption depends on the solubility and permeability of the compound, as well as interactions with transporters and metabolizing enzymes in the gut wall. The considerations at this stage therefore are, ensuring solubility (hydrophilicity) and lipophilicity (hydrophobicity) for optimal absorption. The hydrophilicity lipophilicity balance (HLB) refers to a subtle balance that the drug must possess. It is measured on an empirical scale of 0-20, where an HLB value of 0 corresponds to a completely hydrophobic molecule and a value of 20 to a molecule made up completely of hydrophilic components [173]. Another consideration which has a bearing on solubility and transformations is the pKa of the functional group(s) on the drug and their ionization state in the stomach / small intestines. Because of the difficulty in obtaining human permeability data, the Caco-2 cell monolayer or Madin-Darby canine kidney (MDCK) monolayer models are employed as references [174]. Caco-2 or MDCK cell lines are routinely used in pharmaceutical industry and form a substitute for measuring actual intestinal permeability.

Considerable efforts have gone into the development of *in silico* models for the prediction of oral absorption [175]. However, predicting oral bioavailability is not an easy task, as it depends on the superposition of two processes - absorption and liver first-pass metabolism. Simple models are based on descriptors such as log P or log D, or polar surface area, size of the molecule, shape and flexibility [176-179]. Different multivariate approaches such as, multiple linear regression analysis, partial least squares and artificial neural networks have been used to

develop quantitative structure-human-intestinal-absorption relationships [179]. In all approaches, hydrogen bonding is considered to be a property with an important effect on oral absorption. Lipinski's rule of five arrived at in a retrospective analysis of the marketed drugs has been an extremely useful empirical guide in predicting oral bioavailability. Absorption simulation programs, such as GastroPlus [180] and Idea [181] have become valuable tools in lead optimization and compound selection. They are based on advanced compartmental absorption and transit (ACAT) models, in which physicochemical concepts, such as solubility and lipophilicity are more readily incorporated. The predictive approaches to permeability/absorption prediction have largely been confined to compounds that are transported across the intestinal mucosa by predominantly passive absorption mechanisms. However, there are classes of drugs like ACE inhibitors and beta-lactam antibiotics that rely on active transport systems to convey them from gut to the bloodstream [182].

4.2. Distribution

After absorption, drug enters the blood circulation and binds to blood plasma proteins nonspecifically and is distributed to various tissues and organs in the body. The volume of distribution is defined as [172,183],

$$V_d \text{ (in litres)} = D_{body} / D_{plasma}$$

where D_{body} is the amount of drug in body (mg), D_{plasma} is plasma concentration of drug (mg/L)

The extent of the distribution depends on structural and physicochemical properties of the compound. The primary goal of the drug however, is to reach and bind to its molecular target for which it is tailor-made. If the affinity of the drug is high for the target then the drug molecule will preferentially reach the target site obeying law of mass action and as the drug leaves after eliciting its response, more drug molecules reach the site with blood plasma proteins acting as reservoir. High affinity to the target and optimal binding strength for plasma

proteins is required to ensure nonspecific binding with affinities comparable to solvent or less. The volume of distribution, together with the clearance rate, determines the half-life of a drug and therefore its dose regimen and so an early prediction of both the properties would be of considerable benefit. The log-log plot of unbound volume of distribution, V_d against distribution D at pH 7.4 (with the data corrected for plasma-protein binding), reveals a clear linear trend, with log V_d increasing at higher lipophilicities [184]. This can be used as a simple guide in modifying and optimizing the V_d. It is important to estimate the fraction of drug bound to plasma proteins, because only the unbound drug can cross the membranes and bind to the intended molecular target. In addition to plasma proteins like albumin, glycoproteins and lipoproteins, drug can bind to a variety of particles in the blood, including red blood cells, leukocytes, platelets and globulins.

In silico approaches to predict plasma protein binding have been critically reviewed by several authors [185,186]. Recently, chromatographic retention data has been used to generate a predictive QSPR comprising various E-state and molecular connectivity indices [187]. Using the multiple computer-automated structure evaluation (M-CASE) program and protein affinity data for 154 drugs, models were generated that correctly predicted the percentage of drug bound in plasma for ~ 80% of the test compounds with an average error of ~ 14% [188]. For a drug to exert a therapeutic effect at a central nervous system (CNS) target, it must be able to cross from the systemic circulation into the CNS. There are two interfaces at which this may occur: the blood-brain barrier (BBB) and the blood-cerebrospinal fluid barrier. In the case of CNS-targeted drugs, signs of good BBB permeation will be sought; conversely, for systemically targeted drugs, minimal BBB permeation will help reduce the likelihood of CNS side effects. For this reason, there has been a great interest in the computational prediction of BBB permeation as indicated by recent reviews [189,190]. The computational models developed for BBB permeation can be grouped into three classes. First, there are simple "rules of thumb" that have

been derived by examining the molecular properties of compounds that do and do not cross the BBB [191-195]. Second are classification models that typically predict whether or not a compound is a BBB permeator [196-198]. The final class comprises models predicting continuous values of BBB permeation based on either logBB or logPS data [199].

4.3. Metabolism

A major concern in drug design is the possible *in vivo* metabolic transformations and ensuring that the small molecules (hits) designed remain intact. The drug molecule through blood may also reach besides the target, the sites of biotransformations, usually liver, where the drug metabolizing enzymes (DME) present (Table 1) convert it into metabolites. Several aspects of metabolism are relevant to drug discovery, including the rate and extent of metabolism, the enzymes involved and the products formed, each of which can give rise to different concerns. The extent and rate of metabolism affect clearance, whereas the involvement of particular enzymes might lead to issues related to the polymorphic nature of some of these enzymes and to drug-drug interactions.

In silico approaches to predicting metabolism can be divided into QSAR and three-dimensional QSAR studies [200], protein and pharmacophore models [201,202] and predictive databases. Computational techniques for the prediction of possible metabolites through structure based [203] or rule based methods [204], and the compilation of xenobiotic metabolite databases [204] are a significant development in computer aided drug design.

There has been much interest, in the prediction of interactions of organic compounds with individual cytochrome P450 (CYP450) enzymes, which constitute the major drug metabolizing enzyme system in the human body. Two broad approaches have been adopted to model these interactions: those using available X-ray structures to create homology models of important CYP450s and those that are ligand based, studying known inhibitors/substrates in an

attempt to generate pharmacophore or QSAR models [205]. Availability of three-dimensional structures of all the enzymes responsible for biotransformations combined with rules of design to ensure only nonspecific binding to these enzymes except for the target, is a conceivable pathway within the framework of structure based drug design. About 51 enzymes are identified as responsible for biotransformations and of these, structures for 33 are available in the PDB facilitating a start for an affinity-based elimination of compounds likely to be transformed into inactive metabolites (Table 1). Also, a catalogue of enzymic reactions *in vivo* and substrate structures together with preferred cleavage / modification site information could suggest guidelines for drug designers in proposing candidate molecules to ensure that preempting modifications do not occur. The role of cofactors and coenzymes could pose some hurdles or failures in this scheme, which only a better appreciation of metabolomics can help alleviate.

Metabolomics is gaining increasing interest in drug discovery and disease diagnostics and treatment [206]. The concept of global analysis of all metabolites in a sample and the analysis of metabolic responses to drugs or diseases was recently introduced. Additional non-enzymatic modifications can also occur due to pH, coenzymes or other molecules *in vivo*. Conjugation is another possibility. A database of potential breakdown/modification pathways of a representative set of small molecules, based on bond strengths, quantum mechanical charge distributions and organic reaction mechanisms may facilitate this step in suggesting a few do's and don'ts in design.

