CHAPTER 1

General Introduction
1.1. Preamble

The pursuit of health remedies is as old as mankind itself. The present-day medical science has its root in every human civilization. Folklore, archaeological excavations, paleopathological evidences and literary or artistic resources of ancient civilizations suggest that all ancient civilizations like those in Egypt, Mesopotamia, India or China had developed large, varied and fruitful medical traditions of their own. Interestingly enough, many of such traditions such as the trend of using different plant extracts or herbal medicines, using heat to relieve discomfort or applying cold to deaden pain are being continued even in this modern era of medicines. However, like all other subjects of studies, medical practices and drug discovery have also passed through a series of paradigm shifts – starting from the paradigm of the mysticism and rituals in antiquity, through the prevalence of serendipity and empirical observations in the medieval age, the era of the chemical/biochemical/pharmaceutical experimentations in the past millennium, including the concept of rational drug design during the last quarter of the twentieth century, to the realm of the fourth paradigm of data-intensive science of the current age - a brief account of which is given below.

1.1.1. The ancient & medieval ages – the era of mysticism, heuristics and serendipity

All ancient civilizations had their own traditional medicines. Ancient Egyptians had a system of medicine that was very advanced for its time and influenced later medical traditions. The Egyptians and Babylonians also introduced the concepts of diagnosis, prognosis, and medical examination. In the first millennium BCE, in the post-Vedic India emerged the traditional medicine system known as Ayurveda, meaning the "complete knowledge for
long life”. Traditional cults of medicine were also developed in ancient Chinese and Roman civilizations.

However, starting from the Neolithic Age till the medieval days, there was no clear distinction between philosophy and science, especially the science of medicine. Medicinal practices in those days were often eclectic blends of heuristic observations, folk remedies, and religious cult practice. By the fifth century B.C., logical investigations for probable causes of ailments and their remedies began to emerge. Hippocrates, the famous Greek physician known as the "father of medicine," (Grammaticos and Diamantis 2008) was the first to separate medicine from philosophy and to disprove the idea that disease was a punishment for sin. He also established a code of what should be expected of physicians. Even today, the Hippocratic Oath is well-respected by physicians and medical practitioners.

As already mentioned, till the medieval ages, much of the traditional treatment for injuries and ailments stemmed from folk medicine, a practice which used the knowledge of herbs, plant extracts and other earthy substances, collected piece by piece through the ages, to cure everything from toothaches to infertility. The ancient Egyptians were among the first to use certain herbs and drugs as a form of medicine. Since then till the 1800s, prevailed the age of botanicals, where use of herbal medicines was the common practice (Somberg 1996). Usages of such herbals were, in general, stemmed from intuition and empirical observations and most of the successful results emerged as the products of rare “fortunate accidents” (Kaul 1998). By trial and empiricism, people in various cultures accumulated knowledge on diagnostics and medicines. Many folk remedies are placebos, many are effective, and many are still in use. However, it was not until the 19th century that the systematic applications of the scientific methods to medical research
opened up vistas for development of new generation medicines and drugs like antibiotics, aspirin or morphine.

1.1.2. The nineteenth century & beyond - the era of serendipitous observations followed by logical experimentation & analysis

If necessity be the mother of invention, then serendipity has been the naissance of many, if not most, important discoveries of the nineteenth century that had laid the foundation of modern biomedical sciences. Had Van Leeuwenhoek not focussed a magnifying glass on a drop of water instead of a fly's leg - the discovery of bacteria could have been delayed by an indefinite period (Madigan 2006). Had Pasteur not inoculated chickens with an old cholera culture - the discovery of vaccine against cholera might not have taken place in 1879. Had Fleming not made the error of leaving a Petri dish uncovered, the world might have to wait for many more years to witness the onset of the penicillin or antibiotic era (Haven 1994). These are just a few of many "happy accidents" in medicine and drug discovery.

But the major difference between the serendipitous discoveries in antiquity and those made during the last two centuries had been that the later were followed by systematic scientific investigations in an attempt to provide logical explanations for such serendipitous observations - a practice started around the beginning of the 19th century. During this time, medicine was revolutionized by advances in chemistry and laboratory techniques and equipment; old ideas of infectious disease epidemiology were replaced with bacteriology and virology; diseases were being identified by its symptoms and those symptoms were started being treated by different means (Lesney 2000). And thus came a paradigm shift in the science of medicine and therapeutics.
During the second half of the nineteenth century, use of patent medicines especially homeopathy came into play and that in turn gave birth to the era of modern pharmaceutical industry. The trend of extracting biologically-active organic molecules in its pure form from different herbs was introduced instead of using the whole natural source. As already mentioned, quite a few drugs were accidentally discovered during this period, which formed basis of many new modern medicines. Few such examples include anti-malarial agent quinine from cinchona bark (Toovey 2004), digitalis from foxglove plant leaves, potent painkillers cocaine and morphine from opium poppy. A benchmark of drug discovery in this era is the isolation of salicylic acid form willow bark in 1874 which is the precursor of aspirin – a drug which is still used for various therapeutic purposes (Sneader 2000). Antipyretic drug antipyrine is another important drug which was discovered during this period and still in use.

