Bioinformatics: The effects on the cost of drug discovery

Dibyajyoti Saha, Talha Bin Emran, Swati Paul

Department of Pharmacy, BGC Trust University of Bangladesh, Chittagong, Bangladesh

Correspondence: *Dr. Dibyajyoti Saha (saha.dibyajyoti@gmail.com)*

ABSTRACT

The pharmaceutical industry has been a major respected industry player since becoming organized in the 19th century churning out life saving drugs such as penicillin. Since then many breakthroughs have taken place such as the making of vaccines. So in essence this sector has been associated with saving lives and thus has been held in high esteem. But now with the high cost of drug discovery, rising costs to the consumers, the product recalls being done and adverse side effects, all this topped with more awareness and education on the part of the consumers, the industry image has been battered. Thus it is imperative that the sector finds ways of reducing the costs of drug discovery as well as the time it takes to get the medicine from the laboratory to the patient and at the same time producing drugs which are target specific and with minimal side effects. Bioinformatics is one of the tools the industry has recently engaged to aid in the drug discovery process as well as to cut costs and the time-lines and indeed it was time that the industry caught on with information technology.

Keywords: Bioinformatics, Drug discovery, Pharmaceutical industry, Human Genome Project, Information technology.

Introduction

The birth of Bioinformatics as a result of the explosion of raw data after the completion of the Human Genome Project (HGP) has added another dimension to the drug discovery process. The pharmaceutical industry has all along operated without bringing together the disciplines of biology, chemistry and information technology. These fields, though complimentary had no common interface. The pharmaceutical industry appears to have been left behind when other industries were implementing information technology to improve their operations. But due to the genome project and the resultant data explosion, it was then imperative to join these fields of science together to exploit the available data and thus expedite the drug discovery process. Traditionally, the drug discovery process takes an average of 15 years and costs about $880 million to develop each new medicine that does make it to the market. Nearly 75% of drug candidates currently being tested by pharmaceutical companies will fall short of expectations and never reach the market (1). Added to this is the recent negative perception of the pharmaceutical industry due to the ever spiraling drug prices, recalls and recent warnings about popular prescription medications. In an attempt to improve and reduce the cost of drug discovery, the pharmaceutical industry has recently turned to Bioinformatics. Some analysts predict that Bioinformatics could help cut in half the cost of creating a drug and shave two to three years off its development (1).

History and Definition of Bioinformatics

Bioinformatics started over a century ago when Gregor Mendel, an Austrian monk cross-fertilized different colors of the same species of flowers. Mendel illustrated that the inheritance of traits could be more easily explained if it was controlled by factors passed down from generation to generation. Since Mendel, Bioinformatics and genetic record keeping have come a long way (2).

In 1988, the Human Genome organization (HUGO) was founded. The first complete genome map was published of bacteria *Haemophilus Influenza*. In 1990, the Human Genome Project was started. By 1991, a total of 1879 human genes had been mapped. In France, in 1993, Genethon, a human genome research center produced a physical map of the human genome. Three years later, Genethon published the final version of the human genetic map. This concluded the end of the first phase of the Human Genome Project (2).

Bioinformatics was fuelled by the need to create huge databases, such as Genbank, EMBL and DNA Database of Japan to store and compare the DNA sequence data erupting from the human genome and other genome sequencing projects (2). It enables researchers to analyze the terabytes of data being produced by the Human Genome Project. Gene sequence databases and related analysis tools all help scientists to determine whether and how a particular molecule is directly involved in a disease process. That in turn, helps them find new and better drug targets. Bioinformatics can be thought of as a central hub that unites several disciplines and methodologies as shown below (3). It brings together several activities and this may explain why we get so many definitions for Bioinformatics. The diagram below graphically represents the several methodologies which together make up the discipline of Bioinformatics.
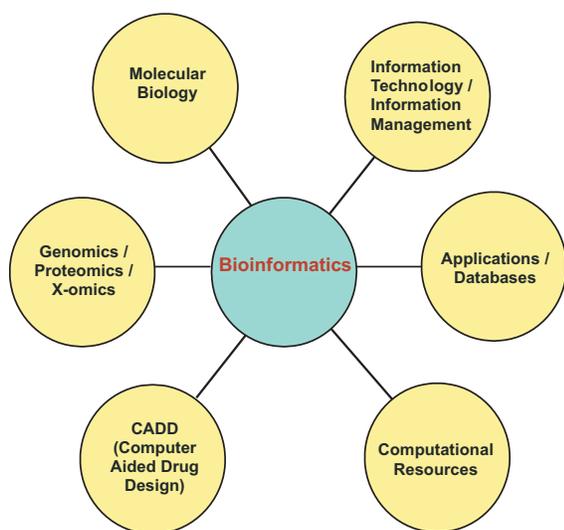
Figure 1: Several methodologies which together make up the discipline of Bioinformatics

