BIOINFORMATICS : A New Era of Drug Design

Madhu Yadav[1] , Gurmit Singh[2]

[1]*Department of Computational Biology & Bioinformatics,* [2]*Department of Computer Science , Sam Higginbottom Institute of Agriculture, Technology & Sciences, (Formerly Allahabad Agricultural Institute), Allahabad –* 211007*, Uttar Pradesh, India.*
*madhuyadav*2003*@gmail.com.*

Abstract

The Human Genome is fundamentally about information, and computers were essential both for the determination of the sequence and for the sequence and for the applications to biology and medicine that are already flowing from it. For the researchers focused on developing bioinformatics methods, use computer programs to make inferences from the archives of modern molecular biology, to make connection among them, and to derive useful and interesting prediction. Bioinformatics technology is used to solve complex biological questions related to metabolic pathways, genes, protein function and pharmacological/developmental aspects of drugs and medicines.Companies invest millions of money and decades of time to develop a new drug. Bioinformatics helps to accelerate this process and make the drug more efficient and specific at the same time. It has significant advantages over traditionally expensive and time consuming "wet lab" research methods, because computational tools give the most predictive and accurate information about genes and proteins with regards to mediating aspects of drug action.

Keywords: Human Genome, Drug Designing, Homology modeling,Cheminformatics,SBDD.

1. Introduction

On 26-June,2000, the sciences of biology and medicine changed forever,announce the completion of the draft of Human Genome,carried out great change in technology.The new way are open for the researchers.The Human Genome is fundamentally about information, and computers were essential both for the determination of the sequence and for the applications to biology and medicine that are already flowing from it. Computing contributed not only the raw capacity for processing and storage of data, but mathematically sophisticated methods required to achieve the results.*The marriage of biology and computer science has created a new field called BIO-Informatics.* We use computer programs to make inferences from the achieves of modern molecular biology, to make connection among them, and to derive useful and interesting prediction(Lesk, 2002).

Bioinformatics plays an important role in the design of new drug compounds. Rational Drug Design(RDD) is a process used in the biopharmaceutical industry to discover and develop new drug compounds (Flower, 2002). RDD uses a variety of computational methods to identify novel compounds, design compounds for selectivity, efficacy and safety, and develop compounds into clinical trials candidate. The word drug which derives from the Middle English word "drogge", first appears in the English language during the 14th centuary,the dictionary definition of a drug is: a substance used medicinally or in preparation of a medicine, a substance described by an official formulary or pharmacopoeia. A substance used in the diagnosis, treatment, mitigation, cure or other prevention of disease.

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 signalling 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 this 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 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 proteinaceous in nature rather than as small molecules. There have also been suggestions to make these using mRNA. Gene silencing may also have therapeutical applications.

2. Different Fields Of Drug Design

To enhance the design and development of new drugs, the field has diversified into different sections as follows:

2.1 Proteomics:

A field involved in studying different proteomes and their protein expression in cells, which assists in identifying disease mechanisms for therapeutic drug targets.It is defined as the qualitative and quantative comparison of proteomes or the study of proteins, particularly their structures and functions (Dale, 2002).

2.2 Pharmacogenomics:

A study of individual's genetic inheritance which is responsible for his/her reaction to different drugs. So as to prescribe the effective and least toxic drugs and decrease overall medical costs. It 's aims to develop rational means to optimize drug therapy with respect to the patients genotype, to ensure maximum efficacy with minimal adverse effect. Such approaches promise the advent of personalize medicine (Dale, 2002).

2.3 Computer based drug design:

Design of drug molecules is done on the basis of the structure of drug receptors, structure activity relationship, toxicity assessment, physical/chemical properties of drugs. The formulation aspects of polymorphism and compatibility with different formulation additives are computed on specially designed software.

