Review

APPLICATION OF BIOINFORMATICS IN DRUG DISCOVERY

Probodh Borah
Department of Animal Biotechnology
College of Veterinary Science, Assam Agricultural University
Khanapara, Guwahati-781022

ABSTRACT

Tools of bioinformatics can supplement and replace the traditional drug discovery process. The traditional process involve extraction, isolation and characterization of a compound or synthesize one in laboratory, followed by a series of pharmacological test, animal model and then trial on human volunteers. All these take 10-15 years, huge sum of money and skilled man power. Tools of bioinformatics, particularly Computer Aided Drug Designing (CADD) can reduce the cost by 33% and the time period by 30%. Molecular models of drug compound and its interaction with target molecule can reveal intricate, atomic scale binding properties which can not be deciphered by other means.

Key words : Drug designing, bioinformatics, CADD, translational bioinformatics.

Bioinformatics is an emerging interdisciplinary area of Science and Technology encompassing a systematic development and application of IT solutions to handle biological information by addressing biological data collection and storage, data mining, database searches, analyses and interpretation, modeling and product design. Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be recognized.

Designing, development and commercialization of a drug is a tedious, time-consuming and cost-intensive process (Kuhlman, 1997). Millions of dollars and man-hours are devoted to the discovery of new therapeutic agents. As the activity of a drug is the result of a multitude of factors such as bioavailability, toxicity and metabolism, rational drug design has been a dream for the scientists for centuries.

While the inclusion of computational approaches is common in many areas of molecular biology and related molecular sciences, it is the specific areas of structural biology and genomics which have exploited the techniques most copiously and are the areas in which the extension of bioinformatics into drug development has been the most active (Kingsbury, 1997). Recent technological advances in areas such as structural characterization of bio-macromolecules, computer sciences and molecular biology have made rational drug design feasible (Ooms, 2000).

The processes of designing a new drug using bioinformatics tools have opened a new area of research. As compared to the traditional method of drug discovery where a compound with potential pharmacological activity is isolated, then tested on animals and subsequently in people during clinical trials; using bioinformatics tools it is now easy to start with the compound which specifically targets proteins. Thus the whole process is no longer on a trial and error basis like the traditional approach (Nyachoto, 2005).

In order to design a new drug one needs to follow the following path:
- Identify target disease
- Study interesting compounds
- Understanding the molecular basis of disease
- Rational Drug Design Techniques
- Refinement of compounds
- Quantitative Structure Activity Relationships (QSAR)
- Solubility of Molecule
- Drug Testing

DRUG DESIGN

It is the inventive process of finding new medications based on the knowledge of the biological target (Madsen *et al.*, 2002). Most commonly a drug is an organic small molecule which activates or inhibits the function of a biomolecule such as a protein which in turn results in a therapeutic benefit to the patient. In the most basic sense, drug design involves design of small molecules that are complementary in shape and charge to the bio-molecular target to which they interact and therefore will bind to it.

THE DRUG DISCOVERY PROCESS

The process of discovery of a drug typically involves the following steps:

1. Drug Target Identification:

The identification of new, clinically relevant, molecular targets is of utmost importance to the discovery of innovative drugs. Generally, the 'target' is the naturally existing cellular or molecular structure involved in the pathology of interest that the drug-in-development is meant to act on. To establish a drug 'target' there is a need to have good scientific understanding, supported by relevant publications, of both how the target functions in physiological state and how it is involved in a pathological condition.

A drug target refers to any native protein (or sometimes DNA/RNA) in the body whose activity is modified by a drug resulting in a desirable therapeutic effect. The drug target may be a protein from the pathogen also, in case of infectious diseases. Generally, a drug target is the naturally existing cellular or molecular structure involved in the pathology of interest that the putative drug is meant to act on. It is such a bio-molecule, inhibition of the native function of which results in recovery from the disease. Drug targets are mostly enzymes. G-protein-coupled receptors - GPCRs (23%) and enzymes (50%) represent the most important target classes of proteins for drug discovery. GPCRs are the largest and most diverse group of membrane receptors in eukaryotes. Sometimes they may be DNA or RNA also.