Computer-Aided Drug Design (CADD) 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 (4).

Bioinformatics uses computers to store, organize, generate, retrieve, analyze and share sequences, structures, functions, pathways and genetic interactions (1). The definition of Bioinformatics is not universally agreed upon. Generally speaking, it is defined as the creation and development of advanced information and computational technologies for problems in biology, most commonly molecular biology (but increasingly in other areas of biology). As such, it deals with methods for storing, retrieving and analyzing biological data, such as nucleic acid (DNA / RNA) and protein genomic, biological and chemical data to support the drug discovery process (5).

Some people construe Bioinformatics more narrowly, and include only those issues dealing with the management of genome project sequencing data. Others construe Bioinformatics more broadly and include all areas of Computational Biology, including population modeling and numerical simulations (5).

It is interesting to note that there is no one single definition of Bioinformatics. Different organizations define it in their own way. A simpler definition of Bioinformatics is that it is the application of computer technology to the management and analysis of biological data. It is an interdisciplinary research area that is the interface between the biological and computational sciences it's ultimate goal being to uncover the wealth of biological information hidden in the mass of data and to obtain a clearer insight into the fundamental biology of organisms (6). Simply put, it is the marriage between Biology and information technology.

Bioinformatics concerns the development of new tools for the analysis of genomic and molecular biological data including sequence analysis, genetic algorithms, phylogenetic inference, genome database organization and mining, optical computation and holographic memory, pattern recognition and image analysis, biologically inspired computational models (7).

Current drug discovery and development problems

Clinical development of medicines, the process leading up to the regulatory approval for new pharmaceutical products has been identified as the greatest cause of increasing costs in the drug development industry and is predominantly attributed to drugs failing the stringent registration approval process (8).

Estimates by Dimasi *et al*, (9), indicate that clinical trials cost an average of $124 million per drug candidate once drug failure rates have been accounted for. The cost escalates to $802 million once research and development and capitalized losses through time spent out of pocket have been included.

Frost and Sullivan Research indicate that the cost of clinical development has increased by nearly 40% over the last ten years. It also estimates that only 40% of drug candidates entering the clinical development pipeline actually receive regulatory approval and reach the market, indicating that investment is wasted through backing late state facilities (8).

According to the Pharmaceutical Research and Manufacturers of America (PhRMA), in 2004 biotechnology and pharmaceutical companies spent a whooping $38.8 billion in Research and Development. This increase has not been matched by an increase in the number of New Drug Applications (NDA) being submitted to the FDA for approval. Since mid - 1990s NDAs have fallen by almost 50%. This is due to the reduced flow of product pipeline. According to the Tufts Center for the Study of Drug Development, a new prescription drug costs on average $802 million and takes up to 15 years to develop and get FDA approval (10).

The traditional approach to drug development is expensive, time consuming and prone to failure. Pharmaceutical industry observers agree that cutting the time and costs involved in drug discovery and development will be essential to ensuring continued productivity and profitability of drug companies in this industry. Currently, the average drug takes approximately ten years to go from the discovery phase to the clinic and costs the company in question $400 million to more than $1 billion to develop-with much of this cost incurred at the later stages of development. According to experts interviewed for an upcoming Cambridge Healthtech Institute (CHI)

report-*Breaking the Bottlenecks: Applying Genomics Throughout Drug Discovery and Development* (November 2001) - judicious application of genomic technologies can help improve efficiencies throughout the process and also help drugs to "fail fast" (before expensive later-phase trials), thus saving time and money (3).