2.4 Cheminformatics:

Cheminformatics (also known as chemoinformatics and chemical informatics) is the use of computer and informational techniques, applied to a range of problems in the field of chemistry. These in silico techniques are used in pharmaceutical companies in the process of drug discovery. It's combines the scientific working fields of chemistry and computer science for example in the area of topology and chemical graph theory and mining the chemical space.A newly emergent disciplines that combine the decades old discipline of chemical information management, which includes substructure searching e.g. with areas of molecular modelling, such as QSAR.

2.5 Lead- discovery:

Lead compounds have possessed key properties, such as activity at a particular receptor or enzyme, but are deficient in others, such as selectivity,metabolic stability or their pharmacokinetics profile. New leads have arisen either as a result of serendipity or from analogy to known compounds. These may be natural ligands enzymes substrates or antagonists.

Together these fields aid in the development of new drug molecules, which reduces cost of conducting clinical trials, which often fail due to variations in animal or human models.

3. Methods Use For Drug Design:

The following methods of drug design fall into several natural categories structure-based drug design, ligand-based drug design, de novo design, homology modeling etc, depending on how much such information is available about drug targets and potential drug compounds. In this article we shall focus on methods of drug design and describe a few salient features.

3.1 Structure-Based Drug Design(SBDD):

Structure-based drug design (Bailey & Brown, 2001) is one of several methods in the rational drug design toolbox. Drug targets are typically key molecules involved in a specific metabolic or cell signaling pathway that is known, or believed, to be related to a particular disease state. Drug targets are most often proteins and enzymes in these pathways. Drug compounds are designed to inhibit, restore or otherwise modify the structure and behavior of disease-related proteins and enzymes. SBDD uses the known 3D geometrical shape or structure of proteins to assist in the development of new drug compounds. The 3D structure of protein targets is most often derived from following methods:

3.1.1 Crystallography:

In X-ray crystallography, a crystal of a substance is placed in an X-ray beam. The electron clouds surrounding the atoms in the crystal reflect X-ray.In a protein crystal, individual protein molecules are arranged in a regular lattice, so the crystal in regular patterns reflects X-rays.The X-ray reflections scattered from a protein crystal can be analyzed to produce an electron density map of the protein.Modeling the best possible way for the atom making up the known sequence of the protein to fit into this electron density produces protein atomic coordinates.

3.1.2 NMR:

An increasing number of protein structure are being solved by nuclear magnetic resonance (NMR) spectroscopy. NMR detects atomic nuclei with non zero spin and produced by these nuclei are shifted in the magnetic field depending on their electronic environment. By interpreting the chemical shifts observed in the NMR spectrum of a molecule,distance between particular atoms in the molecule can be estimated.

3.1.3 Homology modeling :

Homology modeling is one method used to predict 3-D structure. In homology modeling, the amino acid sequence of a specific protein (target) is known, and the 3-D structures of proteins related to the target (templates) are known.Bioinformatics software tools are then used to predict the 3-D structure of the target based on the known 3-D structures of the templates.MODELLER is a well-known tool in homology modeling, and the SWISS-MODEL Repository is a database of protein structures created with homology modeling. Similarity Searches. A common activity in biopharmaceutical companies is the search for drug analogues. Starting with a promising drug molecule, one can search for chemical compounds with similar structure or properties to a known compound. There are a variety of methods used in these searches, including sequence similarity, 2D and 3D shape similarity, substructure similarity, electrostatic similarity and others.

3.1.4 Docking-ligands:

One of the key benefits of SBDD methods is the exceptional capability it provides for docking putative drug compounds (ligands) in the active site of target proteins. Most proteins contain pockets, cavities, surface depressions and other geometrical regions where small-molecule compounds can easily bind. With high-resolution x-ray and NMR structures for proteins and ligands, researchers can show precisely how ligands orient themselves in protein active sites. Open source bioinformatic tools such as VMD and NAMD. Furthermore, it is well known that proteins are often flexible molecules that adjust their shape to accommodate bound ligands. In a process called molecular dynamics, SBDD allows researchers to dock ligands into protein active sites and then visualize how much movement occurs in amino acid sidechains during the docking process. In some cases, there is almost no movement at all (i.e., rigid-body docking); in other cases, there is substantial movement. Flexible docking can have profound implications for designing small-molecule ligands so this is an important feature in SBDD methods.

