CHAPTER - 4

Introduction to
CADD
4.1. Drug Design

Drug design is the approach of finding drugs by design, based on their biological targets. Typically a drug target is a key molecule involved in a particular metabolic or signaling pathway that is specific to a disease condition or pathology, or to the infectivity or survival of a microbial pathogen. Some approaches attempt to stop the functioning of the pathway in the diseased state by causing a key molecule to stop functioning. Drugs may be designed that bind to the active region and inhibit the key molecule. However these drugs would also have to be designed in such a way as not to affect any other important molecules that may be similar in appearance to the key molecules. Sequence homologies are often used to identify such risks. Other approaches may be to enhance the normal pathway by promoting specific molecules in the normal pathways that may have been affected in the diseased state. The structure of the drug molecule that can specifically interact with the biomolecules can be modeled using computational tools. These tools can allow a drug molecule to be constructed within the biomolecule using knowledge of its structure and the nature of its active site. Construction of the drug molecule can be made inside out or outside in depending on whether the core or the residual R-groups are chosen first. However many of these approaches are plagued by the practical problems of chemical synthesis. Newer approaches have also suggested the use of drug molecules that are large and protein in nature rather than as small molecules. There have also been suggestions to make these using mRNA.

4.1.1. Rational drug design

Unlike the historical method of drug discovery, by trial-and-error testing of chemical substances on cultured cells or animals, and matching the apparent effects to treatments, rational drug design begins with a knowledge of specific chemical responses in the body or target organism, and tailoring combinations of these to fit a treatment profile. Due to the complexity of the drug design process two terms of interest are still serendipity and bounded rationality. Those challenges are caused by the large chemical space describing potential new
drugs without side-effects. A particular example of rational drug design involves the use of three-dimensional information about biomolecules obtained from such techniques as x-ray crystallography and NMR spectroscopy (Greer et al., 1994). This approach to drug discovery is sometimes referred to as structure-based drug design. The first unequivocal example of the application of structure-based drug design leading to an approved drug is the carbonic anhydrase inhibitor dorzolamide which was approved in 1995. The activity of a drug at its binding site is one part of the design. Another to take into account is the molecule’s drug likeness, which summarizes the necessary physical properties for effective absorption.

4.1.2. Structure based drug design

Drug discovery referred to, as 'rational' did not take flight until the first structures of the targets were solved. In 1897, Ehrlich suggested a theory called the side chain theory wherein he proposed that specific groups on the cells combine with the toxin. Ehrlich coined these side chains as receptors. Structure-based drug design of protein ligands has emerged as a new tool in medicinal chemistry (Klebe, 2000). The central assumption of structure-based drug design is an iterative one as shown in (Figure 4.1) and often proceeds through multiple cycles before an optimized lead goes into clinical trials. The first cycle includes the

![Figure 4.1. Steps involved in structure based drug design through in silico approach.](image-url)
cloning, purification and structure determination of the target protein or nucleic acid by one of three principal methods: X-ray crystallography, NMR or homology modelling. Using computer algorithms, compounds or fragments of compounds from a database are positioned into a selected region of the structure. These compounds are scored and ranked based on their steric and electrostatic interactions with the target site and the best compounds are tested further with biochemical assays. In the second cycle, structure determination of the target in complex with a promising lead from the first cycle, one with at least micromolar inhibition in vitro, reveals sites on the compound that can be optimized to increase potency. Additional cycles include synthesis of the optimized lead, structure determination of the new target: lead complex, and further optimization of the lead compound. After several cycles of the drug design process, the optimized compounds usually show marked improvement in binding, and often, specificity for the target.

