Journal of Bioinformatics and Sequence Analysis Vol. 2(5), pp. 89-94, June 2011
Available online at http://www.academicjournals.org/JBSA
ISSN 2141-2464 ©2011 Academic Journals

Review

Modern drug discovery process: An *in silico* approach

V. Srinivasa Rao and K. Srinivas*

Department of Computer Science and Engineering, V.R Siddhartha Engineering College, Kanuru, Vijayawada-520 007, India.

Accepted 3 June, 2011

Drug discovery process is a critical issue in the pharmaceutical industry since it is a very cost-effective and time consuming process to produce new drug potentials and enlarge the scope of diseases incurred. Drug target identification, being the first phase in drug discovery is becoming an overly time consuming process. In many cases, such produces inefficient results due to failure of conventional approaches like *in vivo* and *in vitro* to investigate large scale data. Sophisticated *in silico* approaches has given a tremendous opportunity to pharmaceutical companies to identify new potential drug targets which in turn affect the success and time of performing clinical trials for discovering new drug targets. The main goal of this work is to review *in silico* methods for drug discovery process with emphasis on identifying drug targets, where there are genes or proteins associated with specific diseases. This review provides a succinct overview of several recent approaches that employ bioinformatics for the systematic characterization of the targets of bioactive compounds.

Key words: *In silico*, infectious vector, drug, pharmaceutical industry.

INTRODUCTION

In silico is an expression used to mean "performed on computer or via computer simulation (http://en.wikipedia.org/wiki/Binding_site). *In silico* drug designing is a form of computer-based modeling whose technologies are applied in drug discovery processes. Unlike the historical method of drug discovery, by trial-and-error testing of chemical substances on animals, and matching the apparent effects to treatments, *in silico* drug design begins with a knowledge of specific chemical responses in the body or target organism and tailoring combinations of these to fit a treatment profile (http://www.scfbio-iitd.res.in/tutorial/drugdiscovery.htm). A drug is a substance used in the diagnosis, treatment, or prevention of a disease or as a component of a medication. The drug discovery process is aimed at discovering molecules that can be very rapidly developed into effective treatments to fulfill unmet medical needs for both endogenous diseases, that arise from in-born sequence errors in germ cells or spontaneous (or age-related) mutations in somatic cells and exogenous diseases, that arise from an infectious vector (virus, bacterium or parasite) that has its origins outside (Yongliang et al., 2009). In most basic sense, drug design involves design of small molecules that are complementary in shape and charge to the bimolecular target to which they interact and therefore will bind to it. The identification of potential drug target is valuable and significant in the research and development of drug molecules at early stage, safety evaluation and old drugs with new use. Due to the limitation of throughput, accuracy and cost, experimental techniques cannot be applied widely, therefore, the development of *in silico* target identification algorithms, as a strategy with the advantage of fast speed and low cost, has been receiving more and more attention worldwide. It has been of great importance to develop fast and accurate target identification and prediction method for the discovery of targeted drugs, construction of drug-target interaction network as well as the analysis of small molecule regulating network (Markus et al., 2007). Computational tools offer the advantage of delivering new drug candidates more quickly and at a lower cost. Major roles of computation in drug discovery are: 1) Virtual screening and de novo design, 2) *in silico* ADME/T prediction and 3) advanced methods for determining protein-ligand binding and structured based drug design.

*Corresponding author. E-mail: kudipudi72@gmail.com.

The process of finding a drug molecule that attaches itself to the target protein in the body has now moved from the lab to the computer. When a protein target has been identified, it needs to be screened against a large database of known molecules, a method called virtual high-throughput screening (http://www.minglebox.com/blog/l7530/post/in_silico_drug_design). If a molecule fits the target protein, it can be further tested. Knowledge from molecular modeling, molecular biology and combinatorial chemistry is coded into software programs that help predict how the protein target and the drug molecule will come together. The techniques used today, which are continually being improved, do an approximate match to come up with possible hits or pairings. It would be impossible to test millions of molecules in a lab against a given target, but the computerized screening process helps in choosing a small set of molecules that can then be tested in a wet lab (http://www.kauhort.in/Download/SBDD.pdf). This review provides a succinct overview of several recent approaches that employ bioinformatics for the systematic characterization of the targets of bioactive compounds.

TYPES OF DRUG DESIGN

There are two types of drug design; one is “rational drug design” and the other is “structure based drug design”.