However the modern era of drug discovery started with the discovery of biological activity of some natural substances already being used for treatment and also with the introduction of “microbial theory of disease” which explained mechanism of many infections (Knorre 1978). Advances in synthetic organic chemistry and biochemistry with application of medical microbiology (Pizzi 2000) and development of new vaccines, 1920s and 1930s add another dimension to the drug discovery phase. The most important incident of this era is the accidental discovery of penicillin from *Penicillium notatum* by Alexander Fleming leading to the antibiotic era of 1940s (Haven 1994), as mentioned earlier. Some other landmark innovations of this era include discovery of semi-synthetic antibiotic tetracycline from natural chlortetracycline obtained from *Streptomyces aureofaciens* by Benjamin Duggar (Jukes 1985), anti-tubercular aminoglycoside, streptomycin from *Streptomyces griseus* by Selman Waksman (Comroe 1978). The need to treat infected soldiers in World War II also demanded the development and
production of penicillins and other antibiotics. The importance of different vitamins and effect of their deficiency in human health was also uncovered during this period. Discovery of DNA’s structure by Watson and Crick in 1953 (Watson and Crick 1953) and knowledge about human biology and chemistry led to the development of the biotechnology in 1950s which resulted in sophisticated application of different techniques like X-ray crystallography, NMR spectroscopy and mass spectrometry, electrophoresis, ultracentrifugation, HPLC and other technologies into drug discovery. Finally in 1960s the pharmaceutical decade of the century started (Frey 2000), when people became aware of pills. Also the understanding of causes of different diseases by biologists leads to etiology driven strategies for drug discovery. Selected examples of drug discovered during this period include different oral contraceptives (Goldzieher 1982), tranquilizers (e.g. valium), and most importantly poliomyelitis vaccines (Salk and Sabine) (Pearce 2004).

1.1.3. The last half century of the past millennium - the era of evidence-based medicine and rational drug design

Biomedical science had witnessed another paradigm shift with the development of Molecular Biology and Genetic Engineering during the second half of the twentieth century. Especially, the decade of 1970s has been a benchmark in drug discovery as the war against cancer started during this period, the genetic engineering also emerged by using sophisticated biotechnology instruments. Though the fight against cancer has started in 1940s and 1950s by George Hitchings and Gertrude Elion with the invention of DNA-based antimetabolites but different modified purines with anticancer activity were invented in this time period. Another important anticancer drug taxol was discovered in 1973 from *Taxus brevifolia* (Goodman 2001), which is a perfect example of the combination of natural product isolation and organic
synthesis in the development of new drug molecule. The understanding of DNA replication, transcription and translation led to a much better understanding of viral replication (Dimmock 2007) and also leads to the techniques such of recombinant DNA technology (Campbell 2002) and molecular cloning (Hamer and Thomas 1977) and here started the era of evidence-based medicine that attempts to objectively evaluate the quality of clinical research using the techniques reported by researchers in their publications.

Another important incident that shook the world of research in biomedical sciences was the appearance of AIDS during 1980s. As a consequence, the development of immunology and other means to achieve newer drugs against different diseases become much more important as resistance of old diseases to conventional drugs started to increase. Identification of possible biochemical steps for inhibition on the HIV-virus replication cycle led to a newer approach of drug discovery and different nucleoside, non-nucleoside and peptide analogs were invented as targeted anti-HIV agent (D'Cruz and Uckun 2006). Emergence of drug-resistant virus led to the use of combination drug therapy and searching of new biochemical steps to be targeted by drugs. The polymerase chain reaction (PCR) (Bartlett and Stirling 2003) is another most important discovery of 1980s resulted in major advances in biotechnology which added further impact in drug discovery techniques.

The evolution of medical science has been further accelerated, with greater development of large scale genome sequencing projects. With the advent of the evidence-based medicine and parallel advances in information technology, scientists started to invent new strategic approach of drug discovery during this period. Instead of traditional methods of drug discovery that rely on trial-and-error testing of chemical substances on cultured cells or animals, rational drug design (Freter 1988) began with a hypothesis that modulation of a
specific biological target may have significant therapeutic values and in parallel, the concept of combinatorial chemistry were introduced (Xiang, Sun et al. 1995), where large sets of chemically similar agents were produced and then screened for biological activity.

1.1.4. The - omics era of the third millennium – emergence of the fourth paradigm of data-intensive science

Completion of the human genome sequence in February, 2001 (Lander, Linton et al. 2001; Venter, Adams et al. 2001) heralded inception of the - omics era in the field of biomedical sciences (Kiechle, Zhang et al. 2004). Development of advanced technologies, especially the use of robotics and automation has led to routine generation of huge amount of data in diverse areas of life sciences like Genomics, Transcriptomics, Proteomics, Metabolomics etc (Kiechle and Holland-Staley 2003; Daviss 2005). The vital need to harness this deluge of information for medical diagnostic and therapeutic uses as well as other scientific/technological applications has made the *in silico* analysis an integral part of modern life science and thus emerged and flourished the areas like Bioinformatics and Computational Biology. With rapid developments in distributed high-performance computing, with freely-available high-scale data management tools and with advanced open-source data-analysis tools rapidly adapting to the scales of these data sets, the fourth paradigm of data-intensive science has commenced in.