Despite different figures from different sources, one thing is clear and that is the high costs involved in the clinical development of medicines. These costs are of course then passed onto the consumer. The industry has been vilified in recent terms and is seen as only interested in making huge profits at the expense of the common man on the street. There has been news around the world about governments actually intervening to have the pharmaceutical companies cut down drug prices. The pharmaceutical industry has been claimed to be one of the most profitable industries in both the US and Great Britain. Gross Profit margins of some of the leading pharmaceutical companies in recent years has been around 70 to 80 percent (11). In past years, pharmaceutical prices are said to have risen faster than the rate of inflation and because of little price elasticity associated with price increases, pharmaceutical companies have made high profits. Patients on the other hand will not change the demand for a product with a small change in price when there are no close or available substitutes (12). Such high costs and the negative publicity have finally brought the pharmaceutical industry kicking and screaming into the IT age (1).
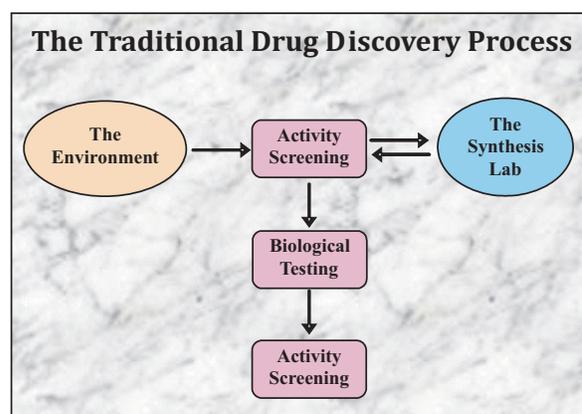
Figure 2: Traditional way of drug discovery used in the Pharmaceutical Industry.

The above diagram is an illustration of the traditional way of drug discovery which has been used in the pharmaceutical industry all along. The process can be divided into several stages. The first stage is the isolation of an active compound with the potential to produce a desired effect. Next studies to determine the pharmacological properties of the drug are determined by testing the product on animals, isolated cell cultures, enzymes and cloned receptor sites and computer models. From these tests, it is determined whether the compound is likely to produce the required benefits and its pharmacological activity. Adverse effects are also observed. A compound which exhibits the most therapeutic potential and minimum side effects then proceeds to the next stage which is the clinical stage.

The clinical stage consists of four phases. Phase I involves the testing of the compound in healthy volunteers, usually men between the ages of 18-30 years. The objective here is to determine its pharmacological effects at given doses and to determine how well it is tolerated in humans and to determine potential toxicological effects. Phase II are controlled trials in which the drug is given to patients with the condition to be treated. Here the effectiveness and safety are observed as well as the optimum dosages. Next Phase III studies which are normally controlled and blinded to reduce bias. Here a larger population of patients, up to 3000, is used and divided into several arms, usually the ones given a placebo, the control group and the ones given the test drug. The test drug may also be compared with the current standard treatment if available to determine effectiveness of the drug. Once a drug proves effective and safe, an application to market the drug is then filed to the relevant regulatory authority, which in the case of the USA is the Food and Drug Administration. Once satisfied the FDA then allows the drug to be marketed. More testing may be required to be done in the post-marketing period as well as the continual monitoring of side effects. This is called Phase IV of the clinical trial.

To carry out all these phases is what take about 15 years and millions of dollars. It is therefore in the interest of the pharmaceutical industry to adopt new ways of developing drugs which are cost-effective and shorten the time-line by only working with potential successful leads early on and eliminating failures quickly. This is where Bioinformatics comes into play and may be 'what the doctor ordered'.

The pharmaceutical industry, though left behind in the information technology boom, has a unique chance to catch up. This however can only happen if the industry fully embraces Bioinformatics and nurture it until it is a well grounded discipline.

Bioinformatics and drug discovery

The drug discovery process pursued by major pharmaceutical companies begins with target identification and validation, assay development and high-throughput screening, the aim being to identify new leads. In pursuit of this paradigm, major pharmaceutical companies have established high-throughput screening facilities and have invested in automation to screen large compound libraries. Advances in Bioinformatics have enabled genome-wide analysis for a broad range of research fields (13).