Rational drug design (RDD)

Rational drug design is a process used in the biopharmaceutical industry to discover and develop new drug compounds. RDD uses a variety of computational methods to identify novel compounds, design compounds for selectivity, efficacy and safety, and develop compounds into clinical trial candidates. These methods fall into several natural categories such as, structure-based drug design, ligand-based drug design, de novo design and homology modeling, depending on how much information is available about drug targets and potential drug compounds.

Structure based drug design (SBDD)

Structure-based drug design is one of the first techniques to be used in drug design. 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 x-ray crystallography or nuclear magnetic resonance (NMR) techniques. X-ray and NMR methods can resolve the structure of proteins to a resolution of a few angstroms (about 500,000 times smaller than the diameter of a human hair) (Suresh et al., 2006). At this level of resolution, researchers can precisely examine the interactions between atoms in protein targets and atoms in potential drug compounds that bind to the proteins. This ability to work at high resolution with both proteins and drug compounds makes SBDD one of the most powerful methods in drug design. Structure- based design refers specifically to finding and complementing the 3D structure (binding and/or active site) of a target molecule such as a receptor protein. Chemists may be guided to subsets of compounds with desired features to complement 3-dimensional shape of the site. From the geometry and functional features of the binding site, complementary structures of a compound (ligand) are so designed as to have high binding affinity with the target molecule. It is a powerful technique to design a corresponding ligand specifically interacting with the target, particularly for the development of a novel therapeutic through stimulation or inhibition of the receptor protein.

MODERN DRUG DISCOVERY PIPELINE

Drug discovery process operates on a target-based approach, in which the organism is seen as a series of genes and pathways and the goal is to develop drugs that affect only one gene or molecular mechanism (that is, the target) in order to selectively treat the deficit causing the disease without producing side effects (api.ning.com/files/.../DrugDiscoveryNewDrugDevelopmentProcess.ppt). Computers can be used to simulate a chemical compound and design chemical structures that might work against it. Enzymes attach to the correct site on a cell’s membrane, which causes the disease. A computer can show scientists what the receptor site looks like and how one might tailor a compound to block an enzyme from attaching there (http://www.microcal.com/drug-discovery-development/small-molecule/target-identification.asp.).

Figure 1 shows the stages of modern drug discovery process pipeline to find out the new drugs. Modern drug discovery process pipeline consists of seven important steps: target identification, target validation, hit and lead identification, lead optimization, pre clinical testing, chemical testing and new drug application (NDA) and food and drug administration (FDA) approval.

Target identification

Target-based drug discovery begins with the identification of the function of a potential therapeutic drug target and understanding its role in the disease process

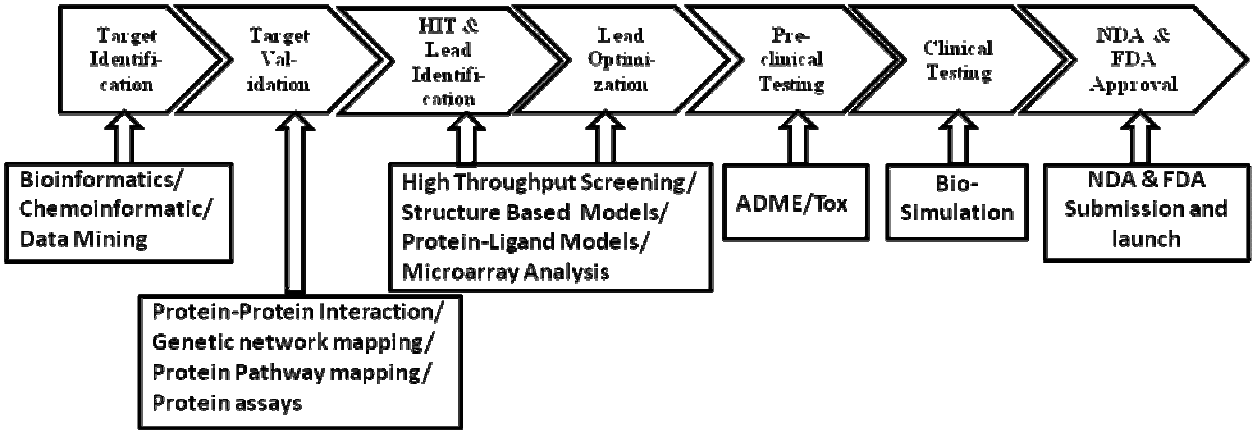

Figure 1. Modern drug discovery process pipeline.