The onset of this - omics era has led to the birth of several new disciplines like Pharmacogenomics (Nagasubramanian 2004), Interactomics (Kiemer and Cesareni 2007) and Systems Biology (Kitano 2002; Bruggeman and Westerhoff 2007) that have revolutionized the study of drug design and development. All these new disciplines rely on global or holistic studies rather than local or
Chapter 1: General Introduction

reductionist approaches to problem solving. For instance, Systems Biology aims at system-level understanding of biological systems, while interactomics deals with the all networks of interactions between and among proteins and other molecules within a cell. The knowledge of the metabolic pathways and complete genome sequences of various disease-causing microbes as well as of human have facilitated development and application of the method of subtractive genome / pathway analysis (Sakharkar, Sakharkar et al. 2008) that rely on mapping between the human and pathogen interactomes to identify genes and pathways exclusively present in the pathogen. Needless to say, such exclusive genes/pathways, if proved to be essential for the survival of the pathogen, may be tapped for development of novel therapeutics. The present dissertation primarily relied on such newly emerging techniques of identification of novel drug targets in pathogenic microbes.

1.2. Drug target

A drug exhibits its action upon interacting with certain receptor inside living body. The macromolecular structures which undergo a specific interaction leading to the clinical effects after interacting with the drug is commonly called as target for drug action or drug target (Swinney 2004). The drug target or receptor is generally protein though few anticancer drugs exert their effects by acting on DNA (Yang and Wang 1999). Drug interacts in specific sites known as active site of the target having specific 3D shapes, which is complementary to the drug structure. Several chemical interactions ranging from polar-polar H-bonds or non-polar hydrophobic bonds may result in a temporary binding of the drug to the receptor however in some cases irreversible covalent bond may also form between the drug and the receptor (Kenakin 2004).
Studies reveal that targets of all drug known till date fall in one of the following categories: enzymes, substrate/metabolites/proteins, receptors, ion channels, transport proteins, DNA/RNA/ribosome, targets of monoclonal antibodies, various physicochemical mechanisms and unknown mechanism of actions (Hopkins and Groom 2002). After the drug interacts with the target, a series of biochemical reactions occur and ultimately the clinical effects come from interference with signal transduction, receptor signalling and biochemical equilibrium. Moreover many drugs are known to interact with more than one target (Jonker, Visser et al. 2005). A good drug should be potent and specific, thus it should have a strong biological effect on the intended target and minimal effect on other targets to reduce the side-effects. In other way if the drug target has some specific structure and which is not present in other proteins then chances of interaction of drug acting on that particular receptor with other proteins are less. Functionally similar proteins are known to possess similar 3D structures, so drugs acting on one protein would affect the other and vice versa.

With the application of different computational structure-based drug designing approach, the knowledge of drug target in detail is really crucial. Therefore identification of the diseased gene and/or protein responsible for any disease is the most important and first step in drug designing. In a similar fashion, designing of a new drug for any pathogenic disease require identification of gene or protein of the particular pathogen in question.

Traditionally the selection of target protein for drug action has been biology driven. Generally, an interesting enzyme or protein whose function in diseased or normal condition is well known is taken into consideration. The protein is then isolated and cloned and then expressed into a recombinant host to confirm its activity and finally used for testing activity of different compounds for high-throughput screening (HTS). This conventional method referred as
“from-function-to-gene” is quite time consuming, and it only focuses to a small set of well established proteins. The availability of whole genome data of human as well as different other pathogenic organisms offers a huge data to be explored in search of novel drug targets. This lead to the involvement of different computational techniques to screen out new potential targets from the whole genome dataset developing the concept of “from-gene-to-screen” process (Debouck and Metcalf 2000).

After the identification of proper drug target, the detail structure and position of the active site of the target is also necessary. Active site of a drug target (mainly enzyme or other protein) is the portion where the drug will bind and exert its action and is mainly found to be present as a pocket on the enzyme surface lined by amino acid residues that is known to interact with the drug and is known as active site residues. Identification of active site is generally carried out from experimentally available 3D structure of the protein-ligand bound form. However for computationally modelled protein structures the same is assessed by different techniques. It involves analyzing the proteins to find binding pockets, deriving key interaction sites within the binding pocket and comparing the same with related known proteins (Gray, Wodicka et al. 1998).

1.3. Human pathogenic bacteria

Infectious diseases caused by different pathogens such as bacteria, virus, prion or fungus are a major threat to human health. Despite of the emergence of AIDS, hepatitis-B and some other life threatening virus inducing infections, bacterial infections are still reasons of major health-issues, making antibiotics most highly used medicines globally. Major genera that contains most of the human pathogenic bacteria species includes: Bacillus, Chlamydia,
Chapter 1: General Introduction

*Clostridium, Helicobacter, Mycobacterium, Neisseria, Pseudomonas, Salmonella, Staphylococcus, Streptococcus* etc.

