ABSTRACT
The development of new drugs with potential therapeutic applications is one of the most complex and difficult processes in the pharmaceutical industry. Millions of dollars and man-hours are devoted to the discovery of new therapeutically agents. As, the activity of a drug is the result of a multitude of factors such as bioavailability, toxicity and metabolism, rational drug design has been utopias for centuries. Very recently, impressive technological advances in areas such as structural characterization of biomacromolecules, computer sciences and molecular biology have made rational drug design feasible. The aim of this paper is to give an outline of studies in the field of medicinal chemistry in which molecular modeling has helped in the discovery process of new drugs. The emphasis will be on lead generation and optimization.

KEYWORDS: Virtual screening, ADME/TOX predictions, structure-based design, ligand-based design.

INTRODUCTION
Drug discovery in the 21st century has been greatly enhanced by the additional tools made available through information technology and the increase in computer processor efficiency. In the information-rich post-genomic era, expectations are high for identifying new targets and the rapid development of effective treatments with low side effects.

Animal Health Care Market
The animal health care market could benefit substantially from more streamlined and economic drug discovery processes. There are important challenges in veterinary medicine including maintaining the health of 3.3 billion livestock animals and 16 billion poultry worldwide, which is critical for human health in this age of antibiotic-resistance and emerging diseases that could pose threats to the food supply. In addition, there are 1 billion companion animals that require traditional veterinary treatments such as for parasites but also increasingly are being treated for diseases associated with aging and more recently for behavioural disorders. As a result, the companion animal market has grown strongly over the past decade as pet owners are more willing to spend money on veterinary health care and the availability of therapeutics for heartworm, flea and tick control, non steroidal anti-inflammatory agents for...
canine arthritis as well as behavioural drugs \cite{1}. Despite this growth, the animal health care market is still a small percentage of the human market and cannot support its own primary research.

**Stages of Drug Discovery Process**

Drug discovery is a process that includes identification of a target, development of an assay of target function, screening of compounds and natural products, lead identifications, and lead optimizations. This is followed by animal studies to measure absorption, distribution, metabolism and excretion (ADME) properties as well as toxicity. The time from the target identification to approval of a new drug is typically 10 to 15 years (Figure 1).

![Timeline in a drug discovery project](image)

**Figure 1: Timeline in a drug discovery project**

The overall estimated cost to bring a drug to market is now $800 million or Rs 4 billion (Figure 2) \cite{2}. One of the factors contributing to this high cost is the large number of lead compounds that fail late in the drug discovery process due to either poor ADME/TOX or adverse side effects such as induction of long QT interval. Increasing efficiency of drug discovery by making the overall time for drug development shorter so that the patent life of a compound is extended and elimination of compounds from further development early on in the discovery process that will have poor ADME/TOX properties or undesirable side effects are the major challenges that could enhance profitability of drug development. The drug discovery process, showing the more recent evolution where target equity and compound equity are parallel early key components and lead equity is now considered. Both of these contribute to efforts to tackle productivity and attrition (risk management) \{Figure 3\}. 
Figure 2: The cost of drug development from $4 million in 1962 to over $350 million in 1996

Figure 3: The drug discovery process, showing the more recent evolution where target equity and compound equity are parallel early key components, and lead equity is now considered. Both of these contribute to efforts to tackle productivity and attrition (risk management).
HISTORY OF COMPUTER-AIDED DRUG DESIGN

Computational chemistry is a relatively new discipline and is the foundation of computer-aided drug design (Figure 4). One of the first major advances that led to the development of many of the most powerful techniques in computer-aided drug design today was the development of the quantitative structure activity relationship (QSAR) analysis by Hansch and Fujita which described a new method for analyzing drug actions [3]. This was followed by the development of molecular mechanics by Allinger in 1971 which is the major foundation for energy-based minimizations of molecules [4]. In 1977, Garland Marshall described the active analog approach, another breakthrough in computer-aided drug design and shortly thereafter established the computational chemistry/drug discovery software company, Tripos [5]. Peter Kohlman developed the AMBER force field in 1981 which allowed for energy minimizations of large protein molecules [6].