Identification of a drug target in disorders involving a single gene has been a less complicated process in view of the recent advancements in molecular techniques and bioinformatics. However, the identification of genetic loci encoding potential protein targets in disorders involving multiple genes remains an extremely difficult problem both biologically and computationally. Sophisticated mathematical models have been proposed to correlate genotypes with disease and to estimate the quantitative contribution of a particular locus to the disease conditions. This is clearly a critical issue when the selection of a drug target must be made from a number of possibilities. It is known that the situation is very complex and that some loci are associated with multiple disorders depending on the genotypes (Kingsbury, 1997).

2. Target Validation:

It involves demonstrating the relevance of the target protein in a disease process/ pathogenicity and ideally involves studies related to both gain and loss of functions. This is accomplished primarily with the application of knock-out or knock-in animal models, small molecule inhibitors/ agonists/antagonists, antisense nucleic acid constructs, ribozymes and neutralizing antibodies. Since strong interactions between

a protein and its ligand are characterized by a high degree of complementarity, knowledge of the three dimensional structure of the protein enables the prediction of 'druggability' of the protein, i.e. possibility of using it as a drug.

3. Lead Compound Identification:

A lead compound is a representative of a compound series with sufficient potential to progress to a full drug development programme. This biomolecule is measured by its potency, selectivity, pharmacokinetics, physicochemical properties, absence of toxicity and novelty. Compounds are identified which interact with the target protein and modulate its activity. Compounds are mainly identified using random (screening) or rational (design) approaches.

4. Lead Optimization:

Molecules are chemically modified and subsequently characterized in order to obtain compounds with suitable properties to become a drug. Leads are characterized with respect to pharmacodynamic properties such as efficacy and potency *in-vitro* and *in-vivo,* physiochemical properties, pharmacokinetic properties and toxicological aspects.

5. Clinical Trials:

Once a drug has been shown to be effective by an initial assay technique, more testing must be done before it can be given to human patients. Animal testing is the primary type of testing at this stage. Eventually, the compounds, which are deemed suitable at this stage, are sent on to clinical trials. In the clinical trials, additional side effects may be recorded and human dosages are to be determined. The trials are usually completed in four different phases:

● Phase 0 - This is a recent designation for exploratory, first-in-human trials. It is designed to expedite the development of promising therapeutic agents by ascertaining whether the agent behaves in human subjects as was anticipated from preclinical studies.

● Phase I - A small group of healthy volunteers (20-80) are selected to assess the safety, tolerability, pharmacokinetics, and pharmacodynamics of a therapy. It normally includes dose ranging studies so that doses for clinical use can be set or adjusted. The studies may include Single Ascending Dose (SAD) or Multiple Ascending Doses (MAD).

● <u>Phase II</u> - Performed on larger groups (200-300) and are designed to assess the activity of the therapy and continue Phase I safety assessments.

● <u>Phase III</u> - Randomized controlled trials on large patient groups (hundreds to thousands) aimed at being the definitive assessment of the efficacy of the new therapy, in comparison with standard therapy. Side effects are also monitored. It is typically expected that there should be at least two successful phase III clinical trials.

● <u>Phase IV</u> - Post-launch safety monitoring of a drug. It may be mandated or initiated by the pharmaceutical company. It is designed to detect rare or long term adverse effects over a large patient population and timescale than was possible during clinical trials.

COMPUTER-AIDED DRUG DESIGN (CADD)

Drug design frequently but not necessarily relies on computer modeling techniques (Cohen, 1996). Drug design based on computer modeling is commonly referred to as *computer-aided drug design* (CADD). 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. CADD uses bioinformatics tools and computational chemistry to discover, enhance, or study drugs and related biologically active molecules. The most fundamental goal is to predict whether a given molecule will bind to a target and if so how strongly.