4.1.3. Evaluating a protein structure for structure based drug design

Once a target has been identified, it is necessary to obtain accurate structural information. There are three primary methods for structure determination that are useful for drug-design: X-ray crystallography, NMR, and homology modeling. High-resolution crystal structures are the most common desired source of structural information for drug design, particularly for proteins that range in size from a few amino acids to 998kD (Pellecchia et al., 2002). Another advantage of crystallography is that ordered water molecules are visible in the experimental data and are often useful in drug design. A crystal structure should be evaluated for the resolution of the diffracted amplitudes (often simply called resolution); reliability, or R factors; coordinate error; temperature factors; and chemical correctness. Typically, crystal structures determined with data extending above 2.5 Å are acceptable for drug design purposes since they have a high data to parameter ratio, and the placement of residues in the electron density map is unambiguous. The R factor and R free reported for a model are measures for the correlation between the model and experimental data. The R free value
should be below 28 % and ideally below 25 %, and the R factor should be well below 25 % in order to use the structure in drug design. If the only structure available for a particular target does not meet the resolution or R factor results should be judged carefully.

4.2. Computer Aided Drug Design (CADD)

CADD involves all computer-assisted techniques used to discover, design and optimize biologically active compounds with a putative use as drugs. In principle, the drug discovery process involves three pre-clinical stages before clinical trials, namely target selection, lead identification, and clinical candidate selection. Due to rapid advances in structural biology and computer technology, structure-based computer-aided drug design (CADD) using docking techniques, virtual screening and library design, along with target/structure focusing combinatorial chemistry, has become a powerful tool in the multi-step process of drug discovery. As an emerging technology, CADD accelerates drug development by making use of the accumulated information of existing drugs and diseases, combined with interdisciplinary inputs from other fields. This process extensively uses mathematical models and simulation tools based on the evaluation of potential risks from drug safety and the experimental design of new trials (Kapetanovic, 2008). Fast expansion in this area has been made possible by advances in software and hardware computational power and sophistication, identification of molecular targets, and an increasing database of publicly available target protein structures. CADD is being utilized to identify hits (active drug candidates), select leads (most likely candidates for further evaluation), and optimize leads i.e. transform biologically active compounds into suitable drugs by improving their physicochemical, pharmaceutical, ADMET/PK (pharmacokinetic) properties (Pozzan, 2006). The term biological target is frequently used in pharmaceutical research to describe the native protein in the body whose activity is modified by a drug resulting in a desirable therapeutic effect. In this context, the biological target is often referred to as a drug target. The most common drug targets of currently marketed drugs include as below (Overington et al., 2006) (Figure 4.2).
Figure 4.2. Targets of currently marketed small-molecule drugs, subdivided by biochemical class (Hopkins and Groon, 2002).

- Enzymes
- Ligand-gated ion channels
- Voltage-gated ion channels
- G protein-coupled receptors
- Nuclear hormone receptors
- Transporters
- DNA, Integrins, Miscellaneous

4.2.1. Species-specific genes as drug targets

Comparative analysis of the complete genome sequences of bacterial pathogens available in the public databases offers the first insights into drug discovery approaches of the near future (Galperin and Koonin, 1999). An interesting approach to the prediction of potential drug targets designated as the differential genome display has been proposed by Huynen and co-workers (Huynen et al., 1997). This approach relies on the fact that genome of parasitic microorganisms are generally much smaller and code for fewer proteins than the genomes of free-living organisms. The genes that are present in the genome of a parasitic bacterium but absent in a closely related genome of free pathogenicity can be considered as potential drug targets. Exhaustive comparison of *H. influenzae* and *E. coli* gene products identified forty *H. influenzae* genes that have been exclusively found in pathogens and thus constitute potential drug targets.
4.2.2. DNA as drug targets

Nucleic acids are the repository of genetic information. DNA itself has been shown to be the receptor for many drugs used in cancer and other diseases. These work through a variety of mechanisms including chemical modification and cross linking of DNA (cisplatin) or cleavage of the DNA (bleomycin). Much work either by intercalation of a polyaromatic ring system into the double stranded helix (actinomycin D, ethidium) or by binding to the major and minor grooves of DNA (e.g., netropsin) has been reported (Haq, 2002). DNA has been shown to be the target for chemotherapy with efforts to design sequence-specific reagents for gene therapy.