An algorithm for docking small molecules to receptors that later became the powerful DOCK program was developed by Kuntz in 1982 [7]. In 1984, partial least squares analysis was introduced. This method is commonly employed in QSAR studies today as it allows for the derivation of linear equations from data tables that have more columns than rows [8]. Robert Pearlman published the first description of CONCORD, a program that allowed for the rapid approximation of 3D structures of molecules [9]. The first description of comparative molecular field analysis (CoMFA), a QSAR technique that explicitly incorporates 3D geometries of small molecules and relates them to activity was published by Richard Cramer in 1988 [10]. These are only a few of the breakthroughs that have contributed to modern computer-aided drug design. In addition to innovations in the way we think about drug design, availability of high resolution structural information from X-ray crystallography and NMR, the vast amount of information available from the genomic databases and the related discipline of bioinformatics and the enhancements of computer processors and graphical interfaces have also been key in advancing computer-aided drug design. A multidisciplinary approach to drug design that truly integrates all of these facets is required to address the challenges of drug discovery in the 21st century (Figure 5).
Figure 4: An example of a computer-aided drug design flowchart.
STRUCTURE-BASED DRUG DESIGN

Most drugs on the market today were found either serendipitously or by screening large numbers of natural products and synthetic substances (Figure 6, 7 and 8). These novel compounds were then improved by synthesizing analogs in hopes of enhancing efficacy or reducing unfavorable side effects. In the post-genomic era, where specific drug targets can be identified and their three-dimensional structures determined either by X-ray crystallography or NMR, structure-based design of drugs based on the principles of molecular recognition has become a new paradigm in drug discovery [11]. Ondetti and Cushman were the first to successfully utilize X-ray crystallographic data in drug design. While they didn’t know the structure of their intended target, human angiotensin-converting enzyme (ACE), they used the structure of a related protein as a model to develop the first ACE inhibitor, Captopril [12]. Similar strategies were used to develop inhibitors of HIV protease [13]. The key requirement for structure-based design is having a high resolution structure for the target protein or of a closely related protein, preferably with a bound ligand, to identify the drug receptor site. Once known, the structure of the receptor site can be used to define a pharmacophore for virtual screening of libraries and in docking studies which can be used to design improvements in lead compounds. Finding the active site of the target protein is necessary for structure-based design of drugs (Figure 9). Homology-modeling of related pro-
proteins where the active site is known is the preferred method. Other methods for predicting active sites include algorithms that predict solvent accessible surfaces or pockets of proteins and those that evaluate the solvent accessibility, hydrophilicity, lipophilicity and clustering algorithms to define potential binding sites. There are two major types of drug design. The first is referred to as **structure-based drug design** (Figure 10) and the second, **ligand-based drug design**.

Figure 6: The Iterative Process of Structure-Based Drug Design

Figure 7: Flow chart for structure-based drug design
Figure 8: Flowchart of a usual clustering analysis for structure-based drug design

Figure 9: Protein structure-based drug design cycle. Lead compounds originate from either random screening of a few hundred thousand compounds or from design. In the latter case, synthesis can be bypassed by using docking of compounds available commercially or in-house. Design is the result of
docking, linking and building, or any combination of the three. Due to the imperfections of computer scoring, only about 2% of the designed compounds pass the first criterion to become a lead, namely having micromolar affinity. Verification of the structure of the protein–lead complex is essential. New rounds of structure-based design are then performed until a promising compound shows up for pre-clinical trials. At this stage the structure is still useful: knowledge of the essential protein–ligand interactions dictates where structural modifications to improve the pharmacodynamic properties should not be made. After successful clinical trials a new drug is born.

Figure 10: Flow charts of two strategies of structure-based drug design

Structure-based drug design (or direct drug design) relies on knowledge of the three-dimensional structure of the biological target obtained through methods such as x-ray crystallography or NMR spectroscopy. If an experimental structure of a target is not available, it may be possible to create a homology model of the target based on the experimental structure of a related protein. Using the structure of the biological target, candidate drugs that are predicted to bind with high affinity and selectivity to the target may be designed using interactive graphics and the intuition of a medicinal chemist. Alternatively various automated computational procedures may be used to suggest new drug candidates.

As experimental methods such as X-ray crystallography and NMR develop, the amount of information concerning 3D structures of biomolecular targets has increased dramatically. In parallel, information about the structural dynamics and electronic properties about ligands has also increased. This has encouraged the rapid development of the structure-based drug design. Current methods for struc-
structure-based drug design can be divided roughly into two categories. The first category is about “finding” ligands for a given receptor, which is usually referred as database searching. In this case, a large number of potential ligand molecules are screened to find those fitting the binding pocket of the receptor. This method is usually referred as ligand-based drug design. The key advantage of database searching is that it saves synthetic effort to obtain new lead compounds. Another category of structure-based drug design methods is about “building” ligands, which is usually referred as receptor-based drug design. In this case, ligand molecules are built up within the constraints of the binding pocket by assembling small pieces in a stepwise manner. These pieces can be either individual atoms or molecular fragments. The key advantage of such a method is that novel structures, not contained in any database, can be suggested \[16, 17\].