Bioinformatics technology allows researchers to analyze the terabytes of data produced by the Human Genome Project. Gene sequence databases, gene expression databases, protein sequence databases and related analysis tools all help to determine whether and how a particular molecule is directly involved in a disease process. That in turn helps find new and better drug targets (1). This is essentially the essence of using Bioinformatics in drug discovery; identifying and validating targets.

Bioinformatics thus has definite 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 (14). The pharmaceutical industry has embraced genomics as a source of drug targets. Bioinformatics tools help by validating the potential drug targets and determining which ones are the most suitable for entry into the drug development pipeline.

It is interesting to note that while the traditional process of drug discovery uses synthetic organic molecules which are then tested in animals or whole organ preparations, nowadays with increased understanding of molecular biology, molecular target approach is used. Here virtual High-Throughput Screening (vHTS) is employed. This involves the screening of protein targets against databases of small molecules to determine the ones which fit into the target. The compounds which fit

are then further tested thus allowing researchers to only work with promising leads only. This results in cutting costs and time for the research process. Virtual High-Throughput Screening enables millions of compounds to be screened daily.

In computer aided drug design research, one often knows the genetic sequence of multiple organisms or the amino acid sequence of proteins from several species. It is very useful to determine how similar or dissimilar the organisms are based on gene or protein sequences. 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.

Determining the 3-D structure of proteins is an important aspect of drug design. 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. 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. Examples of tools used in homology modeling are Modeller, and Swiss-Model Repository (6).

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 (4).

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. Open Eye'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 (4).

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 hydro-phobicity and polarity that have a profound influence on how drugs bind to proteins (4).

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).

Benefits of Bioinformatics

CADD methods and Bioinformatics tools offer significant benefits for drug discovery programs which are discussed briefly below:

Costs

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. Growth of the Bioinformatics market is primarily attributed to its increased usage in the pharmaceutical industry. 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 (15). That is a valuable proposition in the global drug discovery market expected to be worth $25.1 billion in 2006. The global Bioinformatics market is forecast to grow to $3 billion in 2010 from its current $1.4 billion, for a compound annual growth rate (CAGR) of 15.8 percent, says BCC Research. Analysis software and services should drive this growth, rising to $1.2 billion in 2010 from $450 million in 2005 (16).

Time-line

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 (4).

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. 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 benefit that can help design research programs (4).

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 (4).

Bioinformatics thus clearly allows exploitation of the data that is available and this together with increased understanding of molecular biology and the molecular basis of disease greatly improves the drug discovery process. The data from the Human Genome Project has availed great opportunities for drug discovery and streamlining the choice of targets to support the drug discovery pipeline. The methods outlined above which are used in computer aided drug design mean that finding an attractive target is not an issue. The only concern is validating those targets to come up with the ones likely to succeed and here Bioinformatics tools are 'the saviour'

Bioinformatics tools can be used to gather all the necessary information about potential targets. This information includes nucleotide and protein sequencing, homologue mapping, function prediction, pathway information, disease associations, variants, structural information, gene and protein expression data and species distribution among others. The accumulation of this information into databases about potential targets means pharmaceutical companies can save themselves much time, effort and expense exerting bench efforts on targets that will ultimately fail (6).

As compared to the traditional method of drug discovery where a compound with potential pharmacological activity is isolated and 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. This is the way to go for pharmaceutical companies. Now armed with the resources from information technology and the human genome data, it only makes good economic sense to invest in the Bioinformatics sector and help make it work to their advantage. No more do scientists have to work hard coming up with a lead only for it to fail when finally tested in humans, resulting in incurring losses. Using Bioinformatics is more like a marketing aspect where one assesses the needs of the consumer and then comes with a product to meet those needs, instead of making a product first and then imposes it onto the consumer hoping that it will meet their needs. Surely the first method is the winner and Bioinformatics presents such a unique opportunity.

