Innovare Academic Sciences

International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 7, Issue 4, 2015

Review Article

DRUG DISCOVERY: AN APPRAISAL

V. UMASHANKAR[1*], S. GURUNATHAN[2]

[1]Head of the Department, Center for Bioinformatics, Vision Research Foundation, Sankara Nethralaya, Chennai, Tamilnadu, India 600006,
[2]Assistant Professor, Department of Biotechnology, National College, Tiruchirapalli, Tamilnadu, India 620001
Email: vumashankar@gmail.com

Received: 15 Jan 2015 Revised and Accepted: 10 Feb 2015

ABSTRACT

The process of drug discovery and development has undergone radical changes over the years. Introduction of several novel technologies in genomics, proteomics and other omics areas have enabled drug target identification and validation more specific. *In silico* virtual screening and other computational chemistry methods like QSAR (Quantitative Structure-Activity Relationship) and QSPR (Quantitative Structure-Property Relationship) have enabled the emergence of new drugs with minimal toxicity and higher efficacy in this post-genomics era. Moreover, initiative like Open Source Drug Discovery (OSDD) is playing a promising role in accelerating the pace of drug discovery process. Better understanding of these methods and initiatives by researchers will kindle interest towards adopting it. Hence, by this review, we intend to present a comprehensive view of overall transition and modernization of the drug discovery process and it's impacts on the scientific community.

Keywords: Computational Drug Discovery, Drug Discovery Pipeline, Modern Drug Discovery, Open Source, OSDD.

INTRODUCTION

In the past decade, emergence of microbial drug resistance [1] and complicated, new diseases and unexpected adverse drug effects has accelerated the search for potential therapeutic molecules [2]. From being necessary, the process of discovering new drugs, thus has grown into an inevitable and ever evolving phenomenon.

In this modern era, drug discovery has developed into an interdisciplinary scientific field integrating diverse disciplines of biology, chemistry, mathematics and computers [3]. Any novel chemical entity with potential therapeutic value is extensively studied for its safety and efficacy before it is marketed for public use. This multi-stage process is commonly referred as "Drug Discovery Pipeline" or "Development Chain" (fig. 1) [4]. All the early stages of the pipeline phenomenon viz. Identification and validation of the drug target, lead discovery and lead identification is collectively represented by the term "Drug Discovery".

Late stage processes of the pipeline viz. Preclinical testing, clinical studies and regulatory approval are collectively denoted as "Drug Development" [5, 6]. Drug discovery pipeline commonly is known to be a complex, lengthy and expensive phenomenon. An estimated one billion dollars is spent in developing a new drug [7]. The average time span from the point of identifying a clinical candidate for approval of a new drug is approximately around ten years [8].

Historical developments in drug discovery

Traditionally, drugs were discovered by employing predominantly chemistry and pharmacology-based cautious approach [9, 10]. Most of these approaches were blind/random screening that is not only laborious but time-consuming too [11]. Since the early days, there has been a tremendous and continuous evolution of modernized drug discovery. For ease of understanding, historical developments of the drug discovery process can be briefly discussed under three time periods: Early drug discovery period (before 19th century), modern drug discovery period (19th century) and Drug Discovery and Drug Development (DD) period (20th century and beyond).

Early drug discovery till 19th century

Natural products have been and still are the most important source of drugs or drug precursors [12]. In addition to their historical success in drug discovery, natural products continue to be sources of new commercially viable drug leads [6]. A simple literature search indicates the existence of the drug discovery process in early civilizations.

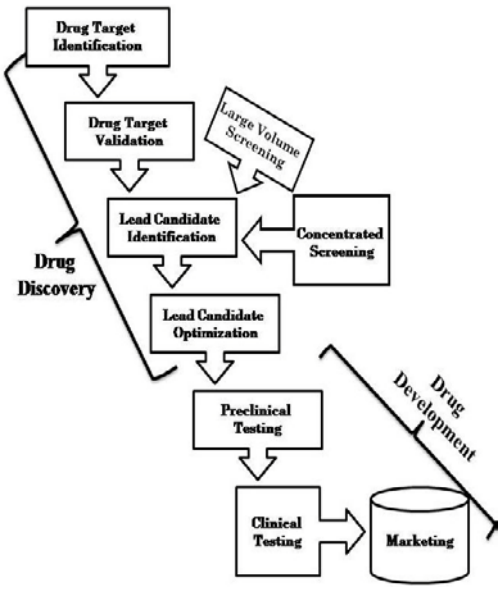

Fig. 1: Modern Drug Discovery Pipeline with seven consecutive steps. The process is lengthy and time-consuming. The application of computational approaches to early stages of the pipeline has considerably reduced time and cost involved [13].

Ancient men relied upon nature for their healing needs. This is clearly evident in their reports describing medicinal properties of several plant extracts. However, such descriptions resulted only after several trial and error procedures during therapy of specific illnesses [13]. Thus, historically, the discovery of drugs has been serendipitous. Verbal and written means helped to share and/or maintain the acquired knowledge on extracts as medicines [11]. As an entire plant and/or its parts were utilized in medication, most individuals were consuming multiple chemical entities. Also, they lacked knowledge of its chemical content or their synergistic effect in curing the ailment [13].