A large number of bacteria are found to be present in different parts of human body such as nasopharynx, stomach, large intestine and skin contributing to normal human flora (Savage 1977). However in certain cases some of them become pathogenic which can be termed as conditionally pathogenic bacteria. For example, *Staphylococcus* and *Streptococcus* which are present in nasopharynx of normal populations can potentially cause skin infection when allowed into blood through an open wound. Some opportunistic pathogens such as *Pseudomonas aeruginosa, Burkholderia cenocepacia* or *Mycobacterium avium* might cause diseases in decreased immune conditions (Heise 1982).

Another group of organisms known as obligate intracellular parasites, invariably cause diseases when transmitted into intracellular regions and they are able to grow and reproduce only within cells of human and/or other organisms (Amann, Springer et al. 1997). Few examples of such group are *Chlamydia*, which can cause pneumonia, urinary tract infection and sometime even coronary heart disease, (Belland, Ouellette et al. 2004) and *Rickettsia*, causative agent of typhus disease. Another group of pathogenic bacteria causes infections when present intracellular, though they can also be present in open environment making them as facultative intracellular parasites. Examples include *Mycobacterium, Brucella, Salmonella, Neisseria* etc.

Different bacterial diseases include, bacterial vaginosis: bacterial infection caused by an imbalance of naturally occurring bacterial flora in vagina (Ferris, Nyirjesy et al. 2002), bacterial meningitis: bacterial inflammation of meninges or outer membrane covering of brain and spinal cord, bacterial pneumonia: bacterial infection of lungs, urinary tract infection: bacterial infection of urinary tract, bladder or sometime even kidney, Bacterial
gastroenteritis: caused by pathogenic enteric bacteria and sometime even by normal gut flora, and different bacterial skin infections.

1.3.1. Anti-bacterial drugs

The combat against human pathogens had started with the invention of the first antibiotic penicillin. An antibacterial is a compound that either kills or decreases the growth and reproduction of bacteria (Waksman 1947). The first antibiotic penicillin was found from the fungus Penicillium notatum as a by-product and acted as growth retardant against other organisms (Fleming 1929). Modern antibiotics are either semi-synthetics, modified from natural compounds such as different beta-lactum antibiotics, aminoglycosides etc (von Nussbaum, Brands et al. 2006), or produced by chemical synthesis such as sulphonamides, quinolones etc. Antibacterial drugs can be classified according to their chemical structure, mechanism of action, or spectrum of activity (Table 1.3.1) (Tripathi 2003).
### Table 1.3.1. Classifications of antibacterial drugs

<table>
<thead>
<tr>
<th>Basis of classification</th>
<th>Type</th>
<th>Drugs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemical structure</td>
<td>Sulfonamides &amp; related drugs</td>
<td>Sulfadiazine, Sulfamethoxazole, Sulfacetamide, Dapsone, PAS</td>
</tr>
<tr>
<td></td>
<td>Sulfonamides &amp; related drugs</td>
<td>Sulfacetamide, Sulfamethoxazole, Sulfacetamide, Dapsone, PAS</td>
</tr>
<tr>
<td></td>
<td>Diaminopyrimidines</td>
<td>Trimethoprim, Pyrimethamine</td>
</tr>
<tr>
<td></td>
<td>Quinolones</td>
<td>Nalidixic acid, Norfloxacin, Ciprofloxacin</td>
</tr>
<tr>
<td></td>
<td>β-lactam antibiotics</td>
<td>Penicillins, Cephalosporins, Monobactams, Carbapenems</td>
</tr>
<tr>
<td></td>
<td>Tetracyclines</td>
<td>Oxytetracycline, Doxycycline</td>
</tr>
<tr>
<td></td>
<td>Nitrobenzene derivatives</td>
<td>Chloramphenicol</td>
</tr>
<tr>
<td></td>
<td>Aminoglycosides</td>
<td>Streptomycin, Gentamycin, Neomycin</td>
</tr>
<tr>
<td></td>
<td>Macrolide antibiotics</td>
<td>Erythromycin, Roxithromycin, Azithromycin</td>
</tr>
<tr>
<td></td>
<td>Polypeptide antibiotics</td>
<td>Polymyxin B, Colistin, Bacitracin, Tyrothricin</td>
</tr>
<tr>
<td></td>
<td>Glycopeptides</td>
<td>Vancomycin, Teicoplanin</td>
</tr>
<tr>
<td></td>
<td>Oxazolidinone</td>
<td>Linezolid</td>
</tr>
<tr>
<td></td>
<td>Nitrofurane derivatives</td>
<td>Nitrofurantoin, Furafoxidone</td>
</tr>
<tr>
<td></td>
<td>Nitroimidazoles</td>
<td>Metronidazole, Tinidazole</td>
</tr>
<tr>
<td></td>
<td>Nicotinic acid derivatives</td>
<td>Isoniazid, Pyrazinamide, Ethionamide</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>Rifampin, Clindamycin, Cycloserine, Viomycin, Ethambutol</td>
</tr>
</tbody>
</table>

