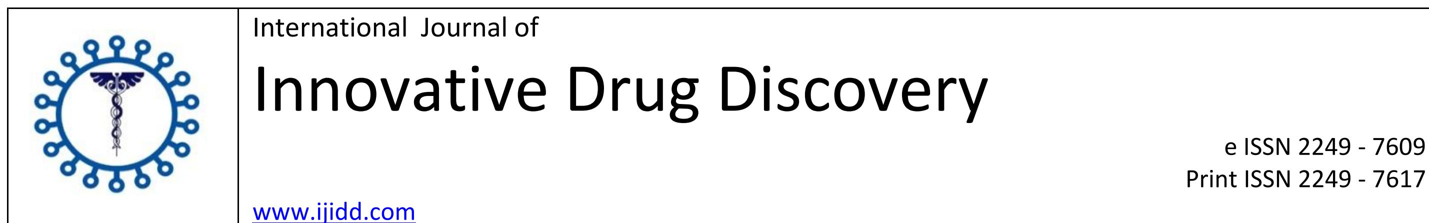
R. Ashok Raj. / Vol 2/ Issue 1/ 2012/ 22-39.



A RECENT TECHNOLOGY IN DRUG DISCOVERY AND DEVELOPMENT

*R. Ashok Raj

*Adamas India Pharmaceuticals Private Ltd., Bangalore, India- 560017.

ABSTRACT

The discovery and development of new medicines is a long, complicated process. Each success is built on many, many prior failures. Advances in understanding human biology and disease are opening up exciting new possibilities for breakthrough medicines. At the same time, researchers face great challenges in understanding and applying these advances to the treatment of disease. As our understanding of disease has increased to the extent that we know how disease and infection are controlled at the molecular and physiological level, scientists are now able to try to find compounds that specifically modulate those molecules, for instance via high throughput screening. Even more recently, scientists have been able to understand the shape of biological molecules at the atomic level, and to use that knowledge to design drug candidates.

Keywords: Drug Discovery, Drug Design, High Throughput Screening, High Content Screening

INTRODUCTION

In the past most drugs have been discovered either by identifying the active ingredient from traditional remedies or by serendipitous discovery. As our understanding of disease has increased to the extent that we know how disease and infection are controlled at the molecular and physiological level, scientists are now able to try to find compounds that specifically modulate those molecules, for instance via high throughput screening. Even more recently, scientists have been able to understand the shape of biological molecules at the atomic level, and to use that knowledge to design drug candidates.

The process of drug discovery involves the identification of candidates, synthesis, characterization, screening, and assays for therapeutic efficacy. Once a compound has shown its value in these tests, it will begin the process of drug development prior to clinical trials. Despite advances in technology and understanding of biological systems, drug discovery is still a lengthy, "expensive, difficult, and inefficient process" with low rate of new therapeutic discovery. Currently, the research and development cost of each new molecular entity (NME) is approximately US$1.8 billion.

Information on the human genome, its sequence and what it encodes has been hailed as a potential windfall for drug discovery, promising to virtually eliminate the bottleneck in therapeutic targets that has been one limiting factor on the rate of therapeutic discovery.[1] However, data indicates that "new targets" as opposed to "established targets" are more prone to drug discovery project failure in general. This data corroborates some thinking underlying a pharmaceutical industry trend beginning at the turn of the twenty-first century and continuing today which finds more risk aversion in target selection among multi-national pharmaceutical companies [1].

Corresponding Author:- **R. Ashok Raj** Email:- revant08@gmail.com

DRUG TARGETS

The definition of "target" itself is something argued within the pharmaceutical industry. Generally, the "target" is the naturally existing cellular or molecular structure involved in the pathology of interest that the drug-in-development is meant to act on. However, the distinction between a "new" and "established" target can be made without a full understanding of just what a "target" is. This distinction is typically made by pharmaceutical companies engaged in discovery and development of therapeutics.