In addition to better filtration of targets in early discovery, bioinformatics can also help with three other aspects of target selection (Searls, 2000):

● The characterization of targets, such as the classification and sub-classification of protein families.

● The understanding of targets, such as their behavior in a larger biochemical and/or cellular context.

● The development of targets, such as making predictions about uptake or reuptake, detoxification, the stratification of patient populations and other genetic variations.

Molecular mechanics or molecular dynamics are most often used to predict the conformation of the small molecule and to model conformational changes in the biological target that may occur when the small molecule binds to it. Molecular mechanics methods may also be used to provide semi-quantitative prediction of the binding affinity. Also, knowledge-based scoring function may be used to provide binding affinity estimates. These methods use linear regression, machine learning, neural nets or other statistical techniques to derive predictive binding affinity equations by fitting experimental affinities to computationally derived interaction energies

between the small molecule and the target (Rajamani and Good, 2007; de Azevedo and Dias, 2008).

Drug design with the help of computers may be used at any of the following stages of drug discovery:
1. Hit identification using virtual screening (structure- or ligand-based design)
2. Hit-to-lead optimization of affinity and selectivity (structure-based design, QSAR, etc.)
3. Lead optimization for other pharmaceutical properties while maintaining affinity

Virtual High-Throughput Screening (vHTS) :

Pharmaceutical enterprises are always searching for new leads to develop drug compounds. One search method is virtual high-throughput screening. 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. Pursuing a handful of promising leads for further development can save researchers considerable time and expense, e.g. ZINC is a good example of a vHTS compound library (Casey, 2005).

There are two main approaches in computer aided drug designing– the first is referred to as structure-based drug design (or direct drug design) and the second, ligand-based drug design (or indirect drug design).

In structure-based drug design approach, the target molecule is known and suitable lead molecules are identified which can inhibit the target molecule. It can be compared to a situation where there is a known lock, for which we are to find the appropriate key from a bunch of available keys. Structure-based drug design relies on knowledge of the three dimensional structure of the biological target obtained through methods such as X-ray crystallography or NMR spectroscopy (Reynolds *et al.*, 2010). If the three dimensional structure of a target is not available, it may be possible to create a homology model of the target based on the structure of a related protein. The Structure Based Drug Designing (SBDD) methods are more obvious and direct and have achieved a high level of sophistication.

In Ligand Based Drug Designing (LBDD), the lead molecules are known and probable bioactivity of these molecules are predicted, i.e. we have a key but we have to search for the lock which can be opened with it. In LBDD, the ligand structure is known. The approach is to find out its probable bioactivity. Quantitative Structure Activity Relationship (QSAR) is one such approach. In QSAR, a model is prepared

with some known molecules with similar known bioactivity. A correlation between calculated properties of molecules and their experimentally determined biological activity may be derived. These QSAR relationships in turn may be used to predict the activity of new analogs (Tropsha, 2010).

SEQUENCE ANALYSIS

In CADD research, one often knows the genetic sequence of many organisms or the amino acid sequence of proteins from several species. It is very useful to determine how species vary based an similarly or dissimilarity of gene structure or amino acid sequence in protein. With this information one can infer the evolutionary relationships of the organisms, search for similar sequences in bioinformatics databases and find related species to those under investigation. There are many bioinformatics sequence analysis tools that can be used to determine the level of sequence similarity.

HOMOLOGY MODELING

Another common challenge in CADD research is determining the 3-D structure of proteins. Most drug targets are proteins, so it is important to know their 3-D structure in detail. It is estimated that the human body has 500,000 to 1 million proteins. However, the 3-D structure is known for only a small fraction of these (Casey, 2005). Homology modeling is a method used to predict the 3-D structure of a protein using *in-silico* approach. 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 softwares 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. A variety of bioinformatics tools and search engines are available for this work.