| Mechanism of action     | Inhibit cell wall synthesis | Penicillins, Cephalosporins, Cycloserine, Vancomycin, Bacitracin |
|                        | Causes leakage from cell wall membrane | Polymyxin B, Colistin, Nystatine |
|                        | Inhibit protein synthesis | Tetracyclines, Chloramphenicol, Erythromycin, Linezolid |
|                        | Cause misreading of mRNA code and affect permeability | Streptomycin, Gentamycin |
|                        | Inhibit DNA gyrase | Ciprofloxacin |
|                        | Interfere with DNA function | Rifampin, Metronidazole |
|                        | Interfere with intermediary metabolism | Sulfonamides, Sulfones, PAS, Trimethoprim, Pyrimethamine, Ethambutol |

| Spectrum of activity    | Narrow spectrum | Penicillin G, Streptomycin, Erythromycin |
|                        | Broad spectrum | Tetracyclines, Chloramphenicol |
1.3.1.1. Antibacterial drug-resistance

The most alarming threat to current antimicrobial therapy is the emergence of antibiotic resistance, i.e., a bacteria is able to survive in exposure of certain antibiotics which were previously lethal to the same. Most resistant bacterial strains were found to acquire new genes from other organisms through horizontal gene transfer by conjugation, transduction, or transformation (Timmis, Gonzalez-Carrero et al. 1986). These genes are mainly responsible for drug resistance as those were primarily evolved by natural selection in other pre-existing antibiotic-resistant bacteria to survive against antibiotics and are mainly found in plasmids to allow for easy transfer into other species. The main reason behind this resistance is the overuse and misuse of antibiotics which is not only helping in evolving resistant-strain but also organisms that are much more competent to survive further insults and may have increased virulence mechanisms, enhanced recombinatorial or gene transfer abilities, and/or improved fitness (Pechere 2001). Also the use of antibiotics in animal food in plenty amounts (Ferber 2002) and in soap and other household items are also found to contribute in the emergence of resistance.

Different studies revealed that the resistance occur mainly by four mechanisms (Tenover 2006): drug inactivation or modification by enzymes produced by the bacteria (deactivation of penicillin G by β-lactamase), alteration of the drug target site (alteration of penicillin-binding-proteins in MRSA), alteration of metabolic pathway where the drug interacts (bacteria using folic acid from hosts instead of PABA for sulfonamides resistant strains) and reduced accumulation of the drug inside the bacterial cell due to decreasing in drug permeability and/or increase in drug efflux by newly acquired gene products (fluoroquinolone resistance) (Li and Nikaido 2009).

To prevent the emergence of resistant strains in this high alarming rate different approaches are now being proposed such as: rational use and
Chapter 1: General Introduction

diminishing the overuse of antibiotics, using narrow-spectrum antibiotics consistently whenever possible, preferring short-course of higher dosed antibiotic to a long course of lower dosed ones, isolation of inpatients or hospital staffs in whom infection or colonization with multi-resistant strain is recognized and controlling prescription of antibiotics (McCusker, Harris et al. 2003). However due to availability of antibiotics as over the counter drugs, it is still highly misused by both hospital and community resulting in failure of different preventive measures. Therefore the only solution to fight against these large numbers of resistant strains is to design new drugs against them.

1.3.1.2. Designing new antibacterials

With the increasing rate of antibiotic resistance, the need of new drug is increasing day by day. The new antibacterial drugs not only demand efficacy against those resistant strains but also reduced side-effects/toxicity of the same. In some cases the drug should be specific to the bacteria such as for long-term therapy to avoid affecting normal human flora and broad-spectrum in other cases to be effective against different organisms. However the main problem in antibacterial designing is to identify new drug targets (Miller 2011). The reason behind this is simply because blocking of previously effective drug targets are no more lethal to the resistant strains as they acquire new genes to survive against those actions.

During identification of drug target, different factors should be taken into consideration (Brown and Wright 2005). The target should be essential for the bacterial survival and if the target is knocked out by blocking with the drug from the corresponding pathway the cell will not be able to function through alternative routes (Kim and Copley 2007). The target should not have any similar human protein as it may lead to the blocking of the host protein by the
drug causing toxicity e.g., trimethoprim a drug being used for UTI acts by inhibiting dihydrofolate reductase and due to the presence of a close human homolog it is responsible in lowering folic acid during therapy and thus reducing the number of platelets (Lawrenson and Logie 2001). However, in case of cholramphenicol which exerts its effect by binding into the large ribosomal subunit of bacteria that do not have any close human homolog, produces its side-effects due to the presence of a functionally conserved area in the ribosomes of mitochondria and bacteria despite of the difference in the structure of the prokaryotic and eukaryotic ribosomes (Yonath 2005). The target should be present in a large group of pathogens for broad-spectrum antibiotics, and should be pathogen specific for narrow-spectrum antibiotics. The most essential pathways are peptidoglycan biosynthesis, cell wall synthesis, transcription and translation in bacteria making enzymes involved in those pathways ideal choice of drug targets. The target protein should have a binding site so that a small molecule can bind into it. The target should ideally be present inside cytoplasm rather than be membrane attached to facilitate easy study of its 3D structure as membrane bound proteins are difficult to crystallize for X-ray spectroscopy (Ozawa 1984). Most of the bacterial drug targets are enzymes such as different transferases, hydrolases or ligase as they pose all the ideal characteristics to be bacterial drug targets.