Modern drug discovery in 19th century

Modern drug discovery consisted of a series of thematic developments beginning in the early years of the nineteenth century [13]. Introduction of techniques for separating individual

components in extracts paved way for availability of single entity drugs. A typical example is quinine. Though, Quinine was found earlier by explorers, it was only isolated in 1823 [14]. Several other molecules were also synthesized and developed into commercial entities during this century. Till late nineteenth century, the focus of drug research was on testing and evaluating existing natural products [13]. In late 1800s, pharmaceutical/chemical companies such as Bayer and Hoechst were established. Scientists at Bayer were successful in chemically synthesizing aspirin in 1895 [15]. Development of quinine analogues as antimalarial drugs also began in early 1900's [14].

Drug discovery and drug development in 20th century and beyond

Development of the modern pharmaceutical industry dates back to the beginning of the twentieth century [15]. Nobel Laureate Paul Ehrlich (considered the father of modern chemotherapy) and other scientists developed chemical processes for synthesis of drugs in laboratories. This concept of synthetic drugs gained further momentum in 1930s following discovery of sulfa drugs. Simultaneously, large-scale manufacturing of insulin also commenced due to the impact of molecular biology methods. These events catalyzed rapid progress of pharmaceutical industries. In 1930, an American Regulatory Authority named Food and Drug Administration (FDA) was established with a primary goal to "protect public health". Penicillin discovery by Alexander Fleming in 1928 was followed by the discovery of several other novel therapeutic molecules from microbes at quick succession in the 1940s. Concurrently, extensive development of antihistamines, analgesics, barbiturates, hormones (e. g., Epinephrine), sedatives, hypnotics and antidepressants happened. Commencement of industrial-scale penicillin manufacture also materialized. Such developments allowed major pharmaceutical industries to flourish, both scientifically and commercially. In spite of biological source of origin, several of these molecules were manufactured by direct chemical synthesis [8, 15]. The majority of drugs from natural sources were isolated in academic laboratories. However, several synthetic drugs were also developed in industrial or research laboratories [13].

Until the middle of the twentieth century, serendipity methods of discovering new drugs were followed, leading to the discovery of several successful drugs such as chlorpromazine, meprobamate and benzodiazepines [11]. Fig. 2 represents the drug discovery process followed by pharmaceutical companies in the middle of the twentieth century. However, important drawbacks such as more time-consuming processes, high expenses in discovering new drugs and reduced success guarantee made conventional randomized drug search phenomenon no longer effective. Also, there was a widespread need for more deterministic approaches to battle diseases. Thus, the concept of "Rational Drug Design" came into existence in the late 1960s.

The concept of QSAR (Quantitative Structure-Activity Relationship) introduced by Hansch and Fujita in 1960 encouraged computer-aided drug design. Approximately after ten years, the use of structural biology in drug discovery was strongly recommended by several scientists. At around the same time, usage and production of semi-synthetic antibiotics gained popularity. Several biopharmaceutical companies emerged in the later quarter of the twentieth century as a result of the infusion of biotechnology into pharmaceutical industries. Eventually, acceleration in discovery of new drug entities occurred. Meanwhile, accumulation of pharmacological bases of drugs and diseases coupled with progresses in biology and chemistry lead to a more rational linear drug development process. Several drug targets identified were protein in nature, particularly enzymes. Due to the emergence of technologies based on computers, drug discovery took a new dimension during the end of 1970s. In 1980s, the introduction of *in vitro* assays using animal tissues instead of more invasive and conventional *in vivo* methods gave valuable information on structure-activity relationships and eventually pharmacophore construction. Additionally, significant reduction in experimentation expenses was observed [11, 16]. Drug discovery process in 1980s is shown in fig. 3.

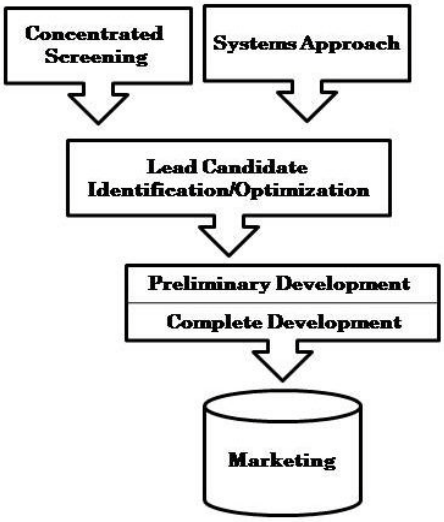

Fig. 2: Drug discovery process during the middle of the twentieth century. Serendipity and blind screening dominated the process [13]

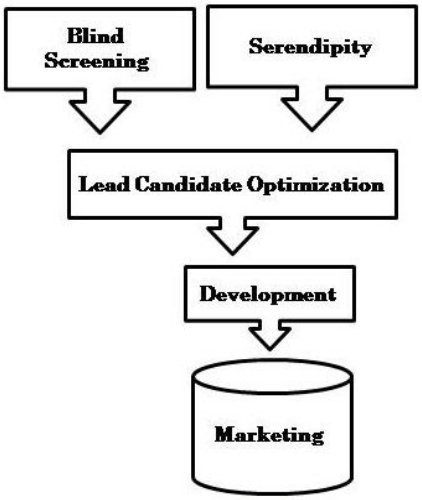

Fig. 3: Drug discovery process during 1980s. Improved screening methods, structure-activity relationships and enhanced safety of the drug molecules were the point of focus in the process [13]

The traditional (i.e., pre-1990s) drug discovery process involved initial lead generation based on the natural ligands, existing drugs, and literature based leads (fig. 2 & 3). New compounds would be synthesized and tested for biological activity and structure–activity relationships. This aided in the optimization of leads using traditional medicinal chemistry techniques [8]. In the past decades, Lipinski's "Rule of Five" was one of the new concepts that drastically altered the traditional paradigm for drug discovery and development [17].

