ABSTRACT
The drug discovery process has been revolutionized with the advent of numerous advanced techniques. Pharmaceutical researchers work to piece together the basic underlying causes of disease at the molecular level to select a potential ‘target’ which is then ‘validated’ to be involved in the disease and can be manipulated by a drug. The researchers then search for a ‘lead compound’ that can act on their target to alter the disease. Collaboration between pharma industry and academia is an essential component in the modern drug discovery program. The main objective of this article is to highlight the recent advances and trends in drug discovery.

INTRODUCTION
The pharmaceutical industry is constantly striving to understand diseases and bring forth safe and efficient drugs into the market to ensure their financial and brand status. Low productivity, expiry of patents, rising R&D costs, and dwindling pipelines are the major challenges threatening the pharma industry today [1]. Outsourcing to the developing nations, mergers and acquisitions and collaboration are the major strategies being employed today to strike a balance between rising development costs and profits. In the past 10-15 years, a lot of changes have taken place in the pharmaceutical industry. Industries are merging with each other to sustain their financial aspects as shown in Table 1. Recent advances in the fields of computational methods, genomics, metabolomics, chemogenomics, and proteomics help in cutting down the expenditure and time scales to improve the quality in the drug discovery and development process.

The global sales generated for pharmaceutical products in 2010 were 856 billion dollars with approximately 60% of these sales in the US and Europe [2]. Novel biological entities (NBEs) constituted approximately 30% of drugs approved during the past two years. The generic products share approximately 10% of the global pharma sales and are expected to rise to approximately 14% by 2015 due to projected patent expiry of proprietary products. In 2011, 35 drug approvals related to rare and/or orphan diseases were granted by the FDA. New molecular entities (NMEs) such as drug combinations, new drug delivery options, enantiomers and polymorphs are being investigated. Several approaches such as phenotypic screening (PS), drug repositioning approach (DRA), high throughput screening (HTS), molecular design (MD), and matching molecular mechanism (MMM) of the ‘off-target’ effects either alone or in collaborative mode is being adopted to identify new use for currently marketed or discontinued drugs.

Drug repositioning or drug repurposing [3], the process of developing new indications for existing pro-drugs, drugs and biologics has been adopted by pharmaceutical companies in order to accelerate the drug discovery and development process and reduce costs and risks of failure. These drugs could directly enter into Phase II and III clinical studies as the bioavailability and safety profiles of drugs are well-known from the preclinical and Phase I studies. For instance, Thalidomide (morning sickness) has received FDA approval for the treatment of multiple myeloma through drug repurposing. Repurposed drugs have recorded a 300 percent increase from around 80 in 2001 to 222 by 2010.

![Figure 1. Risk-reward balance portfolio for various therapeutics.](image-url)
Traditionally, the pharmaceutical industry has operated mainly in the low-risk territory on high incident diseases which generate more revenue whereas academia was involved in rare disorders with high risk, low reward and smaller revenues [4]. The academia has been restricted by the non-availability of advanced technology to manipulate the highly refractory targets. The risk-reward balance portfolio for various therapeutics is shown in Figure 1. The pharma industry is patenting more than ever in efforts to protect its intellectual property and the academia is also patenting vigorously to generate revenue streams (Table 2).

The interdependence of academia, disease foundations and pharma is increasing in the new era of drug discovery [5]. The research at the academic sector contributes to the identification of novel biological targets and disease validation while the pharmaceutical industry develops drugs by high-throughput screening of compounds against these biological targets and optimizing the chemistry involved in the synthesis [6].

**DRUG DISCOVERY PIPELINE**

The process by which a new drug is brought to market stage is referred to as the development pipeline. The drug discovery pipeline involves identification, validation of a disease target followed by the development of a chemical compound to interact selectively with that target. On an average, US$ 1.8 billion and 10-15 years are required to develop a successful drug [8]. Only one among every 5,000-10,000 compounds entering the R&D pipeline receives approval. The drug discovery pipeline is shown in Figure 2.