4.2.3. RNA as drug target

Recent advances in the determination of RNA structure and function have led to new opportunities that will have a significant impact on the pharmaceutical industry. RNA, which, among other functions, serves as a messenger between DNA and proteins, was thought to be an entirely flexible molecule without significant structural complexity. However, recent studies have revealed a surprising intricacy in RNA structure. This observation unlocks opportunities for the pharmaceutical industry to target RNA with small molecules. Perhaps more importantly, drugs that bind to RNA might produce effects that cannot be achieved by drugs that bind to proteins (Ecker and Griffey, 1999). Proof of the principle has already been provided by success of several classes of drugs obtained from natural sources that bind to RNA or RNA-protein complexes.

4.2.4. Membranes as drug targets

Membranes are significant structural elements, both in defining the boundaries of a cell as well as providing interior compartments within the cell associated with particular functions. Cell membranes themselves can also act as targets for molecular recognition. An understanding of the structural and dynamic functions of the membranes (e.g., plasma membranes and intercellular membranes) may add to a more rational design of drug molecules with improved permeation.
characteristics or specific membrane effects. Many general anesthetics are believed to work by their physical effects when dissolved in membranes. Several classes of antibiotics like gramicidin A, antifungals like alamethicin and toxins such as mellitin found in bee venoms have direct effects on planar lipid bilayers, causing transmembrane pores.

4.2.5. Proteins as drug targets

Proteins continue to assume significant attention from the pharmaceutical and biotechnology industries as a valuable source of potential drug targets. Proteins provide the critical link between genes and disease, and as such are the key to the understanding of basic biological processes including disease pathology, diagnosis, and treatment. Researchers have discovered many potential therapeutic targets, and there are currently more than 700 products in various phases of development. However, translating the study of proteins into validated drug targets poses substantial challenges. Genome sequences instruct cells on how and when to make proteins. The proteins in turn are the active players in the cell. Proteins form the machinery of cells, allow cells to communicate, and can control growth or death of an organism. Because of their role in cells, most of the drug targets are proteins. Drugs work by binding specifically to a protein. Extensive knowledge about the function of a protein can guide the selection of targets for pharmaceutical chemists. Studying the complex domain of 2,00,000-3,00,000 distinct and interactive proteins poses substantial challenges. Most target proteins for drug development participate in key regulatory steps in the human body or in an infectious organism. As such, they tend to be present in few copies only and often within specific cells. Their isolation and purification using traditional preparative biochemical means and in quantities required for routine assays has been a formidable challenge. This situation has been radically changed by the ability to clone and express proteins. Thus many key target proteins are now becoming available in sufficient amounts to make them amenable not only to biological assays but also to NMR studies in solution and to crystallization for X-ray analysis. The number of protein structures solved using
X-ray or NMR has begun to rise sharply and more than 69,162 protein three-dimensional structures have been deposited in the Protein Data Bank till date (November, 2010) (www.rcsb.org/pdb/) shown in (Figure 4.3). Various classes of proteins can be categorized as potential drug targets. Small molecules such as drugs, insecticides or herbicides usually exert their effects by binding to protein targets. In the past, many of these molecules were found empirically with little or no knowledge of the mechanism of action involved. In many cases, the targets that are modified by these substances were identified in retrospect. Interestingly, the majority of drugs currently in use modulate either enzymes or receptors, most of them G-protein-coupled receptors.

4.2.5. a. Enzymes - The macromolecule responsible for the catalysis of biochemical reactions are an obvious target when a disease state is associated with production of a biologically active species. Enzymes are a classic target for therapeutic intervention and numerous well-studied examples exist. Traditional medicinal chemistry enzyme targets include kinases, phosphodiesterases, proteases and phosphotases. Some of the examples of drug targeted against enzymes are listed in Table 4.1.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dihydrofolate reductase</td>
<td>Methotrexate</td>
</tr>
<tr>
<td>HIV-1 protease</td>
<td>Saquinavir, Indinavir</td>
</tr>
<tr>
<td>Acetyl choline esterase (ACE)</td>
<td>Captopril</td>
</tr>
<tr>
<td>Neuraminidase</td>
<td>Oseltamivir</td>
</tr>
<tr>
<td>Cyclin dependent Kinase (CDKs)</td>
<td>Flavopiridol</td>
</tr>
<tr>
<td>Cyclooxygenase</td>
<td>Diclofenac, Indomethacin</td>
</tr>
<tr>
<td>Thymidylate synthase</td>
<td>Tomudex</td>
</tr>
<tr>
<td>Guanine phosphoribosyltransferase (GPRT)</td>
<td>Allopurinol</td>
</tr>
<tr>
<td>Inosine 5'-monophosphate dehydrogenase</td>
<td>Tiazopurin</td>
</tr>
</tbody>
</table>

Table 4.1. Some enzymes as drug targets.