Table 1: Drug metabolizing enzymes with their family and availability of 3D structures

S. No.	Drug Metabolizing Enzyme	Family	Structure in PDB
1	Human Cytidine deaminase	Hydrolases	Yes
2	Cholinesterase	Hydrolases	Yes
3	ECOLI Beta-lactamase	Hydrolases	Yes
4	Human Adenosine deaminase	Hydrolases	Yes
5	Human Pancreatic alpha-amylase precursor	Hydrolases	Yes
6	Human Arylsulfatase A precursor	Hydrolases	Yes
7	Human Liver carboxylesterase 1 precursor	Hydrolases	Yes
8	Human Glutamine synthetase	Ligases	Yes

9	Human Cytochrome P450 3A4	Oxidoreductase	Yes
10	Human Cytochrome P450 2D6	Oxidoreductase	Yes
11	Human Cytochrome P450 2C19	Oxidoreductase	No
12	Human Cytochrome P450 2B6	Oxidoreductase	No
13	Human Amine oxidase	Oxidoreductase	Yes
14	Human Cytochrome P450 2C9	Oxidoreductase	Yes
15	Cytochrome P450 19	Oxidoreductase	No
16	Aldehyde oxidase and P450	Oxidoreductase	Yes
17	Human Aldehyde oxidase	Oxidoreductase	Yes
18	Human Cytochrome P450 1A2	Oxidoreductase	No
19	Cytochrome P450 3A4	Oxidoreductase	Yes
20	Human Cytochrome P450 2C19	Oxidoreductase	No
21	Human Cytochrome P450 2C8	Oxidoreductase	Yes
22	Human Cytochrome P450 CP2D6	Oxidoreductase	No
23	Human Cytochrome P450 CYP2D6	Oxidoreductase	No
24	Human Cytochrome P450 2A6	Oxidoreductase	Yes
25	Human Cytochrome P450 2E1	Oxidoreductase	No
26	Human Cytochrome P450 2A13	Oxidoreductase	No
27	Human Alcohol dehydrogenase 6	Oxidoreductase	Yes
28	Human Cytochrome P450 11A1	Oxidoreductase	No
29	Human Cytochrome P450 24A1	Oxidoreductase	No
30	Human Cytochrome P450 1A1	Oxidoreductase	No
31	Human Cytochrome P450, subfamily IIIA	Oxidoreductase	No
32	Human Xanthine dehydrogenase/oxidase	Oxidoreductase	Yes
33	Human Cytochrome P450 3A4	Oxidoreductase	Yes
34	Human Cytochrome P450 1A2	Oxidoreductase	No
35	Human Cytochrome P450 11B2	Oxidoreductase	No
36	RAT Cytochrome P450 3A1	Oxidoreductase	No
37	RAT Cytochrome P450 2C11	Oxidoreductase	No
38	Human Carbonyl reductase	Oxidoreductase	Yes
39	Human Proline oxidase	Oxidoreductase	Yes
40	Human Tryptophan 2,3-dioxygenase	Oxidoreductase	Yes
41	Aminoglycoside 2'-N-acetyltransferase	Transferases	Yes
42	Kanamycin nucleotidyltransferase	Transferases	Yes
43	Aminoglycoside 3'-phosphotransferase	Transferases	Yes
44	Human Glutathione S-transferase A1	Transferases	Yes
45	Human Glutathione S-transferase A2	Transferases	Yes
46	COMT (catecol-O-methyl-transferase)	Transferases	No
47	Human Nucleoside diphosphate kinase A	Transferases	Yes
48	Human Thymidine phosphorylase	Transferases	Yes
49	Human Deoxycytidine kinase	Transferases	Yes
50	Human Histamine N-methyltransferase	Transferases	Yes
51	UDP Glucosyltransferases	Transferases	No

4.4. Excretion

Clearance / excretion is an important parameter that defines, together with the volume of distribution, the half-life and thus the frequency of dosing of a drug. Clearance, Cl, is related to distribution and elimination in the following manner [183],

$$Cl\ (L/hr) = R_e / D_{plasma}$$

where R_e is the rate of elimination (mg/hr), Dplasma is drug concentration in blood plasma (mg/L)

R_e is given by, $R_e = k_e \times D_{body}$

where k_e is the elimination rate constant and D_{body} is the amount of drug in body (mg).

Thus, $Cl = (0.693 \times V_d) / t_{1/2}$

where V_d is the volume of distribution defined in section 4.2 and $t_{1/2}$ is drug half life defined in section 4.5.4.

Excretion of the drug from the body mainly takes place via the liver (hepatic clearance or metabolism and biliary excretion) and the kidney (renal excretion). Except highly polar substances, most drugs are lipid soluble and are reabsorbed from the kidney back into the bloodstream. These compounds undergo metabolism, generating more polar species that may avoid renal absorption and be excreted in the urine [207]. The design must incorporate enough solubility of the drug and its metabolites to facilitate this process. In a plot of plasma concentration against time, the area under the curve relates to dose, bioavailability and clearance [9]. Renal clearance in humans may be predictable from rat renal clearance that has been corrected for species differences in glomerular filtration rate [208]. Allometric relationships for clearance tend to be most successful for compounds undergoing renal clearance or high hepatic extraction where clearance approaches liver blood flow [209]. A multiple linear regression method combining clearance data from two species and readily calculated structural parameters (MW, clogP and number of hydrogen bond acceptors) predicts

human clearance much better (q^2 = 0.682, RMSE = 0.35) [164]. Excretion related properties have not received much attention in drug design so far [185]. Software for the prediction of possible metabolites of the candidate molecule and a strategy to ensure HLB of the candidate and higher hydrophilicity of the metabolites should help.

4.5. Toxicity

Enumerating molecular origins of toxicity is a difficult task but one could envisage the following factors as contributory and propose a computational route to overcome them (Fig. 8).

4.5.1. Tight binding to non-targets. A repository of the three dimensional structures of all biomolecules inside the target cell can help establish specificity to target vis-à-vis non-targets and this, that is, selective binding to target is a necessity.

4.5.2. Accumulation at wrong sites. This could be due to nonspecific binding. Proper HLB will ensure reentry into blood. Both (4.5.1 and 4.5.2) above also apply to metabolites of the drug.

4.5.3. Tight or irreversible binding to target with multiple functions. Firstly, advances in metabolomics should help in identifying a target that does not interfere with different functions. Metabolic pathways help in understanding the point of interception by the drug and its consequences. The ideal target must have a single function that the drug is attempting to interfere with. Irreversible binding to targets exclusive to pathogens is acceptable so also to targets on viral DNA/RNA. Exclusive nucleic acid based targets in humans for cancer cells are probably difficult to establish without interference with normal cells. Where targets have multiple functions, half-life of the drug needs to be fine-tuned. In a nutshell, the computational pathways need to address proper affinity, specificity besides HLB and high solubility of the metabolites for minimizing toxicity.

The existing commercially available *in silico* tools for predicting potential toxicity issues can be roughly classified into two groups. The first group uses expert systems that derive

models on the basis of abstracting and codifying knowledge from human experts and scientific literature. The second group relies primarily on the generation of descriptors of chemical structures and statistical analyses of the relationships between these descriptors and the toxicological end-points [9]. A recent review discusses the advances in toxicology software [210].

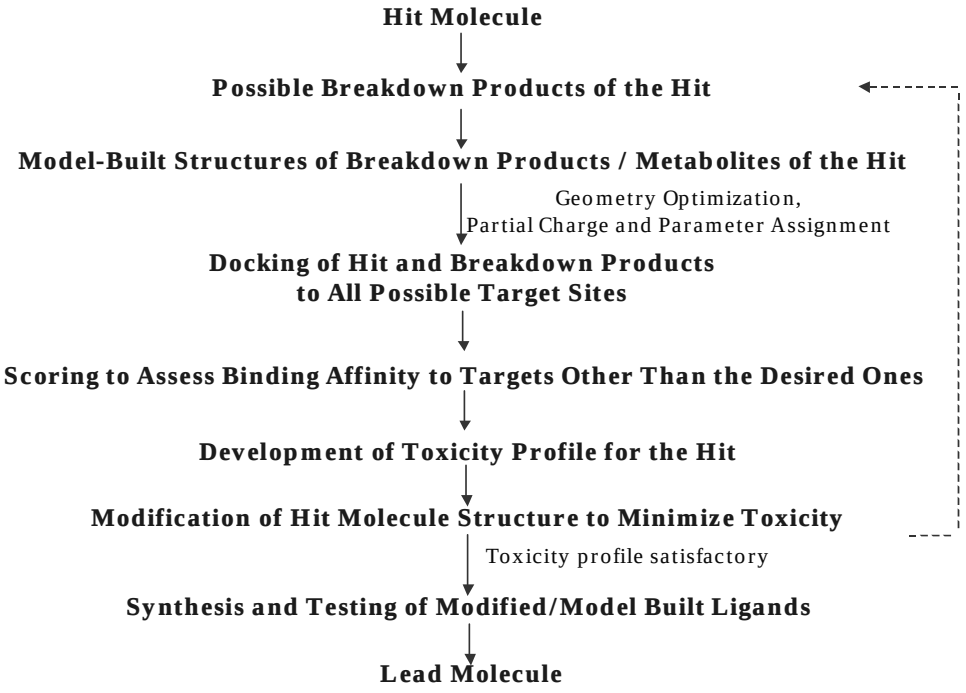

Figure 8. A methodology to assess the possible toxicity of a lead-like molecule taking off from hits in Figure 2 to arrive at a lead molecule.