3.2 Ligand-Based Drug Design:

Many receptors are not readily amenable to receptor-based drug design. For example, many important receptors are membrane-bound proteins, which are notoriously difficult to crystallize. In such cases, a lead compound or active ligand must be found, and then the structure of the ligand guides the drug design process (Bo¨hm, 1996). There are following few methods of ligand-based drug design:

3.2.1 Pharmacophore Based Design:

The pharmacophore is an important concept in rational drug design. It represents a set of functional features in 3D geometry that can interact with a specific receptor to gain activity (Hansch & Fujita, 1964). Pharmacophore is the geometrical depiction of fragements or features believed to be responsible for the biological activity. In general, it is atom based and defined as chemical properties such as acid,base,hydrophobe or aromatic group when an unknown structure of a receptor or a membrane associated protein is the target. The pharmacophore model becomes a useful tool for searching lead compound. Pharmacophore based virtual screening has been demonstrated as an efficient method and can reach hit rates of 1-20%. The generation of pharmacophore can be achieved by DISOtech and RECEPTOR & CATAYST software.

3.2.2 Quantative Structure Activity Relationship:

When ligands or inhibitors are available, QSAR was the first method developed for the virtual screeningand has been advanced from a two dimensional(2D) to 3D model. QSAR analysis uses statistical methods to study the correlation of biological activity, physiochemical properties, or toxicities to structural properties of candidate molecules. In the classical or 2D-QSAR techniques, traditional SAR methods such as the Hansch equation. Modern 3D-QSAR methods analyze the interaction fields around a key molecule calculating the interaction energy in a grid. The most well known 3D QSAR technique is Comparative Molecular Field Analysis (CoMFA). The CoMFA approach calculates molecular properties such as steric, electronic, hydrogen bonding and hydrophobic fields (Hansch & Fujita, 1964).

3.3 De-novo design:

Computational chemist, believed they could circumvent much of the time and effort required for drug synthesis and testing by simply generating novel compounds using the computer. Testing would be replaced by merely calculating the ligand receptor

binding affinity using the physical laws of chemistry. The concepts of generating virtual lead compounds entirely through computer simulation was termed de-novo design (Bo¨hm, 1992). With known 3D structure targets (enzyme or protein) new inhibitors or ligands can be designed de-novo. De-novo design programs try to construct novel structures, using set of predefined fragements,into an active site or onto a pharmacophore model. The computational nature of de-novo design leads to a large amount of chemical structures. ERID is the first de-novo method another LEGEND, leap frog, ligBuilder, CONCERTS were developed to build up structures according to the active site of the receptors. De-novo design generates functional groups in suitable form and suitable conformation to interact with the protein surface.

3.4 CADD (Computer Aided Drug Design):

It is a specialized discipline that uses computational methods to simulate drug-receptor interactions.CADD methods are heavily dependent on bioinformatics tools, applications and databases. As such, there is considerable overlap in CADD research and bioinformatics (Bailey & Brown, 2001).

3.5 Virtual High-Throughput Screening (vHTS):

Pharmaceutical companies are always searching for new leads to develop into drug compounds. One search method is virtual high-throughput screening (Gordon, 1994). In vHTS, protein targets are screened against databases of small-molecule compounds to see which molecules bind strongly to the target. If there is a hit with a particular compound, it can be extracted from the database for further testing. With today's computational resources, several million compounds can be screened in a few days on sufficiently large clustered computers. ZINC is a good example of a vHTS compound library. In the era of combinatorial chemistry and HTS the analysis of virtual chemical structures have assumed a position of central importance within computational chemistry, impinging directly on the design of combinatorial chemistry.