Drug regulations came into existence due to the emergence of multitude of drug activity assessment methods. Hence, from early 1950s, the Food and Drug Administration (FDA) expanded its activities and initiated enforcement of drug laws focusing drug discovery phenomenon. However, the drug discovery process today is very heavily regulated, mainly to ensure the safety and protection of the general population [8, 11].

Huge investments by pharmaceutical industries in Life Sciences over the last two decades of the 20th century had led to the development of new technologies and methodologies. High-throughput screening

(HTS) enabled researchers to quickly identify promising candidates out of large compound libraries for further development. Combinatorial chemistry enabled the chemists to synthesize large amounts of diverse compound libraries required by biologists. Drug discovery process transformed drastically due to continuous development and improvement of molecular biology (e. g. Gene cloning techniques, *in vitro* expression technology and site-directed mutagenesis). Time for the new drug search and reaction mixture volumes was saved significantly by these processes [8, 15, 18]. According to Sneader [13], drug research in the twentieth century was driven by the development of drug analogues; more importantly only a few of which have been isolated in the preceding 50 years.

In order to reduce cost as well as to increase resource utility, drug discovery programs in different therapeutic areas (such as infectious diseases, oncology, immunology, cardiology, etc.) were organized such that biology and sometimes chemistry remain committed to that area, but other functions (screening, animal testing, pharmacology, structural biology, etc.) may be shared. Till date, these approaches are being successfully followed by several pharmaceutical companies [8, 13, 15].

The traditional approach to drug discovery has given way to a more modernized information-based approach – Bioinformatics [9]. According to Drews [19], drug discovery process after 1990s typically involved the following stages: target identification, target validation, lead identification, candidate(s) selection (Flowchart 3). By the end of the twentieth century, several pharmaceutical companies dropped traditional approaches involved in the screening process. Instead, they employed all modern science technologies available at that time [13]. In 1990s and the new millennium, tremendous improvements in biomedical knowledge and technology have necessitated a complete redesign of the drug discovery process [3]. Also, major changes occurred in the pharmaceutical industry from the vantage points of research and development as well as commercial operations. Novel technologies and processes such as "high throughput screening" and "combinatorial chemistry" were widely embraced and were developed to a high state of performance during this period [8].

Modern drug discovery in post-genomics era

Dawn of twenty first century witnessed beginning of the post-genomic era ushered by completion of the most promising Human Genome Project (HGP) announced in 1990s [20, 21]. Apart from the ultimate goal of identifying novel drug targets, obtaining proprietary rights to use those targets were also the goal of this large-scale HGP effort [22, 23]. The "reverse pharmacology" era is defined by the need to first clarify the biology and medical perspectives of the target so as to qualify it as druggable and pharmaceutically exploitable for drug discovery and development scheme [24]. Drug research has been greatly transformed by the "omics revolution". In this post-genomic era, several genomics and proteomics technologies, including microarray and sequencing technologies have enabled a paradigm shift in the drug discovery process [25]. Technological developments in areas such as proteomics, genomics and high-throughput screening are also beginning to impact significantly upon the early stages of drug development [5, 15]. In the post genomic era, Computer-Aided Drug Design (CADD) has found significant applications in almost all stages in the drug discovery pipeline [26].

Identification of novel drug targets in human genome sequence has provided new avenues for discovering new drugs and ensures better success rates for their approval. It is believed that the great majority of drugs reaching the market in the near future would have resulted from genomics. Additionally, advances in genomics enable better computational tools which eventually are expected to reduce cost of genetic testing [25, 27]. Pharmacogenomics implies the application of genomic technologies to drug discovery and development [18]. Pharmacogenomics technologies applied to conventional drug development processes can help us reap several potential benefits [28]. Currently, usage of pharmacogenomics principles is not only restricted to early drug discovery processes, but extends into preclinical and clinical trial studies too [29]. Recently, emerging applications of genomic-driven findings to cancer therapy in oncogenomics have been highlighted [30]. Using genomic technologies judiciously in the drug discovery process can save time and money due to improved efficiencies in several of its stages. On an average, Pharmaceutical companies save US$300 million and two years per drug by employing genomics technology, largely as a result of efficiency gains [25].

The HGP has renewed interest in proteomics eventually reflecting its crucial role in pharmaceutical developments [25]. Since majority of drug targets are proteins, integrating proteomics-based approaches into the conventional drug discovery phenomenon will help to develop novel drugs [31]. Proteomics has consequently assumed a central place in the early stages of modern drug development [5].

CADD in comparison to traditional HTS and combinatorial chemistry employs better targeted search. Hence, it has better hit rates in a screening process to identify novel drug compounds. CADD finds applications in three major sections of drug discovery pipeline viz. In virtual screening process to identify lead compounds for experimental studies, in lead compound optimization for drug metabolism and pharmacokinetics properties and in designing novel drug compounds [32].

Bioinformatics coupled with CADD is a powerful combination in drug research and development. It helps researchers understand more details related to drug-receptor interactions. Thus, enabling design of new analogues with better interactions with the target. Bioinformatics tools have also been employed effectively by pharmaceutical companies in predicting 3D-structure of proteins, as well as ADMET properties of any potential drug molecule. Thus, finding significant applications in improving drug discovery phenomenon [33]. Biochemoinformatics, a combination of Cheminformatics and Bioinformatics, has enabled interaction of biological and chemical information effectively in today's drug discovery environment. Employing biochemoinformatics principles will reduce significant false positive data obtained from omics technologies [34].