4.5.4 Drug retention/residence times. Drug activity and ADME characteristics are related to the residence/retention time of the drug [211] i.e. the time period in which the drug remains bound at the target site, and hence is a crucial factor to be considered during drug design. Non-covalent target-drug complex dissociation typically occurs via a unimolecular dissociation process characterized by the rate equation (first order) [183],

$$[C]= [C]_o \exp(-k_d t)$$

where [C] is the concentration of the drug in complexed form at time, t; $[C]_o$ is the concentration of the drug at t=0, k_d is the dissociation rate constant.

For such a process, the retention time, t_R is obtained from the dissociation rate constant as,

$$t_R = 1/k_d$$

and the half life, $t_{1/2}$ of the drug can be calculated [183] as

$$t_{1/2} = 0.693/k_d$$

The retention time or half-life are important factors which determine the elimination of the drug and hence of significant consideration in toxicity studies. A good binder may not necessarily be a good drug if its retention time is too high, which could cause toxic effects. Also, targets having multiple functions should only be blocked for optimal times or else metabolic pathways other than the targeted pathway may get adversely affected. On the other hand, long retention time could be potentially advantageous in terms of duration of pharmacological effect and target selectivity [211]. Longer half-lives also result in improved drug activity as has been demonstrated in the case of inhibition of viral replication [211].

Drug activity and toxicity can be modulated by controlling its retention time, which depends on both, the structure and charge of the drug as well as external factors like pH [212] and concentration of other solutes [213]. The drug retention time is determined by mainly two dynamic factors, the amount of drug distributed and its elimination processes. Thus, longer retention time can be achieved by either increasing the volume of distribution or decreasing the elimination. The latter is typically easier and may be achieved by means of chemical modifications. For increasing the volume of distribution, sustained-release dosage forms and coadministration of inhibitors of drug-metabolizing enzymes can be employed [173].

Computational methods for the prediction of retention times or dissociation rate constants can be extremely useful in the design of drugs with optimal retention times. Simulation based methods for the prediction of dissociation rate constants [214] may be employed but are highly compute-intensive. QSAR based approaches designed for the study of interaction kinetics may also be adopted for this purpose [215]. An alternative to these can be

the development of an empirical relation based on experimental data for swift prediction of dissociation rate constants.

Keeping such an empirical approach in mind, we carried out a preliminary analysis of experimental data on equilibrium dissociation constants and half-lives derived from experimental dissociation rate constants, and observed a high correlation between the two (Fig. 9). The data set includes DNA [216] as well as protein targets consisting of different enzymes [217-220], receptors [221] and other proteins [220, 222].

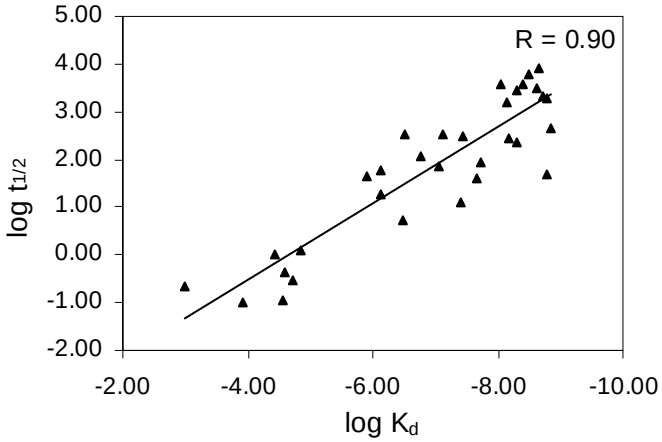

Figure 9. Correlation between drug retention half-life ($t_{1/2}$) and the dissociation constant (K_D) shown as a log-log plot.

From the slope of the linear correlation plot above it may be inferred that, a nanomolar dissociation constant corresponds to a half-life of above one hour. This however, is only an upper limit based on *in vitro* studies.

Intuitively, it is expected that strong binders should result in complexes with longer half-lives. Rates are however related to free energies of activation and not free energies of binding, thus only an empirical correlation can be hoped for at this stage. Also, the effects of competitors and solvent have to be factored into such an analysis.

Thus using an extension of the above approach or similar computational techniques, binding affinity may be fine-tuned to address the retention time issue at the design stage.

5. Some softwares for drug design or intermediate steps thereof

A few comprehensive drug design software are listed in Table 2, some of which are in public domain.

Table 2: Some Drug Design Software

SNo	Software Name	Company/ Institution	Provided Utilities and URL
1	InsightII, Discovery studio Cerius ADME/ Tox Package	Accelrys	Molecular modeling and *de novo* drug design http://www.accelrys.com/products/insight/ Computational models for the prediction of ADME properties derived from chemical structures. http://www.accelrys.com/products/cerius2/cerius2products/c2adme.html
2	Sybyl	Tripos	Computational informatics software for drug discovery http://www.tripos.com/
3	Phase, Glide, Liaison, Qikprop Maestro etc	Schrodinger	Pharmacophore modeling, Ligand –receptor docking, Ligand-receptor binding free energy prediction, ADME prediction, Molecular modeling etc. http://www.schrodinger.com/
4	Bio-Suite	Tata Consultancy Services Ltd	Genomics, protein modeling and structural analysis, simulation and drug Design. http://www.atc.tcs.co.in/biosuite/
5	Sanjeevini	Indian Institute of Technology, New Delhi	Active site directed drug design http://www.scfbio-itd.res.in/research/drugdesign.htm

6. Conclusion and Perspectives

Given the very high attrition rates in drug discovery besides the cost and time factors, the role of computer aided drug design cannot be overemphasized. The key driving forces for current day *in silico* drug design endeavors are the availability of structural information of the targets, emergence of reliable energy functions and force field compatible solvation treatments, as well as free energy methodologies and accessibility of high-end computing clusters. A combination of basic concepts in chemical bonding (generation of candidate molecules from templates), quantum mechanics (geometry optimization and charge derivation), classical mechanics (molecular mechanics and dynamics), statistical mechanics (configurational /

Boltzmann averaging) and thermodynamics (standard free energies of complex formation) allows the development of a rigorous protocol for *in silico* drug design. The overview presented here discusses the advances in and the applicability of predictive *in silico* methods to drug design, from candidate molecule generation, evaluation of their target affinity and specificity, identification of hits, to predicting their fate in the body through ADME and toxicity studies. This review describes the drug design process from a physicochemical perspective as comprising three stages (Fig1). The first stage mainly concerns hit identification on the basis of candidate generation and target affinity, molecular docking, scoring and binding affinity predictions. The next stage involves identification of the target specificity of the candidate molecules, for which a computational protocol is proposed (Fig. 6). This protocol can be easily extended to all known targets with a series of candidate molecules or known drugs. The final stage deals with drug absorption, distribution, metabolism, excretion and toxicity profiles. The significance of these studies to drug design and *in silico* efforts to develop predictive ADMET techniques are discussed. Computational prediction of drug retention times or half-lives, which are strongly related to and also dictate ADMET profiles, is emphasized and a method proposed. If all the steps enumerated in stages I to III (Fig. 1) above could be implemented *in silico*, a drug molecule with desired affinity, high specificity and low toxicity can be discovered. The computational protocols (Fig 2) out-lined can be fine-tuned at each stage to improve accuracies. The major lacunae are in the structural database of biomolecules in target cells, a catalogue of cell specific enzymic reactions *in vivo* and software/methodology to screen the new molecules or their breakdown products for preventing specific binding to wrong sites. Progresses in structural genomics / proteomics and metabolomics are expected to facilitate addressing some of these issues at a molecular level in the near future. Worldwide efforts on genomics and proteomics have given a significant boost to both experimental and computational methods to

march towards personalized medicine with minimal side effects. Automated lead design *in silico* seems a realizable dream in the near future.