**Active site identification**

Active site identification is the first step in this program. It analyzes the protein to find the binding pocket, derives key interaction sites within the binding pocket, and then prepares the necessary data for Ligand fragment link. The basic inputs for this step are the 3D structure of the protein and a pre-docked ligand in PDB format, as well as their atomic properties. Both ligand and protein atoms need to be classified and their atomic properties should be defined, basically, into four atomic types:

1. **hydrophobic atom**: All carbons in hydrocarbon chains or in aromatic groups.
2. **H-bond donor**: Oxygen and nitrogen atoms bonded to hydrogen atom(s).
3. **H-bond acceptor**: Oxygen and sp2 or sp hybridized nitrogen atoms with lone electron pair(s).
4. **Polar atom**: Oxygen and nitrogen atoms that are neither H-bond donor nor H-bond acceptor, sulfur, phosphorus, halogen, metal, and carbon atoms bonded to hetero-atom(s).

The space inside the ligand binding region would be studied with virtual probe atoms of the four types above so the chemical environment of all spots in the ligand binding region can be known. Hence we are clear what kind of chemical fragments can be put into their corresponding spots in the ligand binding region of the receptor.

**LIGAND-BASED DRUG DESIGN**

Ligand-based drug design (or indirect drug design) relies on knowledge of other molecules that bind to the biological target of interest. These other molecules may be used to derive a pharmacophore model that defines the minimum necessary structural characteristics a molecule must possess in order to bind to the target. In other words, a model of the biological target may be built based on the knowledge of what binds to it, and this model in turn may be used to design new molecular entities that interact with the target. Alternatively, a quantitative structure-activity relationship (QSAR) in which a correlation between calculated properties of molecules and their experimentally determined...
biological activity, may be derived. These QSAR relationships in turn may be used to predict the activity of new analogs.

**Virtual screening**

Virtual screening of library compounds is a complementary approach to high throughput screening in the process of lead identification. Defining the pharmacophore, the steric and electrostatic features and their arrangement in space that are required for high affinity binding is a key element for virtual or in silico screening of compounds. The pharmacophore is used as a template for searching virtual libraries of compounds, often using successive “filters” to continue to reduce the number of compounds that will be actually used in a high throughput screening. This can reduce the cost of the actual high throughput screen by reducing a library of 100,000 compounds to 3,000 that meet the pharmacophore criteria \[18\]. Recently, the Kunz research group showed how structure-based design could start with calculating free energies of binding of a combinatorial library with cathepsin D, an aspartyl protease responsible for cleavage of β-amyloid peptide, using the molecular dynamics-based continuum solvent method (MM-PBSA) \[19\]. They were able to predict binding affinities for a set of seven inhibitors within 1 kcal/mol. The molecular dynamics simulations predict a binding conformation of the inhibitors that is in close agreement with the X-ray crystal structure of a peptide inhibitor-cathepsin D complex. In addition, they were able to identify substitutions that improved inhibitor binding. This work demonstrates the utility of virtual screening in a multi-step structure-based drug design process.

<p>| Table 1: Algorithms for docking small molecules or fragments against a target |
|-----------------------------|-----------------------------|-----------------------------|-----------------------------|
| <strong>Program</strong> | <strong>Flexible Protein</strong> | <strong>Flexible Ligand</strong> | <strong>Description</strong> |
| Virtual screening | DOCK | no | yes | docks either small molecules or fragments, includes solvent effects |
| | FlexX | no | yes | incremental construction |
| | FlexE | yes | yes | incremental construction; samples ensembles of receptor structures |
| | SLIDE | yes | yes | anchor fragments placed, remainder of ligand added; backbone flexibility |
| | Flo98 | no | yes | can rapidly dock a large number of ligand molecules, graphically view results |
| | ADAM | no | yes | fragments aligned based on hydrogen bonding |
| | Hammerhead | no | yes | genetic algorithms to link tail fragments to anchor fragments |</p>
<table>
<thead>
<tr>
<th>Software</th>
<th>De novo generation of ligands</th>
<th>MCSS</th>
<th>传说</th>
<th>SMoG</th>
<th>CONCERTS</th>
<th>Legend</th>
<th>DLD</th>
<th>GrowMol</th>
<th>GenStar</th>
<th>GROW</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCSA-PCR</td>
<td>yes</td>
<td>yes</td>
<td>uses simulated annealing to generate conformations of target</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUTODOCK</td>
<td>yes</td>
<td>yes</td>
<td>uses averaged interaction energy grid to account for receptor conformations and simulated annealing for ligand conformations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCDOCK</td>
<td>no</td>
<td>yes</td>
<td>Monte Carlo to sample ligand placement</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProDOCK</td>
<td>yes</td>
<td>yes</td>
<td>Monte Carlo minimization for flexible ligand, flexible site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICM</td>
<td>yes</td>
<td>yes</td>
<td>Monte Carlo minimization for protein-ligand docking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DockVision</td>
<td>no</td>
<td>no</td>
<td>Monte Carlo minimization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LUDI</td>
<td>no</td>
<td>yes</td>
<td>docks and scores fragments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRID</td>
<td>no</td>
<td>yes</td>
<td>calculates binding energies for functional groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCSS</td>
<td>no</td>
<td>yes</td>
<td>exhaustive search of binding site for functional group minima</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMoG</td>
<td>no</td>
<td>yes</td>
<td>knowledge-based scoring function; molecules built by joining rigid fragments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CONCERTS</td>
<td>no</td>
<td>yes</td>
<td>fills active site with molecular fragments, links fragments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Legend</td>
<td>no</td>
<td>yes</td>
<td>grows molecule atom by atom</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DLD</td>
<td>no</td>
<td>yes</td>
<td>saturates binding site with sp3 carbons, later linked</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GrowMol</td>
<td>no</td>
<td>yes</td>
<td>builds ligands from a library of atom types</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GenStar</td>
<td>no</td>
<td>yes</td>
<td>builds ligands from sp3 carbons</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GROW</td>
<td>no</td>
<td>yes</td>
<td>constructs a peptide by residue addition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GroupBuild</td>
<td>no</td>
<td>yes</td>
<td>builds ligand from a predefined library of fragments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>----</td>
<td>-----</td>
<td>--------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOOK</td>
<td>no</td>
<td>yes</td>
<td>searches database of molecular skeletons for fit to binding site; hooks two MCSS functional groups to skeleton</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPROUT</td>
<td>no</td>
<td>yes</td>
<td>generates skeletons that fit site, substitutes atoms into skeleton to give molecule with correct properties</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAVEAT</td>
<td>no</td>
<td>yes</td>
<td>searches database of small molecules to connect fragments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**LIGAND-BASED DESIGN**