3.6 Cheminformatics – Library design:

Parallel synthesis and combinatorial chemistry allow the medicinal chemist to supplement the synthesis of individual compound with the use of the compound library as a tool in drug discovery. These concepts are traced back at least as far as Hank's work in the 1960.The capacity of combinatorial chemistry to generate large number of compound can either be directed towards the generation of a universal library or towards the construction of smaller,more focused libraries of more similar structures. Design of the first of these type is well met by method from the emergent discipline of Cheminformatics (Gordon, 1994). The second type requires knowledge of structural requirement for the activity at a particular kind of receptor; knowledge, which came from SAR or from understanding of the receptor structure itself .

3.7 Lead Optimization :

After a number of lead compounds have been found, SBDD techniques are especially effective in refining their 3D structures to improve binding to protein active sites, a process known as lead optimization. In lead optimization researchers systematically modify the structure of the lead compound, docking each specific configuration of a drug compound in a proteins active site, and then testing how well each configuration binds to the site. In a common lead optimization method known as bioisosteric replacement, specific functional groups in a ligand are substituted for other groups to improve the binding characteristics of the ligand. With SBDD researchers can examine the various bioisosteres and their docking configurations, choosing only those that bind well in the active site.

3.8 Drug Lead Optimization:

When a promising lead candidate has been found in a drug discovery program, the next step (a very long and expensive step!) is to optimize the structure and properties of the potential drug. This usually involves a series of modifications to the primary structure (scaffold) and secondary structure (moieties) of the compound. This process can be enhanced using software tools that explore related compounds (bioisosteres) to the lead candidate.

3.9 Physiochemical Modeling :

Drug-receptor interactions occur on atomic scales. To form a deep understanding of how and why drug compounds bind to protein targets, we must consider the biochemical and biophysical properties of both the drug itself and its target at an atomic level. Swiss-PDB is an excellent tool for doing this. Swiss-PDB can predict key physicochemical properties, such as hydrophobicity and polarity that have a profound influence on how drugs bind to proteins.

3.10 Drug Bioavailability and Bioactivity :

Most drug candidates fail in Phase III clinical trials after many years of research and millions of dollars have been spent on them. And

most fail because of toxicity or problems with metabolism. The key characteristics for drugs are Absorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) and efficacy in other words bioavailability and bioactivity. Although these properties are usually measured in the lab, they can also be predicted in advance with bioinformatics software.

4. Drug Design Based on Bioinformatics Tools:

The processes of designing a new drug using bioinformatics tools have open a new area of research. However, computational techniques assist one in searching drug target and in designing drug in silico, but it takes long time and money. In order to design a new drug one need to follow the following path.

4.1 Identify Target Disease:

One needs to know all about the disease and existing or traditional remedies. It is also important tolook at very similar afflictions and their known treatments. Target identification alone is not sufficient in order to achieve a successful treatment of a disease. A real drug needs to be developed.This drug must influence the target protein in such a way that it does not interfere with normal metabolism. One way to achieve this is to block activity of the protein with a small molecule. Bioinformatics methods have been developed to virtually screen the target for compounds that bind and inhibit the protein. Another possibility is to find other proteins that regulate the activity of the target by binding and forming a complex.

4.2 Study Interesting Compounds:

One needs to identify and study the lead compounds that have some activity against a disease. These may be only marginally useful and may have severe side effects. These compounds provide a starting point for refinement of the chemical structures.

4.3 Detect the Molecular Bases for Disease:

If it is known that a drug must bind to a particular spot on a particular protein or nucleotide then a drug can be tailor made to bind at that site. This is often modeled computationally using any of several different techniques. Traditionally, the primary way of determining what compounds would be tested computationally was provided by the researchers\'understanding of molecular interactions. A second method is the brute force testing of large numbers of compounds from a database of available structures.

4.4 Rational drug design techniques:

These techniques attempt to reproduce the researchers, understanding of how to choose likely compounds built into a software package that is capable of modeling a very large number of compounds in an automated way. Many differentalgorithms have been used for this type of testing, many of which were adapted from artificial intelligence applications. The complexity of biological systems makes it very difficult to determine the structures of large biomolecules. Ideally experimentally determined (x-ray or NMR) structure is desired, but biomolecules are very difficult to crystallize.