"Established targets" are those for which there is a good scientific understanding, supported by a lengthy publication history, of both how the target functions in normal physiology and how it is involved in human pathology. This does not imply that the mechanism of action of drugs that are thought to act through a particular established targets is fully understood. Rather, "established" relates directly to the amount of background information available on a target, in particular functional information. The more such information is available, the less investment is (generally) required to develop a therapeutic directed against the target. The process of gathering such functional information is called "target validation" in pharmaceutical industry parlance. Established targets also include those that the pharmaceutical industry has had experience mounting drug discovery campaigns against in the past; such a history provides information on the chemical feasibility of developing a small molecular therapeutic against the target and can provide licensing opportunities and freedom-to-operate indicators with respect to small-molecule therapeutic candidates [2].

In general, "new targets" are all those targets that are not "established targets" but which have been or are the subject of drug discovery campaigns. These typically include newly discovered proteins, or proteins whose function has now become clear as a result of basic scientific research.

The majority of targets currently selected for drug discovery efforts are proteins. Two classes predominate: G-protein-coupled receptors (or GPCRs) and protein kinases.

SCREENING AND DESIGN

The process of finding a new drug against a chosen target for a particular disease usually involves high-throughput screening (HTS), wherein large libraries of chemicals are tested for their ability to modify the target. For example, if the target is a novel GPCR, compounds will be screened for their ability to inhibit or stimulate that receptor (see antagonist and agonist): if the target is a protein kinase, the chemicals will be tested for their ability to inhibit that kinase.

Another important function of HTS is to show how selective the compounds are for the chosen target. The ideal is to find a molecule which will interfere with only the chosen target, but not other, related targets. To this end, other screening runs will be made to see whether the "hits" against the chosen target will interfere with other related targets - this is the process of cross-screening. Cross-screening is important, because the more unrelated targets a compound hits, the more likely that off-target toxicity will occur with that compound once it reaches the clinic [3].

It is very unlikely that a perfect drug candidate will emerge from these early screening runs. It is more often observed that several compounds are found to have some degree of activity, and if these compounds share common chemical features, one or more pharmacophores can then be developed. At this point, medicinal chemists will attempt to use structure-activity relationships (SAR) to improve certain features of the lead compound:

- increase activity against the chosen target
- reduce activity against unrelated targets
- improve the druglikeness or ADME properties of the molecule.

This process will require several iterative screening runs, during which, it is hoped, the properties of the new molecular entities will improve, and allow the favoured compounds to go forward to in vitro and in vivo testing for activity in the disease model of choice.

Amongst the physico-chemical properties associated with drug absorption include ionization (pKa), and solubility; permeability can be determined by PAMPA and Caco-2. PAMPA is attractive as an early screen due to the low consumption of drug and the low cost compared to tests such as Caco-2, gastrointestinal tract (GIT) and Blood-brain barrier (BBB) with which there is a high correlation.

A range of parameters can be used to assess the quality of a compound, or a series of compounds, as proposed in the Lipinski's Rule of Five. Such parameters include calculated properties such as cLogP to estimate lipophilicity, molecular weight, polar surface area and measured properties, such as potency, in-vitro measurement of enzymatic clearance etc. Some descriptors such as ligand efficiency (LE) and lipophilic efficiency (LiPE) combine such parameters to assess druglikeness.

While HTS is a commonly used method for novel drug discovery, it is not the only method. It is often possible to start from a molecule which already has some of the desired properties. Such a molecule might be extracted from a natural product or even be a drug on the market which could be improved upon (so-called "me too" drugs). Other methods, [4] such as virtual high throughput screening,

where screening is done using computer-generated models and attempting to "dock" virtual libraries to a target, are also often used.

Another important method for drug discovery is drug design, whereby the biological and physical properties of the target are studied, and a prediction is made of the sorts of chemicals that might (eg.) fit into an active site. One example is fragment-based lead discovery (FBLD). Novel pharmacophores can emerge very rapidly from these exercises. In general, computer-aided drug design is often but not always used to try to improve the potency and properties of new drug leads.