**Three-dimensional quantitative structure-activity relationship techniques (3D QSAR)**

QSAR techniques have been important in the design of pharmaceuticals since they were first proposed by Hansch and Fujita in 1964 [3]. More recently, QSAR analyses of ligand receptor interactions have included three dimensional properties of molecules such as comparative molecular field analysis (CoMFA) [10], comparative molecular similarity indices analyses (CoMSIA) [20] and comparative molecular surface analysis (COMSA) [21]. CoMFA is based on the premise that steric and electrostatic fields around an aligned set of molecules can be used to predict biological activity using partial least squares analysis. This technique has been successfully employed to develop predictive models of activity for a wide range of compounds and, although not always successful, it has become a standard tool in computer-aided drug discovery. In CoMSIA, similarity is expressed in terms of different physiochemical properties like steric occupancy, H-bond donor-acceptor properties, local hydrophobicity and partial atomic charges and uses a Gaussian type distance dependent function as opposed to the grid approach taken in CoMFA. COMSA is based on the mean electrostatic potential along with a neural network approach and partial least squares analysis. These methods vary in their success and are often used in combination with other techniques to help establish their validity. Recent studies which combined CoMFA, CoMSIA and docking studies to design selective COX-2 inhibitors demonstrate how using multiple approaches in computer aided drug design are particularly effective [22]. Another novel use of CoMFA published recently showed how 3D QSAR can be used to identify a pharmacophore for LQT-inducing effects from a set of chemically diverse compounds [23].

**PREDICTIVE MODELS OF ADME/TOX**

Drug-like properties include aqueous solubility, ability to cross membranes, metabolic stability and safety. These properties are described by the absorption, distribution, metabolism, excretion and toxicity (ADME/TOX) parameters. The primary reason for failure of drugs late in the drug discovery
process is due to poor ADME/TOX at which point there has already been a substantial financial investment in its development. It is thus desirable to discover early on in the drug discovery process which compounds have poor ADME/TOX properties. Recently, advances have been made in modeling ADME/TOX characteristics, so that compounds can be eliminated from screening\[24\]. One particularly successful method is VolSurf, which correlates 3D structures with physiochemical properties and pharmacokinetics\[25\]. More recently, this technique has been applied in an integrated framework that predicts both activity and ADME/TOX simultaneously, a strategy that would guide lead optimization to increase efficacy while designing in favorable ADMET/TOX properties as well \[26\]. Making reliable predictive models of ADME/TOX will reduce development time and will avoid investment in leads that would make poor drugs. This will be a major breakthrough that would also facilitate the development of companion animal drugs from leads by developing species-specific models of ADMET/TOX based on known differences in CYT P450 structure\[27\].

CONCLUSION

In the post-genomic era, we are faced with new opportunities based on the wealth of information about drug targets that is available. We are also faced with new challenges that include antibiotic resistance, emerging diseases that require novel treatments and strategies for developing drugs for companion animals that are economically feasible. Computer-aided drug design is a tool that can help us to meet these challenges.

REFERENCES


For Correspondence:
Nitesh Kumar
Email: niteshprof@gmail.com