1.3.1.3. Genomics in bacterial drug target identification

With the advance in sequencing techniques added to development of computer and information technology offers a large number of whole genome data of different bacteria in public domain. This leads to the technique of studying the whole gene dataset and comparing the same with host human genome in search of new bacterial drug targets. The traditional technique of discovering new antibacterials by screening synthetic compound
or natural product extracts by studying inhibition of bacterial cell growth in culture media fail tremendously with the emergence of new multidrug resistant strains. The fact that most of the drugs discovered by this technique affect relatively few cellular processes including transcription, translation, DNA replication and cell wall biosynthesis (Fleischmann, Adams et al. 1995) made it an interesting choice to explore other essential metabolic pathways for identifying new targets for drug action and to fight against multi-resistant species.

Development of system biology contributes to a better understanding of the complexity and flexibility of cellular architectures and the interactions between different cellular processes such as biochemical and chemical reactions, signal transduction and also interaction with different external signals (Schrattenholz and Soskic 2008). Therefore, targeting a cellular function of the bacteria by blocking the major check point lead to a better understanding of the action and side effects of the drug. The availability of metabolic pathway data of different pathogenic bacteria and human in public domain (such as KEGG) (Kanehisa, Goto et al. 2002) made it easy to compare both to find out pathways specific to the bacteria but absent in human. This concept was used successfully to shortlist potential drug targets in *Mycobacterium tuberculosis* (Anishetty, Pulimi et al. 2005).

The most important criteria of a bacterial drug target is the essentiality of the same for the growth, replication, viability, or survival of the organism therefore studying essentiality of genes from a whole genome dataset can produce a list of putative drug targets. The concept behind the fact is a molecule which inhibits the activity of an essential gene product would either kill or inhibit the growth of the bacterium which requires that functional protein. Essentiality studies are traditionally being carried out by different mutation studies such as temperature sensitive mutation generation,
transposon mutagenesis etc but with the rapid development of insertion mutagenesis (Apfel, Takacs et al. 1999) it is now possible to identify all the essential genes form a whole genome dataset rapidly. Publicly available databases containing essential gene information of different organisms such as DEG (Zhang and Lin 2009) offers a huge data to study computationally for searching suitable drug targets. This concept was used by Sakharkar et al. (Sakharkar, Sakharkar et al. 2008) to successfully shortlist putative drug targets of *Pseudomonas aeruginosa* (Sakharkar, Sakharkar et al. 2004) where they have shortlisted all the essential gene products of the pathogen and excluded the proteins having any similarity with human protein to identify a list of putative drug targets.

Identifying drug target from metabolic pathway data only may result in targets which lack essentiality whereas targets searched through essentiality study only do not offer detail knowledge of their function and interaction with other cellular processes. In the current dissertation both these approaches are combined to overcome the addressed issues. These drug targets are then studied in detail in structural level to determine the active site which may further employ structure-based designing to discover new drugs.

### 1.4. Rational drug design

Rational drug design can be defined as the invention of new drug molecule based on the knowledge of its biological target (Freter 1988) frequently involving different computer modelling techniques and thus can also be termed as computer aided drug design (CADD) (Veselovsky and Ivanov 2003). In this omics era, with the availability of large number of genomics and proteomics data, the drug design strategies are not random rather it involve strategic steps involving detail knowledge about the biological target of the
Chapter 1: General Introduction

drug or the diseased gene and/or protein (Johnson, Blower et al. 2001). The whole rational drug design procedure is depicted in Figure 1.4.1 (http://www.bmb.leeds.ac.uk/illingworth/cardio/index.htm).

![Diagram of rational drug design process]

**Figure 1.4.1. Rational drug design**

### 1.4.1. Computer-aided drug design

Although a drug can not be designed entirely by computational techniques, the contribution and thus use of these techniques can no longer be eliminated from a successful drug discovery process and all the major pharmaceutical and biotechnology companies use these techniques for successful interventions of new drug molecules. Apart from these, the theoretical studies also permit calculations of free energies and other relevant molecular
properties such as empirical molecular mechanics, quantum mechanics and, more recently, statistical mechanics (King 1990).

These computational techniques involve uses of different chemoinformatics and bioinformatics approach to discover, study or enhance the efficacy of different drug and active molecules. The basic principle behind all these approaches are based on the determination of the ability of a certain chemical molecule to bind to a certain target and the strength of such bindings (Gund 1974). Application of molecular dynamics study ensures prediction of the interaction of small molecular drug with the biological target and thus the changes in conformation. Different molecular mechanics approaches are being used to study the binding affinity applying different scoring functions which employ methods such as linear regression, artificial neural network (ANN) or other statistical techniques (de Azevedo and Dias 2008). Computational methods are vastly used in the area of hit identification using virtual screening, hit-to-lead optimization using QSAR and other approaches for testing efficacy, binding affinity etc (Rajamani and Good 2007).