Acknowledgements. The authors wish to thank the Department of Science and Technology and the Department of Biotechnology, Govt. of India for funding and Dabur Research Foundation for a collaborative research project on the computational design of inhibitors for nuclear receptors. Thanks are also due to Ms. Gandhimathi, Ms. Nidhi Nigam, Ms. Jayamala Pandey, Mr. Shailesh Tripathi, Mr. Anuj Gupta and Mr. Pankaj Sharma for help in preparing the manuscript. Dr. N. Latha is a Reader in Sri Venkateswara College, Delhi University. Ms. Saher Afshan Shaikh is a recipient of Senior Research Fellowship from CSIR, Govt. of India.

7. References

1. Kennedy T. Managing The Drug Discovery/Development Interface. Drug Discov Today 1997; 2: 436-444.

2. Hileman B. Accounting for R&D. Many doubt the $800 million pharmaceutical price tag. Chem. Eng. News 2006; 84: 50-51.

3. Debouck C, Metcalf B. The Impact of Genomics on Drug Discovery. Annu Rev Pharmacol Toxicol 2000; 40: 193-208.

4. Burbaum J, Tobal GM. Proteomics in Drug Discovery. Curr Opin Chem Biol 2002; 6: 427-433.

5. Gatto, JG. The Changing Face of Bioinformatics. Drug Discov Today 2003; 8:375-376.

6. Corbett PT, Leclaire J, Vial L, West KR, Wietor JL, Sanders JKM, Otto S. Dynamic Combinatorial Chemistry. Chem Rev 2006; 106: 3652 - 3711.

7. Bajorath J. Integration of Virtual and High-Throughput Screening. Nat Rev Drug Discov 2002; 1: 882-894.

8. van de Waterbeemd H, Gifford E. ADMET *In Silico* Modeling: Towards Prediction Paradise? Nat Rev Drug Discov. 2003; 2: 192-204.

9. Schneider G, Fechner U. Computer-Based *De Novo* Design of Drug-Like Molecules. Nat Rev Drug Discov 2005; 4: 649-663.

10. Marrone TJ, Briggs JM, McCammon JA. Structure-Based Drug Design: Computational Advances. Annu Rev Pharmacol Toxicol 1997; 37: 71-90.

11. Kuntz ID, Meng EC, Shoichet BK. Receptor-Based Molecular Design Acc Chem. Res 1994; 27: 117-123.

12. Klebe G. Recent Developments in Structure-Based Drug Design. J Mol Med 2000; 78: 269-281.

13. Whittle PJ, Blundell TL. Protein Structure-Based Drug Design. Annu Rev Biophys Biomol Struct 1994; 23: 349-375.

14. Latha N, Jain T, Sharma P, Jayaram, B. A Free Energy Based Computational Pathway from Chemical Templates to Lead Compounds: A Case Study of COX-2 Inhibitors. J Biomol Struct Dyn 2004; 21: 791-804.

15. Shoichet, BK. Virtual Screening of Chemical Libraries. Nature 2004; 432: 862-865.

16. Schneider G, Bohm HJ Virtual Screening and Fast Automated Docking Methods, Drug Discov Today 2002; 7: 64-70.

17. Augen J. Bioinformatics and Information Technology: Reshaping the Drug Discovery Process. Drug Discov Today 2002; 7: S39-S40.

18. Kahn S. Bioinformatics: A Holistic Approach to Drug Discovery. Drug Discov Today 2002; 7:633-634.

19. Claus BL, Underwood DJ. Discovery Informatics: Its Evolving Role in Drug Discovery. Drug Discov Today 2002; 7: 957-966.

20. Jorgensen WL. The Many Roles of Computation in Drug Discovery. Science 2004; 303: 1813-1818.

21. Good A. Structure-Based Virtual Screening Protocols. Curr Opin Drug Discov Devel 2001; 4: 301-307.

22. Walters WP, Murcko MA Prediction of 'Drug-Likeness', Adv Drug Deliv 2002; 54: 255-271.

23. Beresford A, Selick H, Tarbit M. The Emerging Importance of Predictive ADME Simulation in Drug Discovery. Drug Discov Today 2002; 7: 109-116.

24. Hann MM, Oprea TI. Pursuing the Leadlikeness Concept in Pharmaceutical Research. Curr Opin Chem Biol 2004; 8: 255-263.

25. Ajay W, Murcko MA. Computational Methods to Predict Binding Free Energies in Ligand-Receptor Complexes. J Med Chem 1995; 38: 4953-4967.

26. Kubinyi H. QSAR and 3D-QSAR in Drug Design. Part 1. Methodology. Drug Discov Today 1997; 2: 457-467.

27. Kubinyi H. QSAR and 3D QSAR in Drug Design Part 2: Applications and Problems Drug Discov Today 1997; 2: 538-546.

28. Latha N, Jayaram B. A Binding Affinity Based Computational Pathway for Active-Site Directed Lead Molecule Design: Some Promises and Perspectives. Drug Des Rev Online 2005; 2: 145-165.

29. Bajorath J. Rational Drug Discovery Revisited: Interfacing Experimental Programs with Bio- and Chemo-Informatics. Drug Discov Today 2001; 6: 989-995.

30. Anderson AC. The Process of Structure-Based Drug Design. Chem Biol 2003; 10: 787-797.

31. Antel J. Integration of Combinatorial Chemistry and Structure-Based Design. Curr Opin Drug Discov Dev 1999; 2: 224-233.

32. Verlinde CL, Hol WG. Structure-Based Drug Design: Progress, Results and Challenges. Structure 1994; 2: 577-587.

33. Knowles J, Gromo G. A Guide to Drug Discovery: Target Selection in Drug Discovery. Nat Rev Drug Discov 2003; 2: 63-69.

34. Hopkins AL, Groom CR. The Druggable Genome. Nat Rev Drug Discov 2002; 1: 727-730.

35. Lindsay MA. Finding New Drug Targets in the 21st Century. Drug Discov Today 2005; 10: 1683-1687.

36. Mestres J. Representativity of Target Families in the Protein Data Bank: Impact for Family-Directed Structure-Based Drug Discovery. Drug Discov Today 2005; 10: 1629-1637.

37. Zheng CJ, Han LY, Yap CW, Xie B, Chen YZ. Progress and Problems in the Exploration of Therapeutic Targets. Drug Discov Today 2006; 11: 412-420.

38. Sams-Dodd F. Target-Based Drug Discovery: is something wrong? Drug Discov Today 2005; 10: 139-147.

39. Ghosh I. Target Based High Throughput Screening and Lead Designing in Pharmaceutical Drug Industry. Indian J Chem 2006, 45-A, 163-173.

40. Imming P, Sinning C, Meyer A. Drugs, Their Targets and the Nature and Number of Drug Targets. Nature Rev Drug Discov 2006; 5: 821-834.

41. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, Chang Z, Woolsey J. Drugbank: A Comprehensive Resource for *In Silico* Drug Discovery and Exploration. Nucleic Acids Res 2006; 1: 34.

42. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000; 28: 235-242.

43. Drews J. Drug Discovery: A Historical Perspective. Science 2000; 287: 1960-1964.

44. Scapin, G. Structural Biology and Drug Discovery. Curr Pharm Des. 2006; 12: 2087-2097.

45. Vajda S, Guarnieri F. Characterization of Protein-Ligand Interaction Sites using Experimental and Computational Methods. Curr Opin Drug Discov Devel 2006; 9: 354-362.