Once a lead compound series has been established with sufficient target potency and selectivity and favourable drug-like properties, one or two compounds will then be proposed for drug development. The best of these is generally called the lead compound, while the other will be designated as the "backup" [5].

Historical background

The idea that effect of drug in human body are mediated by specific interactions of the drug molecule with biological macromolecules, (proteins or nucleic acids in most cases) led scientists to the conclusion that individual chemicals are required for the biological activity of the drug. This made for the beginning of the modern era in pharmacology, as pure chemicals, instead of crude extracts, became the standard drugs [6]. Examples of drug compounds isolated from crude preparations are morphine, the active agent in opium, and digoxin, a heart stimulant originating from Digitalis lanata. Organic chemistry also led to the synthesis of many of the cochemicals isolated from biological sources.

Nature as source of drugs

Despite the rise of combinatorial chemistry as an integral part of lead discovery process, natural products still play a major role as starting material for drug discovery. A report was published in 2007, covering years 1981-2006 details the contribution of biologically occurring chemicals in drug development. According to this report, of the 974 small molecule new chemical entities, 63% were natural derived or semisynthetic derivatives of natural products. For certain therapy areas, such as antimicrobials, antineoplastics, antihypertensive and anti-inflammatory drugs, the numbers were higher.

Natural products may be useful as a source of novel chemical structures for modern techniques of development of antibacterial therapies.

Despite the implied potential, only a fraction of Earth’s living species has been tested for bioactivity.

Plant-derived

Prior to Paracelsus, the vast majority of traditionally used crude drugs in Western medicine were plant-derived extracts. This has resulted in a pool of information about the potential of plant species as an important source of starting material for drug discovery. A different set of metabolites is sometimes produced in the different anatomical parts of the plant (e.g. root, leaves and flower), and botanical knowledge is crucial also for the correct identification of bioactive plant materials.

Microbial metabolites

Microbes compete for living space and nutrients. To survive in these conditions, many microbes have developed abilities to prevent competing species from proliferating. Microbes are the main source of antimicrobial drugs. Streptomyces species have been a source of antibiotics. The classical example of an antibiotic discovered as a defense mechanism against another microbe is the discovery of penicillin in bacterial cultures contaminated by Penicillium fungi in 1928.

Marine invertebrates

Marine environments are potential sources for new bioactive agents. Arabinose nucleosides discovered from marine invertebates in 1950s, demonstrating for the first time that sugar moieties other than ribose and deoxyribose can yield bioactive nucleoside structures. However, it was 2004 when the first marine-derived drug was approved. The cone snail toxin ziconotide, also known as Prialt, was approved by the Food and Drug Administration to treat severe neuropathic pain. Several other marine-derived agents are now in clinical trials for indications such as cancer, anti-inflammatory use and pain. One class of these agents are bryostatin-like compounds,under investigation as anti-cancer therapy.

Chemical diversity of natural products

As above mentioned, combinatorial chemistry was a key technology enabling the efficient generation of large screening libraries for the needs of high-throughput screening. However, now, after two decades of combinatorial chemistry, it has been pointed out that despite the increased efficiency in chemical synthesis, no increase in lead or drug candidates has been reached. This has led to analysis of chemical characteristics of combinatorial chemistry products, compared to existing drugs or natural products. The chemoinformatics concept chemical diversity, depicted as distribution of compounds in the chemical space based on their physicochemical characteristics, is often used to describe the difference between the combinatorial chemistry libraries and natural products [7]. The synthetic, combinatorial library compounds seem to cover only a limited and quite uniform chemical space, whereas existing drugs and particularly natural products, exhibit much greater chemical diversity, distributing more evenly to the

chemical space. The most prominent differences between natural products and compounds in combinatorial chemistry libraries is the number of chiral centers (much higher in natural compounds), structure rigidity (higher in natural compounds) and number of aromatic moieties (higher in combinatorial chemistry libraries). Other chemical differences between these two groups include the nature of heteroatoms (O and N enriched in natural products, and S and halogen atoms more often present in synthetic compounds), as well as level of non-aromatic unsaturation (higher in natural products). As both structure rigidity and chirality are both well-established factors in medicinal chemistry known to enhance compounds specificity and efficacy as a drug, it has been suggested that natural products compare favourable to today's combinatorial chemistry libraries as potential lead molecules [8].