CADD can be classified into two major categories based on their underline strategic techniques.

1.4.1.1. **Ligand-based drug design**

Ligand-based drug design, also referred as indirect drug design is based on the knowledge of interactions and structures of molecules already known to interact with the biological target. This method is applied when the detail structure and/or the binding site of the target are unknown but a series of compounds have been identified that exert the activity of interest. This approach is most effective when structurally similar compounds with high activity, with no activity, and with a range of intermediate activities are
available. A pharmacophore, the minimum necessary structural characteristics a molecule must possess in order to bind to the target, can then be identified from a template derived from structures of these compounds (Güner 2000). The pharmacophore can thus be determined as a collection of functional groups in three-dimensional space and is complementary to the geometry of the receptor site (Hansch, McClarin et al. 1985).

While applying this approach, conformational analysis is required to study the flexibility of all compounds and then identifying the lowest energy conformers and superimpose them. The pharmacophore is then generated from these superimposed structures (Marshall 1979). A quantitative structure activity relationship (QSAR) is then calculated by a correlation between calculated properties of molecules and their experimentally determined biological activity for the known active compounds (Roy 2007). This pharmacophore may then be used to design new compounds by introducing different functional groups into desired positions and their QSAR is also predicted to determine the most effective compounds.

1.4.1.2. Structure-based drug design

Structure-based or direct drug design is the most widely used approach in computational drug design area and it involves the detail knowledge of 3D structure of the biological target (Leach 2007). The 3D structure of the target is generally obtained by different experimental methods such as X-ray crystallography or NMR spectroscopy. However molecular modelling techniques such as homology modelling or de novo protein designing are also used based on the structures of related proteins for successful determination of target structure. In addition to the structure of the target, the binding site or active site should also be known and the same can be accomplished by
identifying binding pockets on protein surfaces (Hansch, Klein et al. 1986; Kuntz 1992). Structure-based drug design employs two major strategies as described in Figure 1.4.2.

![Figure 1.4.2. Structure-based drug design](image)

In one method some already known and well established pharmacophore is modified and its interaction with the active site is studied to shortlist a group of active compounds. This approach can also be termed as finding method, where a large number of compounds are screened from different ligand databases in order to find those fitting the binding pocket of the receptor (Greer, Erickson et al. 1994). This principle is the basis for virtual screening.
where a list of potential ligand, most likely to bind to the receptor is identified from large libraries of chemical structures (Schneider 2010).

Another newly developed principle uses the building ligand approach or de novo approach which is sometime also referred as receptor-based approach. The ligand molecule is generated or built up in small fragments (either individual atoms or molecular fragments), matching different portions of the binding pockets and those fragments are then constrained and added to build the whole ligand in a stepwise manner (Schneider and Fechner 2005). This method provide added advantage of designing new classes of compounds that present similar substituents to the target using a template or scaffold, which is chemically distinct from previously characterized leads (Dean, Lloyd et al. 2004).

As already described, the structure-based design methods use the receptor structure and the designing of new ligand is done by using molecular recognition. The main principle behind this lies on the assumption that the stronger the affinity of the receptor-ligand interaction the efficient the ligand will be. Thus one of the most important principle is to predict the binding affinity of a certain ligand to its receptor and using this parameter as a selection criteria while assessing its efficacy (Kahraman, Morris et al. 2007). The receptor-ligand interaction is studied using molecular docking techniques and the interactions are ranked based on different scoring functions. The early method proposed by Böhm (Bohm 1994) to develop a general-purposed empirical scoring function to predict the binding energy resulted in the following master equation:

\[
\Delta G_{\text{bind}} = -RT \ln K_d
\]

\[
K_d = [\text{Receptor}] [\text{Acceptor}] / [\text{Complex}]
\]

\[
\Delta G_{\text{bind}} = \Delta G_{\text{desolvation}} + \Delta G_{\text{motion}} + \Delta G_{\text{configuration}} + \Delta G_{\text{interaction}}
\]
Where desolvation is enthalpic penalty for removing the ligand from solvent, motion is entropy penalty for reducing the degrees of freedom when a ligand binds to its receptor, configuration is conformational strain energy required to put the ligand in its active conformation and interaction is enthalpy gain for resolvating the ligand with its receptor. Generally multiple linear regression is used to fit the experimental data into the master equation and the interaction energy is found to be dependent on change in non polar surface, number of H-bonds etc (Gohlke, Hendlich et al. 2000; Clark, Strizhev et al. 2002).

1.5. Homology modelling

Homology modelling is the most widely used protein modelling technique to predict a structure from its primary amino acid sequence with an accuracy that is comparable to the best results achieved experimentally. This in silico method helps to generate protein model in absence of experimentally available structure or when the experimental methods fail such as when the protein is too large for NMR analysis or cannot be crystallized for X-ray diffraction. The rapid progression in sequencing lead to formation of a huge amount of primary sequence data, however with lack in the availability of the experimental protein structure, the usefulness of homology modelling technique increases in the process of structure-based drug design, hypotheses about the location of ligand-binding sites, substrate specificity (Klinkert, Cioli et al. 1994) etc.