46. Blundell TL, Sibanda BL, Montalvao RW, Brewerton S, Chelliah V, Worth CL, Harmer NJ, Davies O, Burke D. Structural Biology and Bioinformatics in Drug Design: Opportunities and Challenges for Target Identification and Lead Discovery. Philos Trans R Soc Lond B Biol Sci. 2006; 361: 413-423.

47. Jiang Z, Zhou Y. Using Bioinformatics for Drug Target Identification from the Genome. Am J Pharmacogenomics. 2005; 5:387-396.

48. 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: Struct Func Bioinfo 2004; 54: 671-680.

49. Narang P, Bhushan K, Bose S, Jayaram B. A Computational Pathway or Bracketing Native-Like Structures for Small Alpha Helical Globular Proteins. Phys Chem Chem Phys 2005; 7: 2364-2375.

50. Clark DE, Frenkel D, Levy SA, Li J, Murray CW, Robson B, Waszkowycz B, Westhead DR. PRO-LIGAND: An Approach to *De Novo* Molecular Design. 1. Application to the Design of Organic Molecules. J Comput Aided Mol Des 1995; 9: 13-32.

51. Laskowski RA, Thornton JM, Humblet C. RASSE: A New Method for Structure-Based Drug Design. J Chem Inf Comput Sci 1996; 36:1187-1194.

52. Wang R, Gao Y, Lai L. LigBuilder: A Multiple-Purpose Program for Structure-Based Drug Design. J Mol Model 2000; 6: 498-516.

53. Beavers MP, Chen X. Structure-Based Combinatorial Library Design: Methodologies and Applications. J Molec Graph Mod 2002; 20: 463-468.

54. Merlot C, Domine D, Cleva C, Church DJ. Chemical Substructures in Drug Discovery. Drug Discov Today 2003; 8: 594-602.

55. Hou T, Xu X, Recent Development and Application of Virtual Screening in Drug Discovery: An Overview. Curr Pharm Des 2004; 10: 1011-1033.

56. Anderson AC, Wright DL. The Design and Docking of Virtual Compound Libraries to Structures of Drug Targets. Curr Comp Aided Drug Des 2005; 1: 103-127.

57. Irwin JJ, Shoichet BK. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. J Chem Inf Model 2005; 45: 177-182.

58. Michalsky E, Dunkel M, Goede A, Preissner R. Superligands - A Database of Ligand Structures Derived from the Protein Data Bank. BMC Bioinformatics 2005; 6:122.

59. Goede A, Dunkel M, Mester N, Frommel C, Preissner R. SuperDrug: A Conformational Drug Database. Bioinformatics 2005; 21:1751-1753.

60. Lipinski C A. Lead-and Drug-Like Compounds: The Rule-Of-Five Revolution. Drug Discov Today: Tech. 2004; 1: 337-341.

61. Lipinski C A. Drug-Like Properties and the Causes of Poor Solubility and Poor Permeability. J Pharmacol Toxicol Methods 2000; 44: 235-249.

62. Oprea TI. Property Distribution of Drug-Related Chemical Databases. J Comput Aided Mol Des 2000; 14: 251-264.

63. Oprea TI, Davis AM, Teague SJ, Leeson PD. Is there a Difference between Leads and Drugs? A Historical Perspective. J Chem Inf Comput Sci 2001; 41: 1308-1315.

64. Ghose AK, Viswanadhan VN, Wendoloski JJ. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. J Comb Chem 1999; 1: 55-68.

65. Oprea TI. Strategies for Compound Selection. Curr Drug Discov Tech 2004; 1: 211-220.

66. Karplus M, Petsko GA. Molecular Dynamics Simulations in Biology. Nature 1990; 347: 631-639.

67. Jorgensen WL, Tirado-Rives J. Chemical Theory and Computation Special Feature: Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. Proc Natl Acad Sci 2005; 102: 6665-6670.

68. MacKerell AD Jr. Empirical Force Fields for Biological Macromolecules: Overview and Issues. J Comput Chem 2004; 25: 1584-1604.

69. Ponder JW, Case DA. Force Fields for Protein Simulations. Adv Protein Chem 2003; 66: 27-85.

70. Weiner PK, Kollman PA. AMBER: Assisted Model Building with Energy Refinement. A General Program for Modeling Molecules and their Interactions. J Comp Chem 1981; 2: 287-303.

71. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. J Am Chem Soc 1995; 117: 5179-5197.

72. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comp Chem 1983; 4: 187-217.

73. Scott WRP, Huenenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krueger P, van Gunsteren WF. The GROMOS Biomolecular Simulations Program Package. J Phys Chem A 1999; 103: 3596-3607.

74. Jorgensen WL, Tirado-Rives J. The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. J Am Chem Soc 1988; 110: 1657-1666.

75. Wong CF, McCammon JA. Protein Simulation and Drug Design. Adv Protein Chem 2003; 66: 87-121.

76. Halgren TA. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization and Performance of MMFF94. J Comp Chem 1996; 17: 490-519.

77. Ewig CS, Berry R, Dinur U, Hill JR, Hwang MJ, Li H, Liang C, Maple J, Peng Z, Stckfisch TP, Thacher TS, Yan L, Ni X, Hagler AT. Derivation of Class II Force Fields. 8. Derivation of a General Quantum Mechanical Force Field for Organic Compounds. J Comput Chem 2001; 22: 1782.

78. Momany FA, Rone R. Validation of the General Purpose QUANTA 3.2/CHARMm Force Field. J Comp Chem 1992; 13: 888-900.

79. Hwang MJ, Stockfisch TP, Hagler AT. Derivation of Class II Force Fields. 2. Derivation and Characterization of a Class II Force Field, CFF93, for the Alkyl Functional Group and Alkane Molecules. J Am Chem Soc 1994; 116: 2515-2525.

80. Sun H. COMPASS: An *Ab Initio* Force-Field Optimized for Condensed-Phase Applications-Overview with Details on Alkane and Benzene Compounds. J Phys Chem B 1998; 102: 7338-7364.

81. Burkert U, Allinger NL. Molecular Mechanics. J Am Chem Soc 1982; 177: 23.

82. Rappe AK, Casewit CJ, Colwell KS, Goddard WA, Skiff WM. UFF, A Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. J Am Chem Soc 1992; 114: 10024-10035.

83. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and Testing of a General Amber Force Field. J Comput Chem 2004; 25: 1157-1174.

84. Babu CS, Lim Empirical Force Field For Biologically Active Divalent Metal Cations in Water. C. J Phys Chem. B 2006; 110: 691-699.

85. Shoichet BK, McGovern SL, Wei B, Irwin JJ. Lead Discovery Using Molecular Docking. Curr Opin Chem Biol. 2002; 6: 439-446.

86. Taylor RD, Jewsbury PJ, Essex JW. A Review of Protein-Small Molecule Docking Methods. J Comput Aided Mol Des 2002; 16:151-166.

87. Watters WP, Stahl MT, Murcko MA. Virtual Screening - An Overview. Drug Discov Today 1998; 3: 160-178.

88. Kairys V, Fernandes MX, Gilson MK. Screening Drug-Like Compounds by Docking to Homology Models: A Systematic Study. J Chem Inf Mod 2006; 46: 365-379.

89. Brooijmans N, Kuntz ID. Molecular Recognition and Docking Algorithms. Annu Rev Biophys Biomol Struct 2003; 32: 335-373.

90. Rosenfeld R, Vajda S, DeLisi C. Flexible Docking and Design. Annu Rev Biophys Biomol Struct. 1995; 24: 677-700.

91. Teague SJ. Implications of Protein Flexibility for Drug Discovery. Nature Rev Drug Discov 2003; 2: 527-541.

92. Davis AM, Teague SJ. Hydrogen Bonding, Hydrophobic Interactions and Failure of the Rigid Receptor Hypothesis. Angew Chem Int Ed 1999; 38: 736-749.

93. Murray CW, Baxter CA, Frenkel AD. The Sensitivity of the Results of Molecular Docking to Induced Fit Effects: Application to Thrombin, Thermolysin and Neuraminidase. J Comput Aided Mol Des 1999; 13: 547-562.