With the introduction of integration and knowledge management solutions supported by computers, a new era is commencing in drug discovery [35]. Traditionally, pharmaceutical researchers focused on developing one drug to act against one target identified. Due to improved understanding of complex diseases in the post-genomic era, currently the focus has shifted to modulating multiple targets. Production of more efficient and less toxic drugs will soon happen once the drug-target network links to other biological networks [36].

Steps in drug discovery pipeline–brief outline

Modern Day Drug Discovery pipeline can be divided into seven consecutive steps viz. Drug Target Identification, Target validation, Lead discovery, Lead optimization, Preclinical studies, Clinical trials and Marketing (flowchart 1).

Drug target identification

Identification of targets on which any chemical entity can bind and act forms the first step in the drug discovery pipeline. Drug target is the specific binding site of a drug *in vivo* [19]. It belongs to a broad spectrum of moieties that includes molecular entities such as nucleic acids and proteins, disease biomarkers and biological pathways associated with specific disease conditions [37]. Two classes of drug targets are available, namely established targets and potential targets. The former class contains molecules with clear scientific knowledge and has complementary drugs. The latter class encompasses molecules with less or no scientific knowledge and lack drugs targeting them. Information supporting the role of these targets in disease modulation can come from a variety of sources [38]. Traditionally, research on targets and their discovery were largely happening in academic environments rather than in laboratories of pharmaceutical and biotechnology companies. Recently, genomic and proteomic approaches have enhanced drug target identification [5]. Approximate timeline for target identification is one year and a spending of around US$200 million has been documented [4].

Drug target validation

Drug target validation is taken up in the drug discovery pipeline following drug target identification step. It involves demonstrating the relevance and confirmation of the target protein in a disease process. This is accomplished primarily with knock-out or knock-in animal models [39]. Target validation is undertaken with the use of animal and disease models. An appropriate change in the behavior of diseased cells relative to normal cells provides stronger evidence for the target to be considered as potential one. Additionally, characterization of the most appropriate drug target forms a part of the validation process [4]. Apart from helping drug research, target validation provides better understanding of pathogenesis of target related diseases [40]. This stage of the pipeline is estimated to cost close to US$ 250 million and has a timeline of 2 years. Latest technologies have enabled combining both target identification and validation into a single process [5].

Lead candidate discovery / generation

Once drug targets have been identified and validated, identification of a suitable chemical moiety displaying interaction with the target is initiated. Identification of small molecule modulators and developing them into high-content lead series are key activities of contemporary drug discovery [41]. Molecules displaying better interactions are called "hits". Most commonly, hit compounds are derived from high-throughput screening (HTS). In the next step, from the "hits", compounds with attractive pharmaceutical properties such a low toxicity, membrane permeability, genotoxicity, etc. are identified. These compounds are often called ''leads''. Traditionally, ''hits'' have been found by screening, while ''leads'' are developed from ''hits'' by chemical synthesis. Leads can be discovered by any one of the following methods viz. Serendipitous method, random screening method, traditional chemical modification method and rational drug design method [5, 39]. Screening usually is done against a number of sources, namely natural compound libraries from microbes or plants documented from earlier traditional studies, commercially available compound collections, diverse proprietary libraries and combinatorial chemistry libraries. Based on the fact that many existing drugs are derived from natural products, they have served as excellent lead compounds. Hence, natural product screening is widely used as a method in finding lead compounds [4, 5].

Lead candidate optimization

Molecules, identified as "Leads" in the previous stage, are subjected to optimization work. This step is believed to be essential in contributing towards drug discovery process. At this stage, leads are modified to provide "best" analogues displaying improved potency, efficacy, pharmcokinetic and pharmacodynamic properties [5]. The changes are accomplished by chemical modifications chosen by structure activity analysis. If a target structure is known, structure based design could also be employed in introducing the changes [39]. As this process involves simultaneous optimization of multiple parameters, it is quite time consuming and a costly step. In the entire drug discovery process, lead optimization step is thought to be a rate-limiting step [5].

Preclinical studies

The outcome of the discovery phase is a handful of lead candidate compounds that have shown promising activity against a drug target. These are subjected to a battery of tests in pre-clinical studies. Most of these tests performed in animals are considered as a final preparation for the clinical evaluation of a potential drug candidate. Following studies are essentially a part of this stage: develop large scale synthesis; animal safety studies; carcinogenicity tests; drug delivery; elimination and metabolism studies; drug formulation experiments; animal dose-ranging studies [5, 39]. As only one third of candidate compounds develop into drugs, methods such as combinatorial lead optimization that automate and miniaturize toxicity and ADME testing were developed. These enable lead compounds to be tested earlier in the discovery process [4].

Clinical trials

In the entire drug discovery pipeline, this phase is considered the most costly [4]. Metabolic and pharmacological effects of the drug candidates in humans are studied in this phase. About 90% of drug candidates entering clinical trials fail [39]. Pharmaceutical clinical trials are commonly classified into four phases namely phase 1, 2, 3 and 4. In addition to testing safety and tolerability, pharmacodynamic and pharmacokinetic properties of the drug in normal, healthy human volunteers is tested in phase 1. In Phase 2 clinical trials, all tests of phase 1 are performed again. However, the tests are done on patients suffering from the targeted disease rather than healthy individuals. Large numbers of patients (hundreds to thousands) are studied in phase 3 stage. This is done to establish the definitive assessment of therapy in comparison with standard therapy. Following at least two successful phase 3 trials, the drugs are approved by FDA for marketing [5, 39].

Marketing

After successful clearance of a drug candidate in Phase 1, 2 and 3 trials, it is allowed for sale in the market only after approval by relevant authorities (for eg. FDA). However, monitoring post-launch safety and detecting rare or long term adverse effects over a large patient population and time period forms the phase 4 of clinical trials [5, 39].

Computational methods in drug discovery

Delivering new drug candidates more quickly and at lower cost is the need of the hour. To achieve these objectives, computational approaches in drug discovery process have become quite a necessity. Continued progress in the application of computational power to chemical and biological space has significantly impacted modern drug development chain [42]. In today's world, apart from several omics technologies, other modern technologies such as combinatorial chemistry, virtual screening, *in silico* ADMET screening and structure-based drug design (SBDD) have revolutionized the drug discovery process [43]. More importantly, the role of computers and computational techniques in drug discovery and development process have gained popularity and implementation. A simple example is drug-target interaction studies. Since experimental approaches are laborious and costly, prediction of these interactions by *in silico* methods provides valuable information in supporting experimental data. Thus, computational methods are considered complementary to the experimental techniques [36, 44].

Computer Aided Drug Discovery and Development is being utilized in early stages of DD process that includes hit identification, lead selection and optimization [44]. Past three decades have witnessed the development of therapeutic small molecules solely based on Computer-aided drug discovery/design methods [32].

Major pharmaceutical and biotechnology companies worldwide is using computational design tools [45]. Structure-based drug design is considered as one of the most innovative and powerful approaches in drug design [35]. Virtual screening has been shown more efficient than commonly used empirical screening. To significantly reduce the time and resource requirements of chemical synthesis and biological testing, *in silico* modelling is employed. Similarly, QSAR and QSPR are commonly used computational methods in predictive toxicology [44].

Table 1, 2 and 3 lists commonly employed tools and databases in computational drug discovery and development process. Several of them have applications in the early stages of the drug discovery pipeline.

"Open source" concept

The concept of "open source" has hugely impacted the software industry globally [78]. Its roots can be traced back to the beginning of computer software development. A ten point criteria were introduced by Open source initiative to define the term "open source". Of the ten points mentioned, three are considered to be major ones, namely access to source code, free redistribution, and creation of derived works [79].

Table 1: Bioinformatics tools and databases commonly employed in DD process

Tool name	Brief description of the tool	Steps involved in DD process
BLAST (Basic Local Alignment Search Tool) [46]	A DNA and protein sequence alignment tool	Target Identification and Validation
FASTA (Fast Alignment) Tool [47]	A DNA and protein sequence alignment software package	Target Identification and Validation
EMBOSS (European Molecular Biology Open Software Suite) [48]	A free Open Source software analysis package specially developed for the needs of the Molecular biology user community	Target Identification and Validation
BioEdit (Biological Editor) [49]	A biological sequence alignment editor with multiple document interface for easy alignment and manipulation of sequences on a desktop computer.	Target Identification and Validation
ClustalW [47]	A general purpose multiple sequence alignment program to study evolutionary relationships	Target Identification and Validation
RasMol (Raster Molecule) tool [50]	A molecular visualization program tool for DNA/RNA and protein structures.	Structure Based Drug Design, Target Identification and validation
PyMOL [51]	Molecular visualization System for DNA/RNA and protein structures.	Structure Based Drug Design, Target Identification and validation
Swiss-PDB Viewer [52]	Standalone molecular visualization and modeling tool with advanced features to handle nucleic acid, protein and other organic molecules.	Structure Based Drug Design, Target Identification and validation
Discovery Studio [53]	Advanced software focusing on modeling and simulation solutions.	Structure Based Drug Design, Target Identification and validation, Lead selection, Lead optimization, ADME studies
Swiss-Modeller [54]	Fully automated protein structure homology modeling server	Structure Based Drug Design, Target Identification and Validation
Modeller [55]	Stand alone comparative modeling tool for 3D structures of proteins	Structure Based Drug Design, Target Identification and Validation
PHYRE [56]	Automatic Fold recognition server for predicting the structure and function of protein sequence	Structure Based Drug Design, Target Identification and Validation
PubMed [57]	Free search engine accessing primarily the MEDLINE database of references and abstracts on life sciences and biomedical topics	Target Identification and Validation
DDBJ (DNA Data Bank of Japan) [58]	Collects and distributes nucleotide sequence data	Target Identification and Validation
NCBI Genbank [59]	Genetic sequence database	Target Identification and Validation
PDB (Protein Data Bank) [60]	An Information Portal to Biological Macromolecular Structures	Structure Based Drug Design, Target Identification and Validation
KEGG (Kyoto Encyclopedia of Genes and Genomes) [61]	Database resource for understanding high-level functions and utilities of the biological system	Structure Based Drug Design, Target Identification and Validation

Table 2: Chemiinformatics tools and databases commonly employed in DD process

Tool Name	Brief description of the tool	Steps involved in DD process
ISIS Draw [62]	A chemical structure drawing program available free of cost for academic and personal use	Lead structure determination, Lead optimization, ligand based drug design
ChemDraw [63]	A molecule editor to handle chemical molecules and is part of the ChemOffice suite of programs.	Lead structure determination, Lead optimization, ligand based drug design
ACD Chemsketch [64]	Advanced chemical drawing tool available free of cost for academic use.	Lead structure determination, Lead optimization, ligand based drug design
MarvinSketch [65]	Advanced chemical editor for drawing chemical structures, queries and reactions.	Lead structure determination, Lead optimization, ligand based drug design
JME Molecular Editor [66]	A Java applet which allows to draw /edit molecules and reactions and to depict molecules directly within an HTML page.	Lead structure determination, Lead optimization, ligand based drug design
ISIS/Base [67]	A database management system for storing, searching, and retrieving chemical structures and associated scientific data	Lead identification, lead optimization, ligand based drug design, virtual screening
ACD Chemfolder [68]	Advanced software to create and manage databases with thousands of chemical structures and reactions	Lead identification, lead optimization, virtual screening
Chemspider [69]	Free chemical structure database	Lead identification, validation and optimization
PubChem [70]	Database containing structures and physiochemical properties of chemical compounds	Lead identification, validation and optimization
CSD (Cambridge Structural Database) [71]	Contains experimentally determined 3D structures of potential ligand molecules	Lead identification, validation and optimization, ligand based drug design
ChEMBL [72]	Chemical database of bioactive molecules with drug-like properties	Lead identification, lead optimization, ligand based drug design

Following the tremendous success in software development, attempts to successfully employ the open source model to other areas, including biotechnology is underway [80]. Recently, Maurer and Scotchmer reviewed the role of the emerging open source model in drug discovery [81]. According to Ardal and co researchers [82], Open source is a desirable model for drug discovery. The concept of open source has been discussed in academic environments for almost a decade. Following its application in tropical diseases [83], it

has also been implemented in Cambia's BiOS and CSIR's (Council for Scientific and Industrial Research) OSDD initiatives. The OSDD model is a unique amalgamation of open source and patenting principles [84]. Recently, it has gained more importance and appreciation in several other research activities [79]. Recently, World Health Organization's Consultative Expert Working Group has been entrusted with the evaluation of an open source drug discovery (OSDD) concept. To develop new and inexpensive drugs more quickly with wider patient reach, several OSDD has been initiated in several countries [82].

Table 3: Miscellaneous computational tools widely employed in DD process

Tool Name	Brief description of tool	Steps involved in DD process
OpenBabel [73]	Open source Chemical toolbox used primarily for converting chemical file formats	Lead optimization, virtual screening
AutoDock [74]	Molecular modeling simulation software	Molecular docking, virtual screening, molecular simulation
ArgusLab [75]	Molecular modeling, graphics and drug design program	Molecular docking, molecular simulation, ligand based drug design
VegaZZ [76]	Molecular modeling suite	Molecular modeling, molecular docking, simulation and ligand based drug design
HEX [77]	Protein docking and molecular superposition program	Molecular docking, simulation, ligand based drug design

OSDD is application of collaboration and open access concepts of open source computing in the drug discovery process. This influential model has potential for developing new medicines/diagnostics for neglected diseases [79]. In OSDD, all experimental results, wet lab as well as *in silico*, are published along with raw data to enable other experts of the domain to critically review it [82]. Case studies of Cambia and India's OSDD by Masum and coworkers [80] clearly pinpoint the high potential of OSDD in the developing world. They also believe that the open source model in drug discovery will enable huge gains.

CONCLUSION

Drug Discovery and Development process is a highly complex phenomenon. Involvement of computers and related technologies has significantly improved several stages of this lengthy process. Many of the success stories on application of CADD in recent years have demonstrated its indispensable potential. CADD approaches can provide valuable information for target identification and validation, lead selection, small-molecular screening and optimization, design of drug with minimal side effect and high potency [45]. In future, SBDD will be an integral part of drug discovery. It will definitely enable robust high throughput process based drug discovery against multiple homologous targets [35].

Applicability of "open source" concept to drug discovery phenomenon and other biotechnology areas has gained momentum in the recent years. OSDD can accelerate the drug discovery even for emerging diseases in the shortest possible times. Eventually, the availability of cheaper drugs due to reduced input costs in R & D of the drug will soon become a reality. OSDD poses major demands for developing computational drug discovery pipelines.

Several individual "open source" *in silico* tools are available for studying and analyzing chemical compounds or drug target. However, a comprehensive collection of tools targeting drug discovery process is yet to gain major attraction.

CONFLICT OF INTEREST

Nil

ACKNOWLEDGMENT

We acknowledge the Department of Biotechnology (DBT) India, Rapid Grant for Young Investigator (grant number BT/PR6476/GBD/27/496/2013) for the funding and encouragement.

CONFLICT OF INTERESTS

Declared None

REFERENCES

1. Amini S, Tavazoie S. Antibiotics and the post-genome revolution. Curr Op Microbiol 2011;14:513–8.
2. Ashrafuzzaman M. Aptamers as both drugs and drug-carriers. BioMed Res Int 2014. http://dx.doi.org/ 10.1155/ 2014/ 697923.[Articles in Press]
3. Herrling PL. The drug discovery process. Prog Drug Res 2005;62:1–14.
4. Sweeny K. Technology trends in drug discovery and development: implications for the development of the pharmaceutical industry in australia. Pharmaceutical Industry Project Equity, Sustainability and Industrial Development, Working Paper Series; 2002.
5. Prakash N, Devangi P. Drug Discovery. J Antivirals Antiretrovirals 2010;2:63-8.
6. Harvey AL. Natural products in drug discovery. Drug Discov Today 2008;13:894-901.
7. Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: a systematic review. Health Policy 2011;100:4-17.
8. Smith CG, O'Donnell JT. Overview of the current process of new drug discovery and development. In: Smith CG, O'Donnell JT, editors. The process of new drug discovery and development. Informa healthcare; 2006. p. 7-13.
9. Lim HA. Bioinformatics and cheminformatics in the drug discovery cycle. In: Hofestädt R, Lengauer T, Löffler M, Schomburg D, editors. Bioinformatics. Berlin: Springer-Verlag; 1997. p. 30–43.
10. Duggan G, Fejes A, Ingram C, Salazar M, Sirski M. Drug Discovery in the Bioinformatics age; 2000.
11. Ratti E, Trist D. Continuing evolution of the drug discovery process. Pure Appl Chem 2001;73:67-75.
12. Verpoorte R. Exploration of nature's chemodiversity: the role of secondary metabolites as leads in drug development. Drug Discov Today 1998;3:232-8.
13. Sneader W. Drug Discovery: A history. Wiley (NY); 2005.
14. Burger A. Medicinal Chemistry. Wiley (NY); 1970.
15. Walsh G. Pharmaceutical biotechnology: Concepts and applications. John Wiley (UK); 2007.
16. Congreve M, Murray CW, Blundell TL. Structural biology and drug discovery. Drug Discovery Today 2005;10:895-907.
17. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev 1997;23:3-25.
18. Boehm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, *et al*. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. J Med Chem 2000;43:2664-74.
19. Drews J. Drug discovery: a historical perspective. Sci 2000;287:1960–4.
20. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, *et al*. Initial sequencing and analysis of the human genome. Nat 2001;409:860–921.

21. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, *et al*. The sequence of the human genome. Sci 2001;291:1304–51.
22. Broder S, Venter JC. Sequencing the entire genomes of free-living organisms: the foundation of pharmacology in the new millennium. Annu Rev Pharmacol Toxicol 2000;40:97–132.
23. Ward SJ. Drug discovery and genomics technologies: Impact of genomics in drug discovery. BioTechniques 2001;31:626-30.
24. Day M, Rutkowski JL, Feuerstein GZ. Translational medicine-a paradigm shift in modern drug discovery and development: The role of biomarkers: In Guzmán CA, Feuerstein GZ, editors. Pharmaceutical Biotechnology. Landes Bioscience and Springer Science; 2009. p. 1-12.
25. Harikumar SL, Neha S. Use of genomics and proteomics in pharmaceutical drug discovery and development: a review. Int J Pharm Pharm Sci 2013;5:24-8.
26. Tang Y, Zhu W, Chen K, Jiang H. New technologies in computer-aided Drug design: Toward target identification and new chemical entity discovery. Drug discov today. Technol 2006;3:307-13.
27. Gupta D. Pharmacogenomics in drug discovery and development. J Dev Drugs 2013;2:e126.
28. Ganguly NK, Bano R, Seth SD. Human genome project: pharmacogenomics and drug development. Indian J Exp Biol 2001;39:955-61.
29. Burmester JK, Sedova M, Shapero MH, Mansfield E. DMET microarray technology for pharmacogenomics-based personalized medicine. Methods Mol Biol 2010;632:99-124.
30. Costa L, Casimiro S, Gupta S, Knapp S, Pillai MR, Toi M, *et al*. The global cancer genomics consortium's third annual symposium: from oncogenomics to cancer care. Genes Cancer 2014;5:64-70.
31. Hall SE. Chemoproteomics-driven drug discovery: addressing high attrition rates. Drug Discovery Today 2006;11:495–502.
32. Sliwoski G, Kothiwale S, Meiler J, Lowe EW. Computational methods in drug discovery. Pharmacol Rev 2014;66:334–95.
33. Saha D, Emran TB, Saha SP. Bioinformatics: The effects on the cost of drug discovery. Galle Med J 2013;18:44-50.
34. Shi L, Tong W, Lu X. Biochemoinformatics: Integrating bioinformatics and chemoinformatics for drug discovery and development. Eur Pharm Rev Drug Dis Technol 2003. Available from http://chipscreen.gotoip4.com/articleFile/lMSHI_EBR.pdf [Last accessed 18 Sep 2014].
35. Baldi A. Computational approaches for drug design and discovery: an overview. Syst Rev Pharm 2010;1:99-105.
36. Masoudi-Nejad A, Mousavian Z, Bozorgmehr JH. Drug-target and disease networks: polypharmacology in the post-genomic era. In Silico Pharm 2013;1:17.
37. Chen YP, Chen F. Identifying targets for drug discovery using Bioinformatics. Expert Opin Ther Targets 2008;12:383–9.
38. Du GH. Evaluation and validation of drug targets. Acta Pharmacol Sin 2004;25:1566.
39. Mohd A, Jyoti DA. Current trends in drug discovery: target identification to clinical development of the drug. Int Res J Pharm 2012;3:23-7.
40. Wang S, Sim TB, Kim YS, Chang YT. Tools for target identification and validation. Curr Opin Chem Biol 2004;8:371-7.
41. Bleicher KH, Bohm HJ, Muller K, Alanine AI. Hit and lead generation: beyond high-throughput screening. Nat Rev Drug Discovery 2003;2:369-78.
42. Rahman MM, Karim MR, Ahsan MQ, Khalipha ABR, Chowdhury MR, Saifuzzaman M. Use of computer in drug design and drug discovery: A review. Int J Pharm Life Sci 2012;1:1-21.
43. Kalyani G. A review on drug designing, methods, its applications and prospects. Int J Pharm Res Dev 2013;5:15-30.
44. Suryawanshi SB, Nazeruddin GM. Computer aided drug discovery and development-an important need of the hour. Int J Pharm Chem Sci 2013;2:293-96.
45. Kore PP, Mutha MM, Antre RV, Oswal RJ, Kshirsagar SS. Computer-Aided drug design: an innovative tool for modeling. Open J Med Chem 2012;2:139-48.
46. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403-10.
47. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, *et al*. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res 2013;41:W597-600.
48. Rice P, Longden I, Bleasby A. Emboss: the european molecular biology open software suite. Trends Gene 2000;16:276–7.
49. Bio Edit tool: Available from: www.mbio.ncsu.edu/bioedit/bioedit.html. [Last accessed 1 Dec 2014].
50. Rasmol tool: Available from: www.openrasmol.org. [Last accessed 1 Dec 2014].
51. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC: Available from: www.pymol.org (Last accessed 5 Dec 2014).
52. Guex N, Peitsch MC. Swiss-PdbViewer: A fast and easy-to-use PDB viewer for macintosh and PC. Protein Data Bank Quaterly Newsletter 1996;77:7.
53. Accelrys Software Inc. Discovery Studio Modeling Environment, Release 4.0, San Diego: Accelrys Software Inc; 2013.
54. Schwede T, Kopp J, Guex N, Peitsch MC. Swiss-model: an automated protein homology-modeling server. Nucleic Acids Res 2003;31:3381-5.
55. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779-815.
56. Kelley LA, Sternberg MJE. Protein structure prediction on the web: a case study using the Phyre server. Nat Protoc 2009;4:363–71.
57. PubMed: Medline retrieval on the world wide web". Fact Sheet. United States National Library of Medicine. [Last accessed 7 Jun 2002].
58. Tateno Y, Imanishi T, Miyazaki S, Fukami-Kobayashi K, Saitou N, Sugawara H, *et al*. DNA Data Bank of Japan (DDBJ) for genome scale research in life science. Nucleic Acids Res 2002;30:27–30.
59. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. "Gen Bank". Nucleic Acids Res 2012;41:D36–D42.
60. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al*. The protein data bank. Nucleic Acids Res 2000;28:235–42.
61. Kanehisa M, Goto S. Kegg: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28:27–30.
62. ISIS/Draw Tool: Available from: http://mdl-isis-draw.updatestar.com. [Last accessed 5 Dec 2014].
63. Mills N. Chem draw ultra 10.0. J Am Chem Soc 2006;128:13649–50.
64. ACD Chemsketch Tool: Available from: http://www.acdlabs.com/resources/freeware/chemsketch. [Last accessed 1 Dec 2014].
65. ChemAxon Marvinsketch Tool: Available from: http:// www.chemaxon.com/products/marvin/marvinsketch. [Last accessed 1 Dec 2014].
66. JME Molecular Editor Tool: Available from: http://www.molinspiration.com/jme [Last accessed 1 Dec 2014].
67. ISIS/Base Database: Available from: http://mdl-information-system-isis-base.software.informer.com/2.4/. [Last accessed 15 Dec 2014].
68. ACD/Chemfolder Software: Available from: http://www.acdlabs.com/products/km/ackm/chemfolder. [Last accessed 15 Dec 2014].
69. Chemspider Database: Available from http://www. chemspider.com. [Last accessed 15 Dec 2014].
70. NCBI PubChem Database: Available from: https:// pubchem.ncbi.nlm.nih.gov. [Last accessed 15 Dec 2014].
71. Cambridge structural database (CSD): Available from: www.ccdc.cam.ac.uk/products/csd. [Last accessed 15 Dec 2014].
72. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, *et al*. Chembl: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2012;40:D1100-D1107.
73. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open babel: an open chemical toolbox. J Chem 2011;3:33.
74. AutoDock Software: Available from: http://mgltools.scripps. edu/downloads [Last accessed 15 Dec 2014].
75. ArgusLab Software: Available from: http://arguslab.en. softonic.com/download. [Last accessed 15 Dec 2014].
76. DDL VEGAZZ Software: Available from: http:// nova.disfarm.unimi.it/cms/index.php?Software_projects:VEGA_ZZ. [Last accessed 15 Dec 2014].

77. HEX Software: Available from: http://hex.loria.fr. [Last accessed 15 Dec 2014].
78. Amant KS, Still B. editors. Handbook of research on open source software: Technological, economic, and social perspectives. 1st Ed. Hershey: IGI Global; 2007.
79. Ardal C, Rottingen JA. Open source drug discovery in practice: a case study. PloS Negl Trop Dis 2012;6:e1827.
80. Masum H, Schroeder K, Khan M, Daar A. Open source biotechnology platforms for global health and development: two case studies. Inf Tech Int Dev J 2011;7:61–9.
81. Maurer S, Scotchmer S. Open source software: The new intellectual property paradigm. In: Amsterdam WA. editor. Handbook of Economics and Information Systems. Amsterdam: Elsevier; 2006. p. 285-322.
82. Ardal C, Alstadsæter A, Rottingen JA. Common characteristics of open source software development and applicability for drug discovery: a systematic review. Health Res Policy Syst 2011;9:36.
83. Maurer SM, Rai A, Sali A. Finding cures for tropical diseases: Is open source an answer? PLoS Med 2004;1:56.
84. Sugumaran G. Open source drug discovery–redefining IPR through open source innovations. Curr Sci 2012;102:1637-9.