NATURAL PRODUCT DRUG DISCOVERY

Screening

Two main approaches exist for the finding of new bioactive chemical entities from natural sources: random collection and screening of material; and exploitation of ethnopharmacological knowledge in the selection. The former approach is based on the fact that only a small part of Earth's biodiversity has ever been tested for pharmaceutical activity, and organisms living in a species-rich environment need to evolve defensive and competitive mechanisms to survive. A collection of plant, animal and microbial samples from rich ecosystems might give rise to novel biological activities. One example of a successful use of this strategy is the screening for antitumour agents by the National Cancer Institute, started in the 1960s. Paclitaxel was identified from Pacific yew tree Taxus brevifolia. Paclitaxel showed anti-tumour activity by a previously undescribed mechanism (stabilization of microtubules) and is now approved for clinical use for the treatment of lung, breast and ovarian cancer, as well as for Kaposi's sarcoma [9].

Ethnobotany is the study of the use of plants in the society, and ethnopharmacology, an area inside ethnobotany, is focused on medicinal use of plants. Both can be used in selecting starting materials for future drugs. Artemisinin, an antimalarial agent from sweet wormtree Artemisia annua, used in Chinese medicine since 200BC is one drug used as part of combination therapy for multiresistant Plasmodium falciparum [9].

HIGH-THROUGHPUT SCREENING

Structural elucidation

The elucidation of the chemical structure is critical to avoid the re-discovery of a chemical agent that is already known for its structure and chemical activity. Mass spectrometry, often used to determine structure, is a method in which individual compounds are identified based on their mass/charge ratio, after ionization. Chemical compounds exist in nature as mixtures, so the combination of liquid chromatography and mass spectrometry (LC-MS) is often used to separate the individual chemicals. Databases of mass spectras for known compounds are available. Nuclear magnetic resonance spectroscopy is another important technique for determining chemical structures of natural products. NMR yields information about individual hydrogen and carbon atoms in the structure, allowing detailed reconstruction of the molecule's architecture [10].

Early drug discovery involves several phases from target identification to preclinical development. 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. The Hit-to-Lead phase is usually the follow-up of high-throughput screening (HTS). It includes the following steps:

- Hit confirmation
- Hit expansion
- Lead optimization phase

Hit confirmation

The Hit confirmation phase will be performed during several weeks as follows:

- Re-testing: compounds that were found active against the selected target are re-tested using the same assay conditions used during the HTS.
- Dose response curve generation: several compound concentrations are tested using the same assay, an IC50 or EC50 value is then generated. Methods are being developed that may allow the reuse of the compound that generated the hit in the initial HTS step. These molecules are removed from beads and transferred to a microarray for quantitative assessment of binding affinities in a "seamless" approach that could allow for the investigation of more hits and larger libraries.
- Orthogonal testing: Confirmed hits are assayed using a different assay which is usually closer to the target physiological condition or using a different technology.
- Secondary screening: Confirmed hits are tested in a functional assay or in a cellular environment. Membrane permeability is usually a critical parameter.
- Chemical amenability: Medicinal chemists will evaluate compounds according to their synthesis feasibility and other parameters such as up-scaling or costs
- Intellectual property evaluation: Hit compound structures are quickly checked in specialized databases to define patentability
- Biophysical testing: Nuclear magnetic resonance (NMR), Isothermal Titration Calorimetry, dynamic light scattering, surface plasmon resonance, dual polarisation interferometry, microscale thermophoresis (MST) are commonly used to assess whether the

R. Ashok Raj. / Vol 2/ Issue 1/ 2012/ 22-39.

compound binds effectively to the target, the stoïchiometry of binding, any associated conformational change and to identify promiscuous inhibitors.
- Hit ranking and clustering: Confirmed hit compounds are then ranked according to the various hit confirmation experiments [11].

