Virtual screening forms an alternative approach and uses computer-logical target and is still the dominant technique in drug discovery. Screening of large libraries of chemical compounds against 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 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 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 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 \textit{in silico} ADMET profiling studies and those molecules that pass these studies are termed as 'leads'.

![Fig. (1). A flowchart outlining a plausible generalized structure-based \textit{in silico} drug discovery strategy.](image)

2. \textsc{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.

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 discov-

![Fig. (2). A computational strategy and considerations for obtaining lead-like molecules \textit{in silico} (Stage I).](image)
ery/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].

Fig. (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 \textit{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

\textbf{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 \textit{in silico} combinatorial methods involving fragments or templates derived from databases. This technique forms the basis of \textit{de novo} design [50-54].

\textbf{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 data base should be restricted.

2.2.2. Development of a Non-Redundant Database of Small Molecules (NRDBSM)

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 precints 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].

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 \textit{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 \textit{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.
A physicochemically rigorous and rapid computational method for binding affinity prediction or scoring will have widespread 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

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 binding, 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 mis-docked 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)).

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 or Monte Carlo methods. 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

![Fig. (6)](Image)

Fig. (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) (ICQP; 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 (1SEL; Transport protein) and Drug 3 is n-(4-hydroxyphenyl)all-trans retinamide (fenretinide). Target 4 is human cardiac troponin C (1LXF; Metal bind-
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.scfbiio-iitd.res.in/software/drugdesign/bappl.jsp and http://www.scfbiio-iitd.res.in/predicta). 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 follow in computer-aided drug design.

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 genes. 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 specifically and effectively to the site of action. It should not bind to, or interact with related receptors and nor 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 ADMET 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].

Fig. (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 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.

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 ADMET 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.

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],

\[ A = \left( \frac{D_{\text{Do}}}{D_{\text{Div}}} \right) \times 100 \]

where \( D_{\text{Do}} \) is the drug distribution after oral administration, \( D_{\text{Div}} \) 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],

\[ V_d \text{ (in litres)} = \frac{D_{\text{body}}}{D_{\text{plasma}}} \]

where \( D_{\text{body}} \) is the amount of drug in body (mg), \( D_{\text{plasma}} \) 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, \( V_u \) against distribution D at pH 7.4 (with the data corrected for plasma-protein binding), reveals a clear linear trend, with log \( V_u \) increasing at higher lipophilicities [184]. This can be used as a simple guide in modifying and optimizing the \( V_u \). 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 QSAR 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 en-
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.

### 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],

\[
Cl \text{ (L/hr)} = \frac{Re}{D_{\text{plasma}}}
\]

where \(Re\) is the rate of elimination (mg/hr), \(D_{\text{plasma}}\) is drug concentration in blood plasma (mg/L).

\(Re\) is given by:

\[
Re = k_e \times D_{\text{body}}
\]

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

Thus,

\[
Cl = \frac{(0.693 \times V_d)}{t_{1/2}}
\]

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 (\(q^2 = 0.682, \text{RMSE} = 0.35\) [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].

#### 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].

\[C_t = [C_0] \times e^{(-k_d t)}\]

where \([C_t]\) is the concentration of the drug in complexed form at time, \(t\); \([C_0]\) is the concentration of the drug at \(t=0\), \(k_d\) is the dissociation rate constant.
Table 1. Drug Metabolizing Enzymes with their Family and Availability of 3D Structures