4.5 Refinement of compounds:

Once you got a number of lead compounds have been found, computational and laboratory techniques have been very successful in refining the molecular structures to give a greater drug activity and fewer side effects. This is done both in the laboratory and computationally by examining the molecular structures to determine which aspects are responsible for both the drug activity and the side effects.

4.6 Quantitative Structure Activity Relationships (QSAR):

This computational technique should be used to detect the functional group in your compound in order to refine your drug. This can be done using QSAR that consists of computing every possible number that can describe a molecule then doing an enormous curve fit to find out which aspects of the molecule correlate well with the drug activity or side effect severity. This information can then be used to suggest new chemical modifications for synthesis and testing.

4.7 Solubility of Molecule:

One need to check whether the target molecule is water soluble or readily soluble in fatty tissue will affect what part of the body it becomes concentrated in. The ability to get a drug to the correct part of the body is an important factor in its potency. Ideally there is a continual exchange of information between the researchers doing QSAR studies, synthesis and testing. These techniques are frequently used and often very successful since they do not rely on knowing the biological basis of the disease which can be very difficult to determine.

5. Benefit of CADD:

CADD methods and bioinformatics tools offer significant benefits for drug discovery programs.

5.1 Cost Savings

The Tufts Report suggests that the cost of drug discovery and development has reached $800 million for each drug successfully brought to market. Many biopharmaceutical companies now use computational methods and bioinformatics tools to reduce this cost burden. Virtual screening, lead optimization and predictions of bioavailability and bioactivity can help guide experimental research. Only the most promising experimental lines of inquiry can be followed and experimental dead-ends can be avoided early based on the results of CADD simulations.

5.2 Time-To-Market

The predictive power of CADD can help drug research programs choose only the most promising drug candidates. By focusing drug research on specific lead candidates and avoiding potential 'dead-end' compounds, biopharmaceutical companies can get drugs to market more quickly.

5.3 Insight

One of the non-quantifiable benefits of CADD and the use of bioinformatics tools is the deep insight that researchers acquire about drugreceptor interactions. Molecular models of drug compounds can reveal intricate, atomic scale binding properties that are difficult to envision in any other way. When we show researchers new molecular models of their putative drug compounds, their protein targets and how the two bind together, they often come up with new ideas on how to modify the drug compounds for improved fit. This is an intangible enefit that can help design research programs.

CADD and bioinformatics together are a powerful combination in drug research and development. An important challenge for us going forward is finding skilled, experienced people to manage all the bioinformatics tools available to us, which will be a topic for a future article.

6. References

1• Arthur M.lesk.2002 Introduction to Bioinformatics. VII-X:263-266

2• Bailey, D., and Brown, D. (2001). High-throughput chemistry and structure-based design: Survival of the smartest. Drug Disc Today 6, 57–59.

3• Bo¨hm, H. J. (1992a). The computer program LUDI: A new method for the de novo design of enzyme inhibitors. J Comput Aided Mol Des 6, 61–78.

4• Bo¨hm, H. J. (1996). Computational tools for structure-based ligand design. Prog Biophys Mol Biol 66, 197–210.

5• Cynthia Gibas, Per Jamberk. 2002 Developing Bioinformatics computer skills.31-37.

6• Dan E.Krane,Michael L. Raymer.2003 Fundamental Concepts of Bioinformatics.

7• Darren R Flower.2002 Drug design cutting edge approaches. 1-47.

8• Gordon, E. M., Barrett, R. W., Dower, W. J., Fodor, S. P. A., and Gallop, M. (1994). Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions. J Med Chem 37, 1385–1401.

9• Hansch, C., and Fujita, T. (1964). r–s–p analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 86, 1616–1626.

10• Jeremy W Dale, Malcolm von Schantz.2002 From Gene to Genomes, Concepts and Application of DNA Technology.304-342.