Hit expansion

Following hit confirmation, several compound clusters will be chosen according to their characteristics in the previously defined tests. An Ideal compound cluster will:

- have compound members that exhibit a high affinity towards the target (less than 1 μM)
- Moderate molecular weight and lipophilicity (usually measured as cLogP). Affinity, molecular weight and lipophilicity can be linked in single parameter such as ligand efficiency and lipophilic efficiency to assess druglikeness
- show chemical tractability
- be free of Intellectual property
- not interfere with the P450 enzymes nor with the P-glycoproteins
- not bind to human serum albumin
- be soluble in water (above 100 μM)
- be stable
- have a good druglikeness
- exhibit cell membrane permeability
- show significant biological activity in a cellular assay
- not exhibit cytotoxicity
- not be metabolized rapidly
- show selectivity versus other related targets

The project team will usually select between three and six compound series to be further explored. The next step will allow to test analogous compounds to define Quantitative structure-activity relationship (QSAR). Analogs can be quickly selected from an internal library or purchased from commercially available sources. Medicinal chemists will also start synthesizing related compounds using different methods such as combinatorial chemistry, high-throughput chemistry or more classical organic chemistry synthesis [12].

Lead optimization phase

The objective of this drug discovery phase is to synthesize lead compounds, new analogs with improved potency, reduced off-target activities, and physiochemical/metabolic properties suggestive of reasonable in vivo pharmacokinetics. This optimization is accomplished through chemical modification of the hit structure, with modifications chosen by employing structure-activity analysis (SAR) as well as structure-based design if structural information about the target is available [13].

High-throughput screening (HTS) is a method for scientific experimentation especially used in drug discovery and relevant to the fields of biology and chemistry. Using robotics, data processing and control software, liquid handling devices, and sensitive detectors, High-Throughput Screening allows a researcher to quickly conduct millions of chemical, genetic or pharmacological tests. Through this process one can rapidly identify active compounds, antibodies or genes which modulate a particular biomolecular pathway. The results of these experiments provide starting points for drug design and for understanding the interaction or role of a particular biochemical process in biology [14].

Assay plate preparation

The key labware or testing vessel of HTS is the microtiter plate: a small container, usually disposable and made of plastic, that features a grid of small, open divots called *wells*. Modern microplates for HTS generally have either 384, 1536, or 3456 wells. These are all multiples of 96, reflecting the original 96 well microplate with 8 x 12 9mm spaced wells. Most of the wells contain experimentally useful matter, often an aqueous solution of dimethyl sulfoxide (DMSO) and some other chemical compound, the latter of which is different for each well across the plate. (The other wells may be empty, intended for use as optional experimental controls.)

A screening facility typically holds a library of *stock plates*, whose contents are carefully catalogued, and each of which may have been created by the lab or obtained from a commercial source. These stock plates themselves are not directly used in experiments; instead, separate *assay plates* are created as needed. An assay plate is simply a copy of a stock plate, created by pipetteing a small amount of liquid (often measured in nanoliters) from the wells of a stock plate to the corresponding wells of a completely empty plate [15].