94. McCammon JA. Target Flexibility in Molecular Recognition. Biochim Biophys Acta. 2005; 1754: 221-224.

95. Trosset JY, Scheraga HA. PRODOCK: Software Package for Protein Modeling and Docking. J Comput Chem 1999; 20: 412-427.

96. Abagyan RA, Totrov MM, Kuznetsov DN. ICM - a New Method for Protein Modelling and Design. Applications to Docking and Structure Prediction from the Distorted Native Conformation. J Comput Chem 1994; 15: 488-506.

97. Liu M, Wang S. MCDOCK: a Monte Carlo Simulation Approach to the Molecular Docking Problem. J Comput Aided Mol Des 1999; 13: 435-451.

98. Hart TN, Read RJ. A Multiple-Start Monte Carlo Docking Method. Proteins 1992; 13: 206-222.

99. McMartin C, Bohacek R. QXP: Powerful, Rapid Computer Algorithms for Structure-Based Drug Design. J Comput Aided Mol Des 1997; 11: 333-344.

100. Goodsell DS, Olson AJ. Automated Docking of Substrates to Proteins by Simulated Annealing. Proteins 1990; 8: 195-202.

101. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated Docking using a Lamarckian Genetic Algorithm and an Empirical Bindng Free Energy Function. J Comput Chem 1998; 19: 1639-1662.

102. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and Validation of a Genetic Algorithm for Flexible Docking. J Mol Biol 1997; 267: 727-748.

103. Clark KP, Ajay. Flexible Ligand Docking without Parameter Adjustment across Four Ligand-Receptor Complexes. J Comp Chem 1995; 16: 1210-1226.

104. Ewing TJA, Kuntz ID. Critical Evaluation of Search Algorithms used in Automated Molecular Docking. J Comput Chem 1997; 18: 1175-1189.

105. Rarey M, Kramer B, Lengauer T, Klebe G. A Fast Flexible Docking Method using an Incremental Construction Algorithm. J Mol Biol 1996; 261: 470-489.

106. Bohm HJ. The Computer Program LUDI: A New Method for the *de novo* Design of Enzyme Inhibitors. J Comput Aided Mol Des 1992; 6: 61-78.

107. Schnecke V, Kuhn LA. Virtual Screening with Solvation and Ligand-Induced Complementarity. Perspec Drug Discov Des 2000; 20: 171-190.

108. Gabb HA, Jackson RM, Sternberg MJE. Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information. J Mol Biol 1997; 272: 106-120.

109. Wong CF, McCammon JA. Protein Flexibility and Computer-Aided Drug Design. Annu Rev Pharmacol Toxicol 2003; 43: 31-45.

110. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of Docking: an Overview of Search Algorithms and a Guide to Scoring Functions. Proteins: Struct Funct Genet 2002; 47: 409-443.

111. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. Nat Rev Drug Discov 2004; 3: 935-949.

112. 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: 2644-2676.

113. Raha K, Merz Jr. KM. Calculating Binding Free Energy in Protein Ligand Interaction. Annu Reports Comput Chem 2005; 1: 113-130.

114. Gohlke H, Klebe G. Statistical Potentials and Scoring Functions Applied to Protein-Ligand Binding. Curr Opin Struct Biol 2001; 11: 231-235.

115. Tanja Schulz-Gasch T, Stahl M. Scoring Functions for Protein-Ligand Interactions: a Critical Perspective. Drug Discov Today: Tech 2004; 1: 231-239.

116. 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: 411-428.

117. Pang YP, Perola E, Xu K, Prendergast FG. 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.

118. Zou X, Sun Y, Kuntz ID. Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model. J Am Chem Soc 1999; 121: 8033-8043.

119. Bohm HJ, Stahl M. Rapid Empirical Scoring Function in Virtual Screening Applications. Med Chem Res 1999; 9: 445-462.

120. Bohm HJ. The Development of a Simple Empirical Scoring Function to Estimate the Binding Constant for a Protein-Ligand Complex of Known Three-Dimensional Structure. J Comput Aided Mol Des 1994; 8: 243-256.

121. Bohm HJ. Prediction of the Binding Constants of Protein Ligands: A Fast Method for the Prioritization of Hits Obtained from *De Novo* Design of 3D Database Search Programs. J Comput Aided Mol Des 1998; 12: 309-323.

122. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. J Comput Aided Mol Des 1997; 11: 425-445.

123. Jain T, Jayaram B. An All Atom Energy Based Computational Protocol for Predicting Binding Affinities of Protein-Ligand Complexes. FEBS Lett 2005; 579: 6659-6666.

124. Murray CW, Auton TR, Eldridge MD. Empirical Scoring Functions. II. The Testing of an Empirical Scoring Function for the Prediction of Ligand-Receptor Binding Affinities and the

Use of Bayesian Regression to Improve the Quality of the Model. J Comput Aided Mol Design 1998; 12: 503-519.

125. Wang R, Liu L, Lai L, Tang Y. SCORE: A New Empirical Method for Estimating the Binding Affinity of a Protein-Ligand Complex. J Mol Model 1998; 4: 379-394.

126. Head RD, Smythe ML, Oprea TI, Waller CL, Green SM, Marshall GR. Validate: A New Method for the Receptor-Based Prediction of Binding Affinities of Novel Ligands. J Am Chem Soc 1996; 118: 3959-3969.

127. Krammer A, Kirchhoff PD, Jiang X, Venkatachalam CM, Waldman M. LigScore: A Novel Scoring Function for Predicting Binding Affinities. J Mol Graph Model 2005; 23: 395-407.

128. Gohlke H, Hendlich M, Klebe G. Knowledge Based Scoring Function to Predict Protein-Ligand Interactions. J Mol Biol 2000; 295: 337-356.

129. DeWitte RS, Shakhnovich EI. SMoG: *De Novo* Design Method Based on Simple, Fast, and Accurate Free Energy Estimates. I. Methodology and Supporting Evidence. J Am Chem Soc 1996; 118: 11733-11744.

130. Verkhivker G, Appelt K, Freer ST, Villafranca JE. Empirical Free Energy Calculations of Ligand-Protein Crystallographic Complexes. I. Knowledge Based Ligand-Protein Interaction Potentials Applied to the Prediction of Human Immunodeficiency Virus 1 Protease Binding Affinity. Protein Engg 1995; 8: 677-691.

131. Mitchell JBO, Laskowski RA, Alex A, Thornton JM. BLEEP - Potential of Mean Force Describing Protein-Ligand Interactions: II. Calculations of Binding Energies and Comparison With Experimental Data. J Comp Chem 1999; 202: 1177-1185.

132. 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.

133. Wang R, Lu Y, Wang S. Comparative Evaluation of 11 Scoring Functions for Molecular Docking. J Med Chem 2003; 46: 2287-2303.

134. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL. Assessing Scoring Functions for Protein-Ligand Interactions. J Med Chem 2004; 47: 3032-3047.

135. Bissantz C, Folkers G, Rognan D. Protein-Based Virtual Screening of Chemical Databases. I. Evaluation of Different Docking/Scoring Combinations. J Med Chem 2000; 43: 4759-4767.

136. Stahl M, Rarey M. Detailed Analysis of Scoring Functions for Virtual Screening. J Med Chem 2001; 44: 1035-1042.

137. Warren GL, Andrews CW, Capelli A, Clarke B, LaLonde J, Millard H, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, James M, Woolven JM, Peishoff CE, Head MS. A Critical Assessment of Docking Programs and Scoring Functions. J Med Chem 2006; 49: 5912 - 5931.

138. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. J Am Chem Soc 1990; 112: 6127-6129.

139. Qiu D, Shenkin PS, Hollinger FP, Still WC. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. J Phys Chem 1997; 101: 3005-3014.

140. Simonson T, Archontis G, Karplus M. Free Energy Simulations Come of Age: Protein-Ligand Recognition. Acc Chem Res 2002; 35: 430-437.

141. Kollman P, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham III TE. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Accts Chem Res 2000; 33: 889-897.

142. Srinivasan J, Cheatham III TE, Cieplak P, Kollman PA, Case DA. Continuum Solvent Studies of the Stability of DNA, RNA and Phosphoramidate-DNA Helices. J Am Chem Soc 1998; 120: 9401-9409.