4.2.5. b. G-proteins - G-protein–coupled receptors are a super family of seven transmembrane spanning proteins that are activated by a wide range of extracellular ligands and are expressed in virtually all tissues. Signaling through receptors regulate a wide variety of physiological processes such as neurotransmission, chemotaxis, inflammation, cell proliferation, cardiac and
smooth muscle contraction as well as visual and chemosensory perception. In view of their widespread distribution and importance in health and disease, it is not surprising that GPCRs are the most successful class of target proteins for drug discovery research (Cacace et al., 2003). The sequencing of human genome has led to the prediction of as many as 1000 GPCRs, of which 400 are nonchemosensory receptors and can therefore be considered as potential drug-targets (Venter et al., 2001). It has been estimated that up to 50% of all marketed drugs directly target this family of receptors (Drews, 2000), some of which are listed in Table 4.2.

<table>
<thead>
<tr>
<th>GPCR</th>
<th>Indication(s)</th>
<th>Drug(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histamine</td>
<td>Allergies, ulcers</td>
<td>Cimetidine, Ranitidine, Terfenadine</td>
</tr>
<tr>
<td>β-adrenergic</td>
<td>Hypertension, asthma</td>
<td>Atenolol, Albuterol, Salmeterol</td>
</tr>
<tr>
<td>α-adrenergic</td>
<td>Benign prostatic hypertrophy</td>
<td>Terazosin, Doxazosin</td>
</tr>
<tr>
<td>Dopamine</td>
<td>Psychosis, Parkinson’s</td>
<td>Aripiprazole, Ropinerole</td>
</tr>
<tr>
<td>Serotonin</td>
<td>Migraine, anxiety</td>
<td>Zolmitriptan, Clozapine, Buspirone</td>
</tr>
<tr>
<td>Opioid</td>
<td>Pain</td>
<td>Butorphanol</td>
</tr>
<tr>
<td>Angiotensin</td>
<td>Hypertension</td>
<td>Losartan, Eprosartan</td>
</tr>
<tr>
<td>Muscarinic acetylcholine</td>
<td>Alzheimer's disease</td>
<td>Bethanechol, Dicyclomine</td>
</tr>
<tr>
<td>Leukotriene</td>
<td>Asthma</td>
<td>Pranlukast</td>
</tr>
</tbody>
</table>

Table 4.2. Some currently marketed drugs that target GPCRs.

The goal in developing drugs against the targets listed is often to modulate the function of the human protein while the goal in developing drugs against pathogenic organisms is total inhibition, leading to the death of the pathogen. Antimicrobial drugs should be essential to the pathogen, have a unique function in the pathogen, be present only in the pathogen, and be able to be inhibited by a small molecule. The target should be essential, in that it is a part of a crucial cycle in the cell, and its elimination should lead to the pathogen’s death. The target should be unique; no other pathway should be able to supplement the function of the target and overcome the presence of the inhibitor. If the macromolecule satisfies all the outlined criteria to be a drug target but functions in healthy human cells as well as in a pathogen, specificity can often be
engineered into the inhibitor by exploiting structural or biochemical differences between the pathogenic and human forms.

![Yearly Growth of Total Structures](image)

Figure 4.3. Growth of crystal structure proteins deposited in Protein data bank.

Finally, the target molecule should be capable of inhibition by binding of a small molecule. Enzymes are often excellent drug targets because compounds are designed to fit within the active site pocket. Structures determined by nuclear magnetic resonance, using a concentrated protein or nucleic acid in solution are also valuable sources for drug design (Pellecchia et al., 2002). Since the target is in solution it is sometimes possible to interpret the dynamics of the target from the data. If no experimentally determined structure is available, a homology
model can be used for drug design (Enyedy et al., 2001). To evaluate a homology model, SWISS MODEL outputs a confidence factor per residue that reflects the amount of structural information used to create that portion of the model. Using the structural information obtained through the above techniques, the structure is then prepared for drug design programs.

4.2.6. Present state of the art: Computer-aided drug design

Given the vast size of organic chemical space drug discovery cannot be reduced to a simple "synthesize and test" drudgery (Kuntz, 1992). There is an urgent need to identify and/ or design drug-like molecules from the vast expanse of what could be synthesized. *In silico* methods have the potential to reduce both time and cost in developing suggestions on drug/ lead-like molecules. Computational tools have the advantage for delivering new lead candidate more quickly and at lower cost. Drug discovery in the 21st century is expected to be different in at least two distinct ways: development of individualized medicine departing from genomic information and extensive use of *in silico* simulations to facilitate target identification, structure prediction and lead/drug discovery. The expectations from computational methods for reliable and expeditious protocols for developing suggestions on potential leads are continuously on the increase. Several conceptual and methodological concerns remain before an automation of drug design *in silico* could be contemplated. Computational methods are needed to exploit the structural information to understand specific molecular recognition events and to elucidate the function of the target macromolecule (Figure 4.4). This information should ultimately lead to the design of small molecule ligands for the target, which will block/activate its normal function and thereby act as improved drugs. As structural genomics, bioinformatics, and computational power continue to explode with new advances, further successes in structure-based drug design are likely to follow. Each year, new targets are being identified; structures of those targets are being determined at an amazing rate, and capability to capture a quantitative picture of the interactions between macromolecules and ligands is accelerating.
4.3. Success of computer-assisted molecular design

The greatest success of computer-aided structure-based drug design to date is the HIV-1 protease inhibitors that have been approved by the United States Food and Drug Administration and reached the market (Wlodawer and Vondrasek, 1998). An example of the success story is that of SAR work carried out on antibacterial agent, Norfloxacin showed 6-fluro derivative of norfloxacin being 500 fold more potent over nalidixic acid. Other examples of drugs that were developed using computer-assisted drug design include Captopril (antihypertensive), Crixican (anti-HIV) (Greer et al., 1994), Teveten (antihypertensive) (Keenan et al., 1993), and Aricept (for Alzheimers disease).

![Diagram](image)

Figure 4.4. Potential areas for in silico intervention in drug discovery.

4.4. Utility of Homology Models in the Drug Discovery Process

Advances in bioinformatics and protein modelling algorithms, in addition to the enormous increase in experimental protein structure information, have aided in generation of databases that comprise homology models of a significant portion of known genomic protein sequences. Currently, 3D structure information can be
generated for up to 56% of all known proteins. However, there is considerable
controversy concerning the real value of homology models for drug design.
Despite the numerous uncertainties that are associated with homology modeling,
recent research has shown that this approach can be used to significant
advantage in the identification and validation of drug targets, as well as for the
identification and optimization of lead compounds. Homology model-based drug
design has been applied to epidermal growth factor receptor tyrosine (Kawakami
et al., 1996); Trusopt (for Glaucoma) and Zomig (for migraine) (Glen et al., 1995).
kinase protein (Ghosh et al., 2001), Bruton's tyrosine kinase (Mahajan et al.,
1999), Janus kinase 3 (Sudbeck et al., 1999) and human aurora 1 and 2 kinases
(Venkayalapati et al., 2003).

Thus, it can be said that pharmaceutical and biotechnology research has
undergone great change. Traditionally, the crucial impasse in the industry's
search for new drug targets was the availability of biological data. Now with the
advent of human genomic sequence, bioinformatics offers several approaches
for the prediction of structure and function of proteins on the basis of sequence
and structural similarities. The protein sequence → structure → function
relationship is well established and reveals that the structural details at atomic
level help understand molecular function of proteins. Impressive technological
advances in areas such as structural characterization of biomacromolecules,
computer sciences and molecular biology have made rational drug design
feasible and present a holistic approach.