Reaction observation

To prepare for an assay, the researcher fills each well of the plate with some logical entity that he or she wishes to conduct the experiment upon, such as a protein, or an animal embryo. After some incubation time has passed to allow the biological matter to absorb, bind to, or otherwise react (or fail to react) with the compounds in the wells, measurements are taken across all the plate's wells, either manually or by a machine. Manual measurements are often necessary when the researcher is using microscopy to (for example) seek changes or defects in embryonic development caused by the wells' compounds, looking for effects that a computer could not easily determine by itself. Otherwise, a specialized automated analysis machine can run a number of experiments on the wells (such as shining polarized light on them and measuring reflectivity, which

R. Ashok Raj. / Vol 2/ Issue 1/ 2012/ 22-39.

can be an indication of protein binding). In this case, the machine outputs the result of each experiment as a grid of numeric values, with each number mapping to the value obtained from a single well. A high-capacity analysis machine can measure dozens of plates in the space of a few minutes like this, generating thousands of experimental datapoints very quickly.

Depending on the results of this first assay, the researcher can perform follow up assays within the same screen by "cherry picking" liquid from the source wells that gave interesting results (known as "hits") into new assay plates, and then re-running the experiment to collect further data on this narrowed set, confirming and refining observations [16].

Automation systems

Automation is an important element in HTS's usefulness. Typically, an integrated robot system consisting of one or more robots transports assay-microplates from station to station for sample and reagent addition, mixing, incubation, and finally readout or detection. An HTS system can usually prepare, incubate, and analyze many plates simultaneously, further speeding the data-collection process. HTS robots currently exist which can test up to 100,000 compounds per day. Automatic colony pickers pick thousands of microbial colonies for high throughput genetic screening. The term uHTS or *ultra high throughput screening* refers (circa 2008) to screening in excess of 100,000 compounds per day [17].

Experimental design and data analysis

With the ability of rapid screening of diverse compounds (such as small molecules or siRNAs) to identify active compounds, HTS has led to an explosion in the rate of data generated in recent years. Consequently, one of the most fundamental challenges in HTS experiments is to glean biochemical significance from mounds of data, which relies on the development and adoption of appropriate experimental designs and analytic methods for both quality control and hit selection [4]. HTS research is one of the fields which have a feature described by Eisenstein as follows: soon, if a scientist does not understand some statistics or rudimentary data-handling technologies, he or she may not be considered to be a true molecular biologist and thus will simply become a dinosaur.

Quality control

High-quality HTS assays are critical in HTS experiments. The development of high-quality HTS assays requires the integration of both experimental and computational approaches for quality control (QC). Three important means of QC are

(i) good plate design,
(ii) the selection of effective positive and negative chemical/biological controls, and
(iii) the development of effective QC metrics to measure the degree of differentiation so that assays with inferior data quality can be identified. A good plate design helps to identify systematic errors (especially those linked with well position) and determine what normalization should be used to remove/reduce the impact of systematic errors on both QC and hit selection.

Effective analytic QC methods serve as a gatekeeper for excellent quality assays. In a typical HTS experiment, a clear distinction between a positive control and a negative reference such as a negative control is an index for good quality. Many quality assessment measures have been proposed to measure the degree of differentiation between a positive control and a negative reference. Signal-to-background ratio, signal-to-noise ratio, signal window, assay variability ratio, and Z-factor have been adopted to evaluate data quality. Strictly standardized mean difference (SSMD) has recently been proposed for assessing data quality in HTS assays [18].

Hit selection

A compound with a desired size of effects in an HTS screen is called a hit. The process of selecting hits is called hit selection. The analytic methods for hit selection in screens without replicates (usually in primary screens) differ from those with replicates (usually in confirmatory screens). For example, the z-score method is suitable for screens without replicates whereas the t-statistic is suitable for screens with replicate. The calculation of SSMD for screens without replicates also differs from that for screens with replicates.

For hit selection in primary screens without replicates, the easily interpretable ones are average fold change, mean difference, percent inhibition, and percent activity. However, they do not capture data variability effectively. The z-score method or SSMD, which can capture data variability based on an assumption that every compound has the same variability as a negative reference in the screens. However, outliers are common in HTS experiments, and methods such as z-score are sensitive to outliers and can be problematic. Consequently, robust methods such as the z*-score method, SSMD*, B-score method, and quantile-based method have been proposed and adopted for hit selection [19].