Drug Lead Optimization:

When a promising lead candidate has been found in a drug discovery program, the next 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. OpenEye's WABE is one such tool. Lead optimization tools such as WABE offer a rational approach to drug design that can reduce the time and expense of searching for related compounds (Casey, 2005).

Physicochemical Modeling:

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

Drug Bioavailability and Bioactivity:

Most drug candidates fail in Phase III clinical trials after many years of research and expenditure of millions of dollars. Most of these failures are due to toxicity of the drug or problems related to its metabolism. The key characteristics for drugs related to its efficacy are Absorption, Distribution, Metabolism, Excretion, Toxicity (ADMET) and bioavailability. Although these properties are usually measured in the laboratory, they can also be predicted in advance with bioinformatics software.

ADME/Tox prediction is a tool in CADD for prediction of probable toxicity and other ADME profiles of drug molecules. ADMET screening is a method of predicting these characteristics using bioinformatics tools so that the drug that we are going to design becomes more effective and non-toxic.

BENEFITS OF CADD
Cost Savings:

It is reported that on an average $800 million are spent during the process of drug discovery and development and to successfully release the same in the 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 (Nyachoto, 2005). This contribute to cost reduction significantly.

Time-to-Market:

The predictive power of CADD can help drug research programs to 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.

The application of bioinformatics in drug discovery and development is expected to reduce the annual cost of developing a new drug by 33 percent, and the time taken for drug discovery by 30 percent (Lawrence, 2005).

Insight:

One of the non-quantifiable benefits of CADD and the use of bioinformatics tools is the deep insight that researchers acquire about drug-receptor interactions. Molecular models of drug compounds can reveal intricate, atomic scale binding properties that are difficult to envision in any other way. Knowledge of new molecular model of putative drug compounds, their protein targets and interaction of protein and drug compound, often generate new ideas to modify the drug, leading to further improvement. This is an intangible benefit that can help design research programs.

Translational Bioinformatics:

Translational research in drug discovery and development is broadly defined as the closer integration and use of discovery and preclinical activities with clinical applications (Woolf, 2008). Platform technologies, such as genomics, proteomics and metabolomics have been widely used over the years to understand the inherent biology of normal versus diseased states. Various approaches have been proposed to use transcriptomics studies in understanding disease biology and in drug discovery (Buchan *et al.*, 2011).

Drug repurposing:

Drug repurposing is the process of finding new uses for existing drugs outside the scope of the original medical indication (Ashburn and Thor, 2004). Pharmaceutical companies are increasingly under pressure to justify escalating costs of research in producing therapies (Paul *et al.*, 2010). Taking advantage of existing chemical molecules for new or alternative indications is therefore of great interest. Various translational bioinformatics approaches have been used for repurposing drugs for new indications (Buchan *et al.*, 2011).

Laboratory to market : Case examples:

The first successful case of drug, based on computer aided designing is dorzolamide™ . Computer simulation showed that this compound effectively bind to carbonic anhydrase and inhibit its activity. After a series of testing, trial on human volunteer for efficacy and biosafety; finally this drug was approved for commercial production in 1995. This was followed by approval for commercialization of Imatinib™. This particular compound bind and inactivate tyrosine kinase, a cell surface receptor. This particular drug is specifically designed for the *bcr-abl* fusion protein which is characteristic for Philadelphia chromosome positive leukemia (chronic myelogenous leukemia and occasionally acute lymphocytic leukemia). Imatinib is a new generation drug and superior to excising chemotherapy agents, which generally target rapidly dividing cells. Unlike this Imatinib can differentiate between cancerous cells and normal cells and selectively act on cancerous cells. Subsequently there is a steady stream of such computer designed drug appearing in the market. Notable among then are—Enfuvirtide, a peptide HIV entry inhibitor. Raltegravir, an HIV integrase inhibitor. SSRIs (selective serotonin reuptake inhibitors)— a class of antidepressants, Zanamivir an anti-viral drug. In the coming days more and more such computer designed drugs will appear and bring radical changes to healthcare sector. The most promising aspect is that such novel approach will substantially bring down the cost of medicine to the great relief of common man.