143. Aqvist J, Medina C, Samuelsson J-E. A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. Protein Eng 1994; 7: 385-391.

144. Shaikh SA, Ahmed SR, Jayaram B. A Molecular Thermodynamic View of DNA-Drug Interaction: A Case Study of 25 Minor Groove Binders. Arch Biochem Biophys 2004; 429: 81-99.

145. Kalra P, Reddy TV, Jayaram B. Free Energy Component Analysis for Drug Design: A Case Study of HIV-1 Protease-Inhibitor Binding. J Med Chem 2001; 44: 4325-4338.

146. Jayaram B, McConnell K, Dixit SB, Beveridge DL. Free Energy Analysis of Protein-DNA Binding: The EcoRI Endonuclease - DNA Complex. J Comput Phys 1999; 151; 333-357.

147. Honig B, Nicholls A. Classical Electrostatics in Biology and Chemistry. Science 1995; 268: 1144-1149.

148. Åqvist J, Marelius, J. The Linear Interaction Energy Method for Predicting Ligand Binding Free Energies. Combin Chem High Throughput Screen. 2001; 4: 613-626.

149. Åqvist J, Luzhkov VB, Brandsdal BO. Ligand binding affinities from MD simulations Acc Chem Res 2002; 35: 358-365.

150. Oostenbrink C, van Gunsteren WF. Chemical Theory and Computation Special Feature: Free energies of ligand binding for structurally diverse compounds. Proc Natl Acad Sci 2005; 102; 6750-6754.

151. Reddy MR, Erion MD, Agarwal A. Free Energy Calculations: Use and Limitations in Predicting Ligand Binding Affinities. Rev Comput Chem, 6, VCH Publishers: New York, 2000, 217-304.

152. Brandsdal BO, Osterberg F, Almlof M, Feierberg I, Luzhkov VB, Aqvist J. Free Energy Calculations and Ligand Binding. Adv Protein Chem 2003; 66: 123-158.

153. Wang W, Donini O, Reyes CM, Kollman PA. Biomolecular Simulations: Recent Developments in Force Fields, Simulation of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions. Annu Rev Biophys Biomol Struct 2001; 30: 211-243.

154. van Gunsteren WF, Berendsen HJC. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. Angew Chem Int Ed Engl 1990; 29: 992-1023.

155. Lybrand TP. Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods. Rev Comput Chem 1990; 1: 295-320.

156. Gilson MK, Given AJ, Bush BL, McCammon JA. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. Biophys J 1997; 72: 1047-1069.

157. Bash PA, Singh UC, Langridge R, Kollman PA. Free Energy Calculation by Computer Simulation. Science 1987; 236: 564-568.

158. Mitchell MJ, McCammon JA. Free Energy Difference Calculations by Thermodynamic Integration: Difficulties in Obtaining a Precise Value. J Comp Chem 1991; 12: 271-275.

159. Rohs R, Bloch I, Sklenar H, Shakked Z. Molecular Flexibility in *Ab Initio* Drug Docking to DNA: Binding-Site and Binding-Mode Transitions in All-Atom Monte Carlo Simulations. Nucleic Acids Res 2005; 33: 7048-7057.

160. DA Gschwend, W Sirawaraporn, DV Santi, ID Kuntz. Specificity in structure-based drug design: Identification of a novel, selective inhibitor of *Pneumocystis carinii* dihydrofolate reductase. Proteins: Struc Func Genet 1998; 29: 59-67.

161. E. Friere. Designing drugs against heterogeneous *targets.* Nature Biotech, 2002; 20:15-16.

162. Gombar VK, Silver IS, Zhao Z. Role of ADME Characteristics in Drug Discovery and Their *In Silico* Evaluation: *In Silico* Screening of Chemicals for Their Metabolic Stability. Curr Topics Med Chem 2003; 3: 1205-1225.

163. Ajay A, Walters WP, Murcko MA. Can We Learn to Distinguish between "Drug-Like" and "Nondrug-Like" Molecules? J Med Chem 1998; 41: 3314-3324.

164. Davis AM, Riley RJ. Predictive ADMET Studies, the Challenges and the Opportunities. Curr Opin Chem Biol 2004; 8: 378-386.

165. Butina D, Segall MD, Frankcombe K. Predicting ADME Properties *In Silico*: Methods and Models. Drug Discov Today 2002; 7: S83-S88.

166. Weaver DC. Applying Data Mining Techniques to Library Design, Lead Generation and Lead Optimization. Curr Opin Chem Biol 2004; 8: 264-270.

167. Singh SS. Preclinical Pharmacokinetics: An Approach Towards Safer and Efficacious Drugs. Curr Drug Metab 2006; 7:165-182.

168. Patel DS, Bharatam PV. New Leads For Selective GSK-3 Inhibition: Pharmacophore Mapping and Virtual Screening Studies. J Comput Aided Mol Des 2006; 20:55-66.

169. Sadowski J, Kubinyi H. A Scoring Scheme for Discriminating between Drugs and Nondrugs. J Med Chem 1998; 41: 3325-3329.

170. Anzali S, Barnickel G, Cezanne B, Krug M, Filimonov D, Poroikov V. Discriminating between Drugs and Nondrugs by Prediction of Activity Spectra for Substances (PASS). J Med Chem 2001; 44: 2432-2437.

171. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001; 46: 3-26.

172. Alavijeh MS, Chishty M, Qaiser MZ, Palmer AM. Drug Metabolism and Pharmacokinetics, the Blood-Brain Barrier, and Central Nervous System Drug Discovery. NeuroRx 2005; 2: 554-571.

173. Griffin WC. Classification of Surface-Active Agents by HLB. J Soc Cosmet Chem 1949; 1: 311.

174. Weinheim. Methods and Principles in Medicinal Chemistry. Wiley-VCH 2003; 18: 72-89.

175. Norinder U, Osterberg T. Theoretical Calculation and Prediction of Drug Transport Processes Using Simple Parameters and Partial Least Squares Projections to Latent Structures (PLS) Statistics. The Use of Electrotopological State Indices. J Pharm Sci 2001; 90: 1076-1085.

176. Stenberg P, Norinder U, Luthman K, Artursson P. Experimental and Computational Screening Models for the Prediction of Intestinal Drug Absorption. J Med Chem 2001; 44: 1927-1937.

177. Winiwarter, Bonham S, Ax NM, Hallberg F, Lennernaes A, Karlen HA. Correlation of Human Jejunal Permeability (*in vivo*) of Drugs with Experimentally and Theoretically Derived Parameters. A Multivariate Data Analysis Approach. J Med Chem 1998; 41: 4939-4949.

178. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular Properties that Influence the Oral Bioavailability of Drug Candidates. J Med Chem 2002; 45: 2615-2623.

179. Kustrin SA, Beresford R, Yusof APM. Theoretically Derived Molecular Descriptors Important in Human Intestinal Absorption. J Pharm Biomed Anal 2001; 25: 227-237

180. Agoram B, Woltosz WS, Bolger MB. Predicting The Impact Of Physiological And Biochemical Processes On Oral Drug Bioavailability. Adv Drug Deliv Rev 2001:50: S41-S67.

181. Norris DA, Leesman GD, Sinko PJ, Grass GM. Development of Predictive Pharmacokinetic Simulation Models For Drug Discovery. J Contr Rel 2000; 65: 55-62. (2000).

182. Zhang EY, Phelps MA, Cheng C, Ekins S, Swaan PW. Modeling of Active Transport Systems. Adv Drug Deliv Rev 2002; 54: 329-354.

183. Galinsky RE, Svensson CK. In: Gennaro AR, et al, Eds. Remington: The Science and Practice of Pharmacy. Baltimore, Lippincott Williams and Wilkins. 2000; 1125-1144.

184. van de Waterbeemd H, Smith DA, Jones BC. Lipophilicity in pK: Methyl, Ethyl, Futile. J Comput Aided Mol Des 2001; 15: 273-286.

185. Lombardo F, Gifford E, Shalaeva MY. *In Silico* ADME Prediction: Data, Models, Facts And Myths. Mini Rev Med Chem. 2003; 3:861-875.