The process is based on two major observations: firstly, the structure of a protein is uniquely determined by its amino acid sequence (Epstain 1963), therefore in theory, knowledge of protein sequence should suffice to obtain its structure and secondly, during evolution, the protein structure is more stable and changes at much slower rate than its primary sequence, so similar
sequences adopt practically identical structures, and distantly related sequences still fold into similar pattern (Chothia and Lesk 1986; Sander and Schneider 1991). Thus the technique involves taking a known sequence with an unknown structure (target protein) and deriving its structure by mapping it against a known structure of one or several similar homologous proteins (template).

The concept of homology modelling started as early as 1960s and 1970s when the models were constructed using wire or plastic, however the use of computer graphics came to play much later. The first published report on homology modelling was done by Browne et al. (Browne, North et al. 1969) where they have modelled protein structure by mutating the side-chains not identical in the template protein but the existing coordinate were let remained the same. Different techniques were being introduced to model the backbone structure (McLachlan and Shotton 1971), finally the concept of modelling the whole protein structure was introduced by Greer (Greer 1980; Greer 1981).

1.6. Molecular docking

Molecular docking can be defined as the prediction of the structure of the intermolecular complexes formed by two or more constituent molecules. Molecular docking techniques are widely used in recent drug designing area for virtual screening, lead optimization, binding-site identification, protein-protein interaction studies, understanding enzymatic reaction mechanism etc. The main concept behind studying intermolecular complexes lies on widely accepted theory of the drug action according to which the activity is obtained through the molecular binding of one molecule (ligand) into the pocket of another molecule (receptor) partially or totally. In their binding
conformations, the molecules exhibit geometric and chemical complimentary structures and exert its effect (Wodak and Janin 1978).

The concept of molecular docking started as early as 1970s, however the first use of computer for a study of interaction between haemoglobin and sickle cell fibres was reported in 1975 (Levinthal, Wodak et al. 1975), where simple model of interacting proteins were produced by computers and calculations of electrostatic interactions, hydrogen-bonds were carried out manually. With the advances in bioinformatics, a general algorithm for studying different protein-ligand interactions using correlation method was introduced in 1992 (Katchalski-Katzir, Shariv et al. 1992), however the first successful docking study was performed in 1996 to predict the complexed structure of TEM-1 Beta-lactamase with Beta-lactamase inhibitor protein (Strynadka, Eisenstein et al. 1996).

A docking protocol can be depicted as the combination of two components: a search algorithm and a scoring function. The search algorithm elucidates all possible binding conformation between the ligand and the receptor (Welch, Ruppert et al. 1996). Currently available popularly used search algorithms include Monte Carlo methods, genetic algorithms, fragment-based methods, point complementarity methods, distance geometry methods, tabu searches and systematic searches (Taylor, Jewsbury et al. 2002). The scoring function is used to assess different binding conformation and rank them to predict the best possible interaction complex. Scoring functions can be based on molecular mechanics such as AMBER (Cornell 1995), OPLS (Jorgensen 1988) or CHARMM (Brooks 1983), empirical free energy (Eldridge, Murray et al. 1997) or knowledge based functions (Muegge and Martin 1999).
1.7. Aims and scope of the present study

The ever increasing incidents of antibiotic-resistance can only be defeated by developing new drugs against resistant strains. To design and develop new antibacterial drugs, the first requirement is to successful identification of new targets for drugs as using the old targets might be ineffective in resistant strains. The availability of the full genome and proteome information and also information of different metabolic pathways of a large number of pathogenic bacteria as well as human presents an attractive choice to employ different computational techniques to study whole genome of bacteria and comparing the same with host human and identification of a set of possible drug targets. Further investigation of three dimensional structures of these drug targets helps in detail understanding of its binding cavity and protein-ligand interaction studies help in identification of active amino acid residues responsible for the interaction. Drugs against these targets can then be designed which will interact with those active amino acid residues.

In the present dissertation an endeavour has been made to identify a list of putative drug targets in two different human pathogens namely *Helicobacter pylori* HPAG1 and *Neisseria meningitidis* MC58 using a systematic analysis of the whole bacterial genome and proteome employing subtractive metabolic pathway analysis, gene essentiality study, subtractive genomic analysis and homology searching of bacterial protein with the host human proteins.

In order to study the identified drug targets in detail, attempt has also been made to build their three dimensional structures by homology modelling approach. The ligand-binding cavities have been identified in the energy minimized protein structures. In another attempt to study the interactions between the targets and their biological substrates, molecular docking
strategies have been applied and active amino acid residues of the identified targets have also been identified. The reliability of obtained data has been verified using reference protein structures from other organisms and with different experimental data. The potential targets identified and structurally characterized in the dissertation may facilitate development of novel intervention strategies not only against *H. pylori* HPAG1 and *N. meningitidis* MC58, but also against related virulent microbes, if the enzymes harbouring these targets be also present in those microbes, be essential for their survival, participate in the pathogen-exclusive pathways and have no human homologs.