**Target identification and validation**

The targets responsible for a particular disease may be nucleic acids, hormones, transport proteins, enzymes, ion channels, etc. Recent technologies like gene expression profiling (transcriptomics), protein expression profiling (proteomics), metabolic pathways (metabolomics), protein glycosylation (glycosilomics), protein-protein interactions (interactomics), phenotype screening, data mining in silico, in vivo methods using genetic engineering and somatic mutagenesis using RNAi technologies are used to identify novel targets [9]. RNA editing [10], the change in the nucleotide sequence of RNA transcripts relative to that of the template DNA, is a promising technology which provides additional drug targets for drug discovery. Targeting signaling pathways (cancer hallmarks) has had a significant impact on drug discovery and development, especially in the treatment of cancer [11]. Currently marketed drugs, such as imatinib mesylate (Gleevec; Novartis) and dasatinib (Sprycel, Bristol Myers Squibb) modulate the aberrant signaling pathways linked with diseases. Subsequently, siRNA [12] and chemogenomic [13] approaches have been extensively employed in both academic labs and pharmaceutical industries to identify novel signaling pathways for cancer.
The molecular target-based approach for drug discovery (‘reverse pharmacology’ or ‘reverse chemical biology’) begins with the relevant target identification and validation, assay development, followed by identification of hits by HTS of chemical libraries against the target (Figure 3a). The hits which are validated in orthogonal assays are then characterized by the structure-activity relationships (SAR) and modified to develop favorable ADME properties. Molecular target-based screening has some distinct advantages over phenotypic screening; knowledge about a molecular target and its related screening assay are crucial for the optimization of lead compounds, SAR, toxicology studies and biomarker development.

In contrast, in the phenotypic screening [14] approach for drug discovery (‘forward pharmacology’, ‘classical pharmacology’ or ‘forward chemical biology’), the activity and efficacy of a drug are determined even before any knowledge of the target and the molecular mechanism are known (Figure 3b). These approaches involving cellular assay systems based on a characteristic associated with the disease, is usually more physiologically relevant than an in vitro screen since intact cells and their indigenous environment are used and enables lead discovery for many rare diseases in which a drug target has not been identified and/or validated. Cell-based phenotypic screens [15] utilizing primary cells and stem cell derived human cells have recently emerged for lead discovery in early drug discovery in parallel to the molecular target-based screening approach. Validating targets is usually performed after the disease has been established and can be achieved through conditional target gene knockout technology, specific agonists and/or antagonists that act at the transcriptional level (e.g. anti-sense oligonucleotides), the posttranscriptional level (e.g. RNAi) or the protein level (e.g. receptor antagonists, antibodies and aptamers). Microfluidics technology is emerging as a promising application in the drug discovery and development process mainly in genomics and proteomics. It is being increasingly employed in lead synthesis, target identification, crystallization, High-Throughput screening, drug ADME and toxicity studies [16]. This technology is utilized in proteomics for enzymatic assays, immunoassays and peptide mass fingerprinting [17]. Novel organ-on-a-chip [18] platforms fabricated using microfluidics and microfabrication technologies, can be utilized for developing disease models, high-throughput screening and drug testing.

In the recent years, bioinformatics has been exploited extensively in the identification of potential targets for various diseases. Computer-aided drug design (CADD) is a specialized discipline that uses computational methods to simulate drug-receptor interactions. When the crystal structure for a protein is not available, homology modeling tool is used to predict the probable structure. Molecular docking and scoring techniques involve computationally placing a virtual molecular structure into a binding site of a biological macromolecule.

**Hits to leads selection**

Identifying potent molecules that bind selectively to a biological target remains an elusive challenge in the drug discovery field in spite of the significant technological advances in the compound library preparation in the last two decades. Traditionally, there are two classes of compound libraries: (i) natural products and (ii) synthetic compounds. Historically, most of the leads were obtained from nature; because nature only can create and inspire complex bioactive molecules. Natural products constitute for over 28% of the new chemical entities and 42% of the anticancer drugs introduced into the market [19]. Microbes, plants, and marine organisms were extensively exploited to identify interesting compounds for fighting diseases. Over 15,000 natural products with antimicrobial, anti-tumor, anti-inflammatory and anti-cardiovascular activities have been identified from marine microbes and over 30 compounds such as didemnin B (Aplidine™) are currently in clinical or preclinical studies for the treatment of cancer [20]. Generally, natural products display greater structural diversity and complexity when compared to synthetic molecules; whereas synthetic compound libraries have the advantage of high-purity and well-characterized structures. Ideally, a combinatorial library should exhibit elements of complexity of natural products and the well-characterized structural format of synthetic compound libraries. Most commonly, hit compounds are derived by High-Throughput Screening (HTS) [21]. Typical HTS programs have potentials to screen up to 10000 compounds per day, while some laboratories with Ultra High-Throughput Screening (UHTS) systems can perform 100,000 assays per day. A large number of hits can be synthesized by combinatorial chemistry and screened for their biological activity against a target. Quantitative Structure Activity Relationships (QSAR) constitutes immense importance in discovering new drug candidates which shows high affinity with the target. Recent advances in the field of chemistry have enabled scientists to synthesize compounds from scratch (De novo).
Combinatorial chemistry permits medicinal chemists to assemble a large library of compounds in a relatively short time and in a multitude of combinations. Diversity-oriented synthesis (DOS) [24], is similar to combinatorial chemistry, but the order in which diversity is introduced is reversed. In a combinatorial library, diversification is achieved through tandem reactions where different appendages are added on different core scaffolds to access different parts of chemical space.

High-content screening (HCS) [22] approaches are a series of multiparametric approaches at the single-cell level that are employed for increasing throughput in library screening, developing assays, mechanism of action studies, target identification and validation. HCS has been increasingly adopted in the drug discovery field, 61% of the HTS laboratories incorporated HCS in their operations in 2012, compared to 40% in 2007. In 2012, the peptides lucinactant, peginesatide, pasireotide, carfilzomib, linaclotide, and teduglutide received marketing approvals as new molecular entities. All six were approved in the USA and five of the six except peginesatide were also approved in the European Union (EU). In 2013, lixisenatide has been approved, but peginesatide approved in 2012 was withdrawn because of safety issues [23].

High-throughput synthesis techniques

Combinatorial chemistry permits medicinal chemists to assemble a large library of compounds in a relatively short time and in a multitude of combinations. Diversity-oriented synthesis (DOS) [24], is similar to combinatorial chemistry, but the order in which diversity is introduced is reversed. In a combinatorial library, diversification is achieved through varying the appendages of a common core scaffold (Figure 4a). Conversely, in DOS the library has similar appendages on different core scaffolds (Figure 4b), thus covering more chemical space. Biology-oriented synthesis (BIOS) [25], involves the generation of compound libraries based upon iterations similar to scaffolds of known biological activities—often natural products.

In many cases, the phase-tagging technologies and solid-supports have been successfully employed in traditional library synthesis techniques, such as parallel and split-pool synthesis. To overcome the challenges associated with compatibility of reactants and the heterogeneous nature in the solid-phase organic synthesis (SPOS), the liquid-phase organic synthesis (LPOS) [26], which employ precipitation tags, has been developed. Complementary to LPOS, fluorous tags have also been used in a combinatorial format by tagging the starting material with a polyfluorocarbon chain. Microwave-assisted organic synthesis (MAOS) [27] has achieved higher rate accelerations and different product profiles compared to traditional thermal heating methods.

Receptor-assisted combinatorial chemistry (RCC) [28] combines synthesis and screening procedures in one step. Two major RCC methods that have emerged in the last decade are dynamic combinatorial libraries and receptor-accelerated synthesis (RAS). In dynamic combinatorial chemistry, the synthesis of library members is reversible thermodynamically. Thus, the binding of a given library member to the receptor alters the equilibrium of the reaction mixture, which allows its identification. In RAS, two building blocks in a library bind with a given receptor in close proximity to each other, thus increasing their effective molarity. This leads to formation of a covalent bond between the two blocks and generates the inhibitor molecule.

Fragment-based lead generation [29] from less complex and simple molecules offer efficient sampling of chemical space and is also useful in areas such as druggability assessments [30] and HTS evaluations. Moreover, in the case of intractable targets for which HTS methods fail to yield a suitable hit, fragment approaches have been found to be successful. The assessment of druggability of a target can be done by specialized fragment screening approaches, thereby reducing the risk of failure in early projects.

In continuous-flow chemistry [31], a chemical reaction is performed in a continuously flowing stream in a network of interconnecting tubes. Flow chemistry can be used with other technologies such as polymer-assisted solution-phase synthesis (PASPS) [32] and microwave heating. Owing to the possibility of producing a large number of compounds in high purity and in short time, microreactors [33] i.e. flow reactors of micrometer scale have been employed successfully in combinatorial chemistry. The concept can be extended further to include high-throughput assays in the flow system, in order to speed up the entire hit identification and lead optimization process. Advances in robotics and computational power allow researchers in high-throughput screening to screen hundreds of thousands of compounds against a target. Biotechnology can be utilized to produce disease-fighting biological molecules from genetically engineered living systems [34].

High-performance liquid chromatography combined with mass spectrometry (HPLC-MS) [35] is the technique of choice for most assays used in various stages in drug discovery. Advanced technologies, such as ultra performance liquid chromatography (UPLC) [36] and supercritical fluid chromatography (SFC) [37], contributed to enhance analytical efficiency.
In silico Methods in Drug Discovery

The role of computational chemists has evolved from a supporting to that of a mainstream role in the drug discovery arena. When a library of drug-like molecules and the atomic coordinates of the target protein are available, computational methods like molecular docking, virtual screening (VS) and pharmacophore-based virtual screening can be applied. These in silico approaches reduce the lead-discovery time by sifting through large virtual libraries. Molecular docking involves the determination of binding affinity between the receptor and ligand. ‘Receptor based docking’ uses the 3D structure of the receptor to dock each compound from the chemical database in the active site, rank and score them according to the binding affinity. ‘Ligand based docking’ uses methods such as similarity searching, substructure searching, 3D shape matching or pharmacophore matching to identify compounds similar to known inhibitors from chemical databases.