Although Bioinformatics is considered to have generated a lot of excitement and yet failed to deliver what it promised, that does not remove the clear advantages Bioinformatics brings to drug discovery. Bioinformatics may be considered to be a discipline in its infancy and as such it needs time to grow and really get organized. It is important to note that at the end of it, it is the patient buying prescription drugs from a chemist who has to enjoy the benefit of medicine which works and has minimal side effects at most importantly at an affordable price. The pharmaceutical industry should never lose this focus as it is easy to only consider productivity and profitability at the expense of the patient.

It is exciting to think of the possibilities that Bioinformatics may bring forth in terms of finding drugs for conditions such as cancer, AIDS, tuberculosis. The benefit for the industry is obviously in cutting costs and speeding up the process to get new drugs onto the market. For the common man on the street the major benefits would be of course drug affordability as well as access to life saving drugs in a shorter time. Many lives have been lost due to unavailability of such drugs and embracing Bioinformatics may reduce the loss and greatly improve the lives of humanity.

Conclusion

Bioinformatics clearly may be the answer to solve the drug discovery and cost woes of the pharmaceutical industry. Though still in its infancy and having been considered by some as having up to now failed to deliver and live up to its theoretical potential, the advantages and unique opportunities it brings can not be ignored. It holds one of the keys to dramatically cut the costs involved in drug discovery and ultimately the price of the drugs to the patients at the end of the chain. By eliminating potential drug failures early on during the process, it also helps cut the time scientists take to get a drug from the laboratory to the chemist's shop as they only concentrate their efforts on the leads which hold the greatest potential only. Thus if properly utilized, the pharmaceutical industry will increase the number of drugs in their pipelines which has been dwindling, drugs which are more effective due to the Bioinformatics tools employed. Reduced time of drug discovery also benefits the patients immensely as they will have quicker access to life saving drugs at an affordable price.

Thus Bioinformatics has the potential to hugely decrease the risk, cost and expertise required for the early stages of drug development, target selection and validation. Bioinformatics has proven indispensable for applying genomic technologies to drug discovery and development. Today pharmaceutical and life sciences companies are placing increasing emphasis on Bioinformatics investments in order to gain a competitive advantage in the drug discovery process. By fully integrating Bioinformatics in the drug discovery process, pharmaceutical companies could cut the cost of creating a new drug in half and remove two to three years off the development.

References

1. Drug Companies on Speed. www.cio.com/archive/101501/speed.html

2. Bioinformatics web-Comprehensive educational resource on Bioinformatics. www.geocities.com/Bioinformatics web.

3. Di Masi JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. Journal of Health Economics 2003; **22**: 151-85.

4. Clinical costs skyrocket to new levels. Phil Webster, Pharma-Biotech Research Analyst (Published 09/12/04 Source: Frost and Sullivan. www.inpharm.com/External/InpH/1,2580, 1-3-128917

5. Pharmaceutical Industry, Heal thyself. David Handelsman www.sas.com/news/feature/his/mar05dev.html

6. Genomics Provides Crucial Tools for Breaking Bottlenecks in Drug Discovery and Development. www.healthtech.com/ newsarticles/issueII_1.asp

7. IT Use for Drug Discovery to Rise. Stacey Lawrence. June 29, 2005 www.cioinsight.com/article2/0,1540, 1832630,00. Asp

8. What is Bioinformatics? Http://smi-webistanford-edu/people/altman/Bioinformatics.html

9. Faststats: Drug Discovery-Drug Discovery buzz June 19, 2005. www.redherring.com/Article.aspx?a=12388~hed=faststats

10. Bioinformatics in Computer-Aided Drug Design by Dr Richard M. Casey (Published May 10, 2005). www.b-eye-network.com/view/852

11. Bioinformatics and drug discovery www.ebi.ac.uk/2can/disease/genes11.html.

12. www.hypography.com/topics/Bioinformatics_100147.cfm

13. www.biochem.northwestern.edu/holmgren/Glossary/Definitions/Def-B/Bioinformatics.html (27-07-04)

14. Stanley Belkowski, Sergey Ilyin. Integration of automation, Bioinformatics and proteomics spawns a new approach.

15. www.ddd.mag.com

16. www.activemedia-guide.com/pharmaceutical_industry.htm (Industry Analysis, Pharmaceuticals Industry, source: US Business Reporter, 17/11/01).