In a screen with replicates, we can directly estimate variability for each compound; consequently, we should use SSMD or t-statistic that does not rely on the strong assumption that the z-score and z*-score rely on. One issue with the use of t-statistic and associated p-values is that they are affected by both sample size and effect size. They come from testing for no mean difference, thus are not designed to measure the size of compound effects. For hit selection, the major interest is the size of effect in a tested compound.

R. Ashok Raj. / Vol 2/ Issue 1/ 2012/ 22-39.

SSMD directly assesses the size of effects. SSMD has also been shown to be better than other commonly used effect sizes. The population value of SSMD is comparable across experiments and thus we can use the same cutoff for the population value of SSMD to measure the size of compound effects.

Techniques for increased throughput and efficiency

Unique distributions of compounds across one or many plates can be employed to increase either the number of assays per plate, or to reduce the variance of assay results, or both. The simplifying assumption made in this approach is that any N compounds in the same well will not typically interact with each other, or the assay target, in a manner that fundamentally changes the ability of the assay to detect true hits.

For example, imagine a plate where compound A is in wells 1-2-3, compound B is in wells 2-3-4, and compound C is in wells 3-4-5. In an assay of this plate against a given target, a hit in wells 2, 3, and 4 would indicate that compound B is the most likely agent, while also providing three measurements of compound B's efficacy against the specified target. Commercial applications of this approach involve combinations in which no two compounds ever share more than one well, to reduce the (second-order) possibility of interference between pairs of compounds being screened [20].

Recent advances

In March 2010 research was published demonstrating an HTS process allowing 1,000 times faster screening (100 million reactions in 10 hours) at 1 millionth the cost (using 10^{-7} times the reagent volume) than conventional techniques using drop-based microfluidics.[18] Drops of fluid separated by oil replace microplate wells and allow analysis and hit sorting while reagents are flowing through channels.

In 2010 researchers developed a silicon sheet of lenses that can be placed over microfluidic arrays to allow the fluorescence measurement of 64 different output channels simultaneously with a single camera. This process can analyze 200,000 drops per second.

Increasing lab utilization of HTS

HTS is a relatively recent innovation, made lately feasible through modern advances in robotics and high-speed computer technology. It still takes a highly specialized and expensive screening lab to run an HTS operation, so in many cases a small-to-moderately sized research institution will use the services of an existing HTS facility rather than set up one for itself.

There is a trend in academia to be their own drug discovery enterprise. (High-throughput screening goes to school) These facilities, which normally are only found in industry, are now increasingly be found as well at universities. UCLA for example, features an HTS laboratory (Molecular Screening Shared Resources (MSSR, UCLA) which can screen more than 100,000 compounds a day on a routine basis. The University of Illinois also has a facility for HTS, as does the University of Minnesota. The Rockefeller University, has an open access (infrastructure) HTS Resource Center HTSRC (The Rockefeller University, HTSRC) which offers a library of over 165,000 compounds. Northwestern University's High Throughput Analysis Laboratory supports target identification, validation, assay development, and compound screening.

In the United States, the National Institute of Health or NIH has created a nationwide consortium of small molecule screening centers that has been recently funded to produce innovative chemical tools for use in biological research. The Molecular Libraries Screening Center Network or MLSCN performs HTS on assays provided by the research community, against a large library of small molecules maintained in a central molecule repository.