Over the past decade, the evolution of VS strategies is evident by an increasing degree of their integration into the discovery process and their complementarity to high-throughput screening (HTS) methods. A novel classification of VS applications according to its level of integration has been proposed [38]. Parallel VS strategy involves the parallel application of complementary methods like 2D, 3D, ligand-based, structure-based, similarity searching, and molecular modeling and combines the results. The combination of multiple results increases the number of true positives and decreases the number of false positives and thus, helps to improve the enrichment rates. In Iterative VS, VS is sequentially integrated iteratively into the hit identification and hit-to-lead optimization processes. The in vitro screening data flows back into VS and helps to develop the in silico model. Integrated VS involves complete integration of computational techniques with HTS. This strategy is used for targeting vast virtual chemical libraries that are not yet synthesized or commercially available libraries. QSAR (Quantitative structure-activity relationships) correlate the structure of a compound with its biological activity. Various QSAR like 1D QSAR to 6D QSAR are in use based on the data dimensions. The chemical descriptors commonly used for SAR correlation include molecular weight, number of Hydrogen bond donors or acceptors, number of rotatable bonds, etc. Pharmacophore mapping involves the generation of a 3D pharmacophore, a set of molecular features such as Hydrogen bond donors and acceptors, positively and negatively charged groups, hydrophobic regions and aromatic rings and their relative spatial orientation that are essential for interaction and biological activity against a particular receptor. A chemical library can then be searched for members that match these molecular features, which indicates the potential for similar activity.

Lead Optimization

Lead compounds that survive the initial screening are then optimized to make them more potent, selective, safer, metabolically stable and suitable for testing in a clinical setting. This optimization is accomplished through chemical modification of the hit structure by employing structure-activity analysis (SAR) as well as structure-based design if structural information about the target is available. Lead optimization is one of the most-time consuming stages in the drug discovery process, yet it is the most important stage where efforts are put into combining high target potency with high drug efficacy to reduce the active dose. This is a labour intensive process where biologists and chemists collaborate: The biologists test the effects of analogues on biological systems while the chemists take this information to make additional alterations that are then retested by the biologists. Lead optimization involves multiparameter optimization of potency against a target, efficacy, PK/PD, selectivity, activity in an animal model, cellular and toxicity assays, such as P450 inhibition assays, cytotoxicity assays, hERG safety testing, metabolic profiling, etc. Process chemistry, scale up, preformulation and formulation studies are also carried out. The challenge is to reduce timelines for candidate selection while ensuring the highest quality candidates to minimize the expensive late-stage attrition.

DRUG DEVELOPMENT

The drug discovery process is followed by the drug developmental process. The study of the promising drug candidates are carried out in two stages: preclinical pharmacology (animal studies) and clinical pharmacology (human studies).

Preclinical pharmacology (animal studies)

The candidate drug is subjected to extensive pharmacological testing in vitro and in vivo on animal models (mice, rats, pigs, dogs). Major areas of research are:

1. Acute, sub-acute and chronic toxicity studies (toxicity profile)
2. Therapeutic Index (safety and efficacy evolution): it is the ratio of median lethal dose (LD50) for a drug to the median effective dose (ED50).
3. Absorption, distribution.

The safety issues which are a major challenge in the drug discovery are believed to be overcome by the right balance of in vivo, in vitro and computational toxicology [39] predictions applied as early as possible in the discovery process. It can be more predictive than results from animal studies when there are significant genetic differences between human and rodent. The algorithms that can predict the side effects from chemical structure can be put into two classes: expert systems and statistical modeling. Expert systems, such as Oncologic or Derek are a repository of expert knowledge. Statistical modeling software - such as Topkat, PASS, TPS-SVM and Multicase aims to analyze existing data and automatically build models. Gene knockout technology has provided evidence for expected phenotypes and helps in investment in a particular target.

CONCLUSION

Despite all the advances, the drug discovery process still remains an expensive, time-consuming and inefficient process with low rate of new therapeutic discovery. As quoted by Sharpless and co-workers ‘the most fundamental and lasting objective of synthesis is not production of new compounds, but production of properties.’ Thus, the goal of medicinal chemists need not solely be the synthesis of new compounds but those that possess the required biological property. The right science and the right technologies that are available could be utilized in a right manner at the right time to develop the right products for the right patients.

ACKNOWLEDGEMENT

Authors are thankful to Department of Pharmaceuticals, Ministry of Chemicals and Fertilizers, Government of India for financial support to the Research Scholar, Shalini Nekkanti.
REFERENCES