CONCLUSION

Computer-aided drug designing and bioinformatics together are a powerful combination in drug research and development. An important challenge for going forward is finding skilled, experienced people to manage all the bioinformatics tools available.
          CADD approaches aim to increase the speed and efficiency in the drug discovery process. CADD is, however, not a direct route to new drugs, but it provides a

somewhat more detailed road map to the goal. The hope is that providing bit and pieces of information and by helping to coordinate the information, CADD will help to make the drug design process more rational (Ooms, 2000). The application of translational approaches (*e.g.* from bed to bench and back) is gaining momentum in the pharmaceutical industry. By utilizing the rapidly increasing volume of data at all phases of drug discovery, translational bioinformatics is poised to address some of the key challenges faced by the industry (Buchan *et al.*, 2011).

REFERENCES

1. Ashburn T.T. and K.B. Thor (2004) Drug repositioning: identifying and developing new uses for existing drugs. *Nat. Rev. Drug Discov.* 3: 673–683
2. Buchan N.S., D.K. Rajpal, Y. Webster, C. Alatorre, R.C. Gudivada, C. Zheng, P. Sanseau and J. Koehler (2011) The role of translational bioinformatics in drug discovery. *Drug Discovery Today* 16 (9/10): 426-434
3. Casey R.M. (2005) Bioinformatics in Computer-Aided Drug Design. http://swift.cmbi.ru.nl /gv/ students/mtom/how_bioinf_suport_drug.htm (accessed on 24.08.2015)
4. Cohen N. Claude (1996) *Guidebook on Molecular Modeling in Drug Design.* Boston: Academic Press
5. Kingsbury D.T. (1997) Bioinformatics in Drug Discovery. *Drug Development Research* 41:120–128
6. Kuhlman J. (1997) Drug research; from the idea to the product. *International Journal of Clinical Pharmacology and Therapeutics*, 35: 541
7. Lawrence S. (2005) IT Use for Drug Discovery to Rise. www.cioinsight.com/ article2/ 0,1540,1832630,00.asp (Accessed on 24.08.2015)
8. Madsen Ulf; Krogsgaard-Larsen, Povl and Liljefors, Tommy (2002) *Textbook of Drug Design and Discovery.* Washington, DC: Taylor & Francis
9. Nyachoto F.T.D. (2005) Benefits of bioinformatics technology in drug discovery. http:// bioinformaticsanddrugdiscovery.blogspot.in (Accessed on 24.08.2015)
10. Ooms F. (2000) Molecular Modelling and Computer Aided Drug Design. Examples of their Applications in Medicinal Chemistry. *Current Medicinal Chemistry, 7:* 141-158
11. Paul S.M. *et al.* (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* 9: 203–214 (Cited by Buchan *et al.*, 2011)
12. Reynolds C.H., K.M. Merz and D. Ringe, eds. (2010). Drug Design: Structure- and Ligand-Based Approaches (1st ed.). Cambridge, UK: Cambridge University Press. ISBN 978-0521887236
13. Schneider G. and U. Fechner (2005). Computer-based *de novo* design of drug-like molecules. *Nat. Rev. Drug Discov.* 4 (8): 649–63
14. Searls D.B. (2000) Using bioinformatics in gene and drug discovery. *Drug Discovery Today.* 5(4): 135-143
15. Tropsha A. (2010) QSAR in Drug Discovery. *In* Reynolds, C.H.; Merz, K.M. and Ringe, D. *Drug Design: Structure- and Ligand-Based Approaches* (1st ed.). Cambridge, UK: Cambridge University Press. pp. 151–164
16. Woolf S.H. (2008) The meaning of translational research and why it matters. *J. Am. Med. Assoc.* 299: 211–213