(http://en.wikipedia.org/wiki/Drug_design). A target is generally a single molecule, such as a gene or protein, which is involved in a particular disease. A drug target is a key molecule involved in a particular metabolic or signaling pathway that is specific to a disease condition or pathology, or to the infectivity or survival of a microbial pathogen. Some approaches attempt to inhibit 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. Another approach may be to enhance the normal pathway by promoting specific molecules in the normal pathways that may have been affected in the diseased state (Taylor and Francis, 2006). The exact target and the specific patient population are identified with the help of bioinformatics, chemoinformatics and/or data maiming approaches such as, homology based, ligand base, structures base, high throughput screening (HTS), text mining, microarray technologies, pattern matching etc.

Target validation

After a drug target has been identified, a rigorous evaluation needs to occur to demonstrate that modulation of the target will have the desired therapeutic effect. In the drug-discovery process, the major bottleneck is target validation. If this process can be accelerated with computational tools, the target validation step will speed up significantly. The target-validation process includes determining if the modulation of a target's function will yield a desired clinical outcome, specifically the improvement or elimination or a phenotype. *In silico* characterization can be carried by using approaches such as genetic-network mapping, protein-pathway mapping, protein–protein interactions, disease-locus mapping, and subcellular localization predictions. Initial selection of a target may be based on the preliminary results found between cellular location and disease/health condition, protein expression, potential binding sites, cross-organism confirmation, or pathways involved in a disease/health condition (Bleicher et al., 2003).

HIT and lead identification

For many targets in drug discovery, the identification of a small molecule 'hit' as a starting point for the hit-to lead process. The identification of small molecule modulators of protein function and the process of transforming these into high-content lead series are key activities in modern drug discovery (Robert AG 2006). The "hit-to-lead" phase is usually the follow up of high-throughput screening (HTS). Hits can be identified by one or more of several technology-based approaches like high-throughput biochemical and cellular assays, assay of natural products, structure-based design, peptides and peptidomimetics, chemogenomics and virtual HTS, and literature- and patent-based innovations (Suresh et al., 2006). To develop efficient drug discovery practices, it is useful to consider the various strategies that have been reported for hit and lead identification; assay development, where the target is converted to an HTS assay system.

Lead optimization

Lead optimization is the complex, no-linear process of refining the chemical structure of a confirmed hit to improve its drug characteristics with the goal of producing drug candidate. Lead structures are optimized for target affinity and selectivity. Docking techniques are currently applied to aid on structure-based absorption, distribution, metabolism and excretion (ADME). Drug candidates discovered using this approach needs to be validated on

a disease-specific animal model to provide experimental proof of concept. This radical shift in the drug discovery process from physiology-based approach to target-based approach offers high screening capacity and supports to formulate simple, clear requirements to candidate drugs, which allows implementation of rational drug design (Klaus et al., 2001).

Pre-clinical testing

Preclinical studies and testing strategies with and without the use of animal testing methods have the purpose of limiting risks whenever a new active substance is to be used as a medicinal product in humans. They should be designed in such a way as to achieve as early, risk-free, unproblematic, and economic a transition as possible from preclinical to clinical trials in medicinal products development (Glossary of Clinical Trial Terms, NIH Clinicaltrials.gov).Scientists carry out *in vitro* and *in vivo* tests. *In vitro* tests are experiments conducted in the lab, usually carried out in test tubes and beakers ("*vitro*" is "glass" in Latin) and in vivo studies are those in living cell cultures and animal models ("*vivo*" is "life" in Latin). Preclinical testing involves: pharmacology, toxicology, preformulation, formulation analytical and pharmacokinetics.

Clinical testing

A clinical trial (also clinical research) is a research study in human volunteers to answer specific health questions. Carefully conducted clinical trials are the fastest and safest way to find treatments that work in people and ways to improve health. During the clinical trial, the investigators: recruit patients with the predetermined characteristics, administer the treatment(s), and collect data on the patients' health for a defined time period. The U.S. National Institutes of Health (NIH) organizes trials into five (5) different types: prevention trials, screening trials, diagnostic trials, treatment trials, quality of life trials and compassionate use trials or expanded access (Glossary of Clinical Trial Terms, NIH Clinicaltrials.gov).

NDA and FDA approval

The new drug application (NDA) is the vehicle in the United States through which drug sponsors formally propose that the food and drug administration (FDA) approve a new pharmaceutical for sale and marketing. The goals of the NDA are to provide enough information to permit FDA reviewers**.** The NDA includes all of the information from the previous years of work, as well as the proposals for manufacturing and labeling of the new medicine. FDA experts review all the information included in the NDA to determine if it demonstrates that the medicine is safe and effective enough to be approved.

SIGNIFICANCE OF *IN-SILICO* DRUG DISCOVERY PROCESS

As structures of more and more protein targets become available through crystallography, NMR and bioinformatics methods, there is an increasing demand for computational tools that can identify and analyze active sites and suggest potential drug molecules that can bind to these sites specifically. Time and cost required for designing a new drug are immense and at an unacceptable level. According to some estimates it costs about $880 million and 14 years of research to develop a new drug before it is introduced in the market. Intervention of computers at some plausible steps is imperative to bring down the cost and time required in the drug discovery process.

BIOINFORMATICS IN COMPUTER-AIDED DRUG DESIGN

Bioinformatics is the interdisciplinary subject to solve the biological problems in the life science using computational approach. Computer-aided drug design (CADD) is a specialized discipline that uses computational methods to simulate drug-receptor interactions. CADD methods are dependent on bioinformatics tools, applications and biological databases. There are several key areas in bioinformatics regarding CADD research.

Homology modeling

Homology modeling, also known as comparative modeling of protein refers to constructing an atomic-resolution model of the "target" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "template"). Most drug targets are proteins, so it is important to know their 3-D structure. It is estimated that the human body has 500,000 to 1 million proteins. Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 20% sequence identity can have very different structure (http://en.wikipedia.org/wiki/Homology_modeling). Modeller is a well-known tool in homology modeling, and the Swiss-Model Repository is a database of protein

structures created with homology modeling (Richard, 2005).

Interaction networks

Docking is a method used to identify the fit between a receptor and a potential ligand. Docking actually consists of two distinct parts, the “docking” part, which is the search scheme to identify suitable conformations or poses, and “scoring”, which is a measure of the affinity of various poses (http://mndoci.com/2007/02/10/evaluating-protein-ligand-interactions/). Interaction network is a network of nodes that are connected by features. If the feature is a physical and molecular, the interaction network is molecular interactions usually found in cells (http://en.wikipedia.org/wiki/Interaction_network). Protein-ligand docking is a molecular modelling technique. The goal of protein-ligand docking is to predict the position and orientation of a ligand (a small molecule) when it is bound to a protein receptor or enzyme. Docking techniques are employed in the virtual screening of large databases of available chemicals in order to select likely drug candidates (http://en.wikipedia.org/wiki/Protein-ligand_docking). The following are the protein-ligand docking software’s:

i) Gold (http://www.ccdc.cam.ac.uk/products/life_sciences/gold).
ii) AutoDock (http://autodock.scripps.edu/).
iii) Dock (http://dock.compbio.ucsf.edu/).
iv) ZDock (http://zlab.bu.edu/zdock/).
v) Docking server: http://www.dockingserver.com/web.

Microarray analysis

Microarray analysis, known as DNA technology is a new technology now promises to advance biotechnology further. Microarrays are simply ordered sets of DNA molecules of known sequence. Usually rectangular, they can consist of a few hundred to hundreds of thousands of sets. Each individual feature goes on the array at precisely defined location on the substrate. The identity of the DNA molecule fixed to each feature never changes. Scientists use that fact in calculating their experimental results. Microarray analysis permits scientists to detect thousands of genes in a small sample simultaneously and to analyze the expression of those genes. As a result, it promises to enable biotechnology and pharmaceutical companies to identify drug targets. Since it can also help identify individuals with similar biological patterns, microarray analysis can assist drug companies in choosing the most appropriate candidates for participating in clinical trials of new drugs. In the future, this emerging technology has the potential to help medical professionals select the most effective drugs, or those with the fewest side effects, for individual patients. It has potential applications in several fields, such as tissue microarrays for cancer and other diseases, normal tissues and cells during development, and studies of transgenic animals.Microarray technology has the potential to be used to develop new drugs. "It's clearly been demonstrated that the technology can identify genes that have been upregulated or downregulated," says Jeffrey Williams, CEO of Genomic Solutions, based in Ann Arbor, Michigan (http://www.sciencemag.org/site/products/micro.xhtml).
DruTiMine (drug target integrative miner), which is an IBM research project conducted at IBM Centre of Advanced Studies in Cairo (Cairo-CAS) is to identify drug targets for a specific disease: genes, proteins, and chemical compounds by utilizing text mining, association rule mining, and machine learning techniques (Hisham KH (2003).

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 (vHTS). 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. ZINC is a good example of a vHTS compound library (Satyajit et al., 2010).

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 and secondary structure 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 (Satyajit et al., 2010).

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 such as C2-ADME, TOPKAT, CLOGP, DrugMatrix, AbSolv, Bioprint, GastroPlus etc.

CONCLUSION

The drug discovery and development process is a long and expensive one. It stars from target identification, after that, validates the targets and identifies the drug candidates. Before any newly discovered drug is placed on the market, it must undergo extreme preclinical and clinical tests and get the FDA approval. Due to the limitation of throughput, accuracy and cost, experimental techniques cannot be applied widely, therefore, in recent times the drug discovery process has shifted to *in silico* approaches such as homology modeling, protein-ligand interactions, microarrary analysis, vHTS etc. *In silico* approach has been of great importance to develop fast and accurate target identification and prediction method for the discovery.

REFERENCES

api.ning.com/files/.../DrugDiscoveryNewDrugDevelopmentProcess.ppt.
Richard C (2005). Bioinformatics in Computer-Aided Drug Design, May 10.
Bleicher KH, Bohm HJ, Muller K, Alanine AI (2003). "Hit and lead generation: Beyond high-throughput screening." Nat. Rev. Drug Discov., 2(5): 369-378. May.
Suresh KP, Siddharth M, Rao V, Vadlamudi SV (2006). Current approaches in drug discovery. Pharma Times - Vol 38 - No. 8 - August.
Hisham E, Taysir HA, Soliman IE (2006). "Mining Drug Targets: The Challenges and a Proposed Framework." Proc. 11th IEEE Symp. Comput. Commun. (ISCC'06).
http://en.wikipedia.org/wiki/Binding_site
http://en.wikipedia.org/wiki/Drug_design
http://en.wikipedia.org/wiki/Homology_modeling
http://en.wikipedia.org/wiki/Interaction_network
http://en.wikipedia.org/wiki/Protein-ligand_docking.
http://english.simm.cas.cn/rp/201007/t20100701_55978.html
http://mndoci.com/2007/02/10/evaluating-protein-ligand-interactions/
http://www.kauhort.in/Download/SBDD.pdf
http://www.microcal.com/drug-discovery-development/small-molecule/target-identification.asp.
http://www.minglebox.com/blog/l7530/post/in_silico_drug_design
http://www.scfbio-iitd.res.in/tutorial/drugdiscovery.htm
http://www.sciencemag.org/site/products/micro.xhtml
Klaus ODVM, Peter GU, Nzel DVM, Rolf BMD (2001). "Preclinical Testing Strategies." Drug Inf. J. 35: 321–336.
Markus HJ Seifert JK, Bernd K (2007). "Virtual high-throughput screening of molecular databases." Curr. Opin. Drug Discov. Dev., 10(3): 298-307. May
Robert AG (2006). "Hit and lead identification: Integrated technology-based., 3 (4)
Satyajit D, Sovan S, Kapil S (2010). Computer Aided Drug Design- A New Approach in Drug Design and Discovery. Int. J. Pharm. Sci. Rev. Res., 4(3). September – October.
Suresh KP, Siddharth M, Rao V, Vadlamudi SV (2006). "Current approaches in drug discovery." Pharma Times, 38(8). August.
Yongliang Y, Adelstein SJ, Amin IK (2009). Target discovery from data mining approaches. 20 January.
Taylor, Francis (2006). *"In silico* Technologies in Drug Target Identification and Validation." Taylor and Francis Group, LLC.
Xiaofeng L, Sisheng O, Biao Y, Yabo L, Kai H, Jiayu G, Siyuan Z, Zhihua L, Honglin L, Hualiang J (2010). "PharmMapper server: A web server for potential drug target identification using pharmacophore mapping approach." Nucleic Acids Res., 38: W609-W614.