186. Colmenarejo G. *In Silico* Prediction of Drug-Binding Strengths To Human Serum Albumin. Med Res Rev 2003; 3: 275-301.

187. Hall LM, Hall, LH, Kier LB. Modeling Drug Albumin Binding Affinity with E-State Topological Structure Representation J Chem Inf Comput Sci 2003; 43: 2120-2128.

188. Saiakhov RD, Stefan LR, Klopman G. Multiple Computer-Automated Structure Evaluation Model Of The Plasma Protein Binding Affinity Of Diverse Drugs. Perspect Drug Discov Design. 2000; 19: 133-155.

189. Norinder U, Haeberlein M. Computational approaches to the prediction of the blood-brain distribution. Adv Drug Deliv Rev 2002; 54: 291-313.

190. Ecker GF, Noe CR. *In Silico* Prediction Models For Blood-Brain Barrier Permeation. Curr Med Chem 2004; 11: 1617-1628.

191. Kelder J, Grootenhuis PDJ, Bayada DM, Delbressine LPC, Ploemen JP. Polar Molecular Surface As A Dominating Determinant For Oral Absorption And Brain Penetration of Drugs. Pharm Res 1999; 16: 1514-1519.

192. van de Waterbeemd H, Camenisch G, Folkers G, Chretien JR, Raevsky OA. Estimation of Blood-Brain Barrier Crossing of Drugs Using Molecular Size and Shape, and H-Bonding Descriptors. J Drug Target 1998; 6:151-165.

193. Doan KMM, Humphreys JE, Webster LO, Wring SA, Shampine LJ, Serabjit-Singh CJ, Adkison KK, Polli JW. Passive Permeability and P-Glycoprotein-Mediated Efflux Differentiate Central Nervous System (CNS) and Non-CNS Marketed Drugs J Pharmacol Exp Ther. 2002; 303: 1029-1037.

194. Moriguchi I, Hirono S, Liu Q, Nakagome I, Matsushuta Y. Simple Method Of Calculating Octanol Water Partition-Coefficient. Chem Pharm Bull. 1992; 40: 127-130.

195. Leeson PD, Davis, AM. Time-Related Differences in the Physical Property Profiles of Oral Drugs. J Med Chem 2004, 47: 6338-6348.

196. Clark DE. *In Silico* Prediction of Blood-Brain Barrier Permeation. Drug Discov. Today 2003; 8: 927-933.

197. Subramanian G, Kitchen DB. Computational Models to Predict Blood-Brain Barrier Permeation and CNS Activity. J Comput Aided Mol Des 2003; 17: 643-664.

198. Cabrera MA, Sanz MB. *In Silico* Prediction of Central Nervous System Activity Of Compounds. Identification of Potential Pharmacophores By The TOPS-MODE Approach. Bioorg Med Chem 2004; 12: 5833-5843.

199. Cabrera MA, Bermejo M, Perez M, Ramos R. TOPS-MODE approach for the prediction of blood-brain barrier permeation. J Pharm Sci 2004; 93: 1701-1717.

200. Ekins S, Bravi G, Binkley S, Gillespie JS, Ring BJ, Wikel JH, Wrighton SA Three- and Four-Dimensional-Quantitative Structure Activity Relationship (3D/4D-QSAR) Analyses of CYP2C9 Inhibitors. Drug Metab Dispos 2000; 28: 994-1002.

201. de Groot MJ, Ackland MJ, Horne VA, Alex AA , Jones BC. Novel Approach To Predicting P450-Mediated Drug Metabolism: Development of a Combined Protein and Pharmacophore Model For CYP2D6. J Med Chem 1999; 42: 1515-1524.

202. Ekins S, de Groot MJ, Jones JP. Pharmacophore and Three-Dimensional Quantitative Structure Activity Relationship Methods for Modeling Cytochrome P450 Active Sites. Drug Metab Dispos 2001; 29: 936-944.

203. Madden JC, Cronin MT. Structure-Based Methods For the Prediction of Drug Metabolism. Expert Opin Drug Metab Toxicol. 2006; 2: 545-557.

204. Ekins S, Andreyev S, Ryabov A, Kirillov E, Rakhmatuli EA, Bugrim A, Nikolskaya T. Computational Prediction of Human Drug Metabolism. Expert Opin Drug Metab Toxicol 2005; 1: 303-324.

205. Vermeulen NP. Prediction Of Drug Metabolism: The Case of Cytochrome P450 2D6. Curr Top Med Chem 2003; 3: 1227-1239.

206. Frank R, Hargreaves R. Clinical Biomarkers In Drug Discovery And Development. Nat Rev Drug Discov 2003; 2: 566-580.

207. Wessel MD, Mente S. ADME by Computer. Ann Reports Med Chem 2001; 36: 257-266.

208. Lin JH, Applications and Limitations Inter-Species Scaling and *In Vitro* Extrapolation In Pharmacokinetics. Drug Metab Dispos 1998, 26, 1202-1212.

209. Mahmood I, Balian JD. The Pharmacokinetic Principles Behind Scaling from Preclinical Results to Phase I Protocols. Clin Pharmacokinet 1999; 36: 1-11.

210. Green N. Computer Systems For The Prediction of Toxicity: An Update. Adv Drug Deliv Rev 2002; 54: 417-431.

211. Copeland RA, Pompliano DL, Meek TD. Drug-target residence time and its implications for lead optimization. Nat Rev Drug Discov 2006; 5: 730-739.

212. Gossas T, Danielson UH Analysis of The Ph-Dependencies Of The Association And Dissociation Kinetics Of HIV-1 Protease Inhibitors J Mol Recognit 2003; 16: 203-212.

213. Borgna JL, Ladrech S. The Dissociation Rate of Estrogen Receptor-Ligand Complexes Is Increased By High Concentrations Of Steroids and Antiestrogens. Mol Cell Endocrinol 1982; 27:1-15.

214. Zhang Y, McCammon JA. Studying the Affinity and Kinetics of Molecular Association with Molecular Dynamics Simulation. J Chem Phys 2003; 118: 1821-1827.

215. Choulier L, Andersson K, Hamalainen MD, van Regenmortel MHV, Malmqvist M, Altschuh D QSAR studies applied to the prediction of antigen-antibody interaction kinetics as measured by BIACORE. Protein Eng Des Sel. 2002; 15: 373 - 382.

216. Steullet V, Dixon DW, Takenaka S, Wilson WD. Studies of Naphthalene Diimides as DNA-binding Agents. First International Electronic Conference on Synthetic Organic Chemistry (ECSOC-1), 1997, Sept. 1-30, E003.

217. Day YSN, Baird CL, Rich RL, Myszka DG. Direct Comparison of Binding Equilibrium, Thermodynamic, and Rate Constants Determined by Surface- and Solution-Based Biophysical Methods. Protein Sci 2002; 11: 1017-1025.

218. Hood WF, Gierse JK, Isakson PC, Kiefer JR, Kurumbail RG, Seibert K, Monahan JB. Characterization of Celecoxib and Valdecoxib Binding to Cyclooxygenase. Mol Pharmacol 2003; 63: 870-877.

219. Thurmond RL, Wadsworth SA, Schafer PH, Zivin RA, Siekierka JJ. Kinetics of Small Molecule Inhibitor Binding to p38 Kinase. Eur J Biochem 2001; 268, 5747-5754.

220. Mangold U, Dax CI, Saar K, Schwab W, Kirschbaum B, Müllner S. Identification and Characterization of Potential New Therapeutic Targets in Inflammatory and Autoimmune Diseases. Eur J Biochem 1999; 266: 1184-1191.

221. Gillard M, van Der Perren C, Moguilevsky N, Massingham R, Chatelain P. Binding Characteristics of Cetirizine and Levocetirizine to Human H1 Histamine Receptors: Contribution of Lys191 and Thr194. Mol Pharmacol 2002; 61:391-399.

222. Talbert AM, Tranter GE, Holmes E, Francis PL. Determination of Drug-Plasma Protein Binding Kinetics and Equilibria by Chromatographic Profiling: Exemplification of the Method Using L-Tryptophan and Albumin. Anal Chem 2002; 74: 446-452.