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Drug Metabolizing Enzyme</th>
<th>Family</th>
<th>Structure in PDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Human Cytidine deaminase</td>
<td>Hydrolases</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Cholinesterase</td>
<td>Hydrolases</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>ECOLI Beta-lactamase</td>
<td>Hydrolases</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Human Adenosine deaminase</td>
<td>Hydrolases</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>Human Pancreatic alpha-amylose precursor</td>
<td>Hydrolases</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>Human Arylsulfatase A precursor</td>
<td>Hydrolases</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>Human Liver carboxylesterase 1 precursor</td>
<td>Hydrolases</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>Human Glutamine synthetase</td>
<td>Ligases</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>Human Cytochrome P450 3A4</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>Human Cytochrome P450 2D6</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>Human Cytochrome P450 2C19</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>Human Cytochrome P450 2B6</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>Human Amine oxidase</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>14</td>
<td>Human Cytochrome P450 2C9</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>15</td>
<td>Cytochrome P450 19</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>16</td>
<td>Aldehyde oxidase and P450</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>17</td>
<td>Human Aldehyde oxidase</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>Human Cytochrome P450 1A2</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>19</td>
<td>Cytochrome P450 3A4</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>Human Cytochrome P450 2C19</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>21</td>
<td>Human Cytochrome P450 2C8</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>22</td>
<td>Human Cytochrome P450 CP2D6</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>23</td>
<td>Human Cytochrome P450 CYP2D6</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>24</td>
<td>Human Cytochrome P450 2A6</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>25</td>
<td>Human Cytochrome P450 2E1</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>26</td>
<td>Human Cytochrome P450 2A13</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>27</td>
<td>Human Alcohol dehydrogenase 6</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>28</td>
<td>Human Cytochrome P450 11A1</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>29</td>
<td>Human Cytochrome P450 24A1</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>30</td>
<td>Human Cytochrome P450 1A1</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>31</td>
<td>Human Cytochrome P450, subfamily IIIA</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>32</td>
<td>Human Xanthine dehydrogenase/oxygenase</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>33</td>
<td>Human Cytochrome P450 3A4</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>34</td>
<td>Human Cytochrome P450 1A2</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>35</td>
<td>Human Cytochrome P450 11B2</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>36</td>
<td>RAT Cytochrome P450 3A1</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>37</td>
<td>RAT Cytochrome P450 2C11</td>
<td>Oxidoreductase</td>
<td>No</td>
</tr>
<tr>
<td>38</td>
<td>Human Carbonyl reductase</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>39</td>
<td>Human Proline oxidase</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>40</td>
<td>Human Tryptophan 2,3-dioxygenase</td>
<td>Oxidoreductase</td>
<td>Yes</td>
</tr>
<tr>
<td>41</td>
<td>Aminoglycoside 2’-N-acetyltransferase</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
<tr>
<td>42</td>
<td>Kanamycin nucleotidyltransferase</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
<tr>
<td>43</td>
<td>Aminoglycoside 3’-phosphotransferase</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
<tr>
<td>44</td>
<td>Human Glutathione S-transferase A</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
<tr>
<td>45</td>
<td>Human Glutathione S-transferase A2</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
<tr>
<td>46</td>
<td>COMT ( catecol-O-methyl-transferase)</td>
<td>Transferases</td>
<td>No</td>
</tr>
<tr>
<td>47</td>
<td>Human Nucleoside diphosphate kinase A</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
<tr>
<td>48</td>
<td>Human Thymidine phosphorylase</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
</tbody>
</table>
(Table 1) contd….

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Drug Metabolizing Enzyme</th>
<th>Family</th>
<th>Structure in PDB</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>Human Deoxycytidine kinase</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
<tr>
<td>50</td>
<td>Human Histamine N-methyltransferase</td>
<td>Transferases</td>
<td>Yes</td>
</tr>
<tr>
<td>51</td>
<td>UDP Glucosyltransferases</td>
<td>Transferases</td>
<td>No</td>
</tr>
</tbody>
</table>

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

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

\[ t_R = \frac{1}{k_d} \]

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].

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 an hour. This however, is only an upper limit based on in vitro studies.

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

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 (Fig. 1). 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 Darab 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 Afsan Shaikh is a recipient of Senior Research Fellowship from CSIR, Govt. of India.

REFERENCES


From Drug Target to Leads-Sketching A Physicochemical Pathway


From Drug Target to Leads-Sketching A Physicochemical Pathway