HIGH-CONTENT SCREENING

High-content screening is a drug discovery method that uses images of living cells as the basic unit for molecule discovery. Proteins of interest present in the cells are detected using fluorescent tags, such as the green fluorescent protein, or by fluorescent antibodies. Image analysis is then used to measure changes in properties of the cells caused by external treatment such as chemical inhibitors or RNA interference- for example if the cells division is slowed or entry of a protein into the cell is arrested. High content screening is the combination of modern cell biology, with all its molecular tools, with automated high resolution microscopy and robotic handling. It differs from most life science work because experiment evaluation is mostly automated. The technology is mainly used to determine whether a potential drug affects aspects of cell function involved in or that describe a disease. For example, in humans G-protein coupled receptors (GPCRs) are a large family of around 880 cell surface proteins that transduce extra-cellular changes in the environment into a cell response, like triggering an increase in blood pressure because of the release of a regulatory hormone into the blood stream. Activation of these GPCRs can involve their entry into cells and when this can be visualised it can be the basis of a systematic analysis of receptor function through chemical genetics, systematic genome wide screening or physiological manipulation [21].

High Content screening is also known as cell based screening, phenotypic screening or visual screening, and it has a major place in pharmaceutical company drug discovery. All these terms refer to the systematic search for new drugs, small molecule inhibitors or chemical entities

that could have use in biology or medicine. The parallel acquisition of data on different cell properties, for example activity of signal transduction cascades and cytoskeleton integrity is the main advantage of this method in comparison to the faster but less detailed high throughput screening. While HCS is slower, the wealth of acquired data allows a more profound understanding of drug effects.

Automated image based screening permits the identification of small compounds altering cellular phenotypes and is of interest for the discovery of new pharmaceuticals and new cell biological tools for modifying cell function. The selection of molecules based on a cellular phenotype does not require a priori knowledge of the biochemical targets that are affected by compounds and while this may be a benefit for compound discovery, the biochemical target itself must be subsequently identified. Given the increase in the use of phenotypic/visual screening as a cell biological tool, methods are required that permit systematic biochemical target identification if these molecules are to be of broad use. Target identification has been defined as the rate limiting step in chemical genetics/high-content screening.

The history of high-content screening

High-content screening technology allows for the evaluation of multiple biochemical and morphological parameters in cellular systems. Through combining the imaging of cells in microtiter plates with automated image analysis algorithms, researchers can acquire deeper knowledge on multiple biochemical or morphological pathways at the single-cell level at an early stage in the development new drugs.

The utility of automated cell biology requires an examination of how automation and objective measurement can improve the experimentation and the understanding of disease. First, it removes the influence of the investigator in most, but not all, aspects of cell biology research and second it makes entirely new approaches to possible.

In review, classical 20th century cell biology used cell lines grown in culture where the experiments were measured using very similar to that described here, but there the investigator made the choice on what was measured and how. In the early 1990s, the development of CCD cameras (charge coupled device cameras) for research created the opportunity to measure features in pictures of cells- such as how much protein is in the nucleus, how much is outside. Sophisticated measurements soon followed using new fluorescent molecules, which are used to measure cell properties like second messenger concentrations or the pH of internal cell compartments. The wide use of the green fluorescent protein, a natural fluorescent protein molecule from jellyfish, then accelerated the trend toward cell imaging as a mainstream technology in cell biology. Despite these advances, the choice of which cell to image and which data to present and how to analyse it was still selected by the investigator.

By analogy, if one imagines a football field and dinner plates laid across it, instead of looking at all of them, the investigator would choose a handful near the score line and had to leave the rest. In this analogy the field is a tissue culture dish, the plates the cells growing on it. While this was a reasonable and pragmatic approach automation of the whole process and the analysis makes possible the analysis of the whole population of living cells, so the whole football field can be measured.

From one to many

This technology allows a (very) large number of experiments to be performed, allowing explorative screening. Currently its main use is in chemical genetics where large, diverse small molecule collections are systematically tested for their effect on cellular model systems. Novel drugs can be found using screens of tens of thousands of molecules, and these have promise for the future of drug development. Beyond drug discovery, chemical genetics is aimed at functionalizing the genome by identifying small molecules that acts on most of the 21,000 gene products in a cell. High-content technology will be part of this effort which could provide useful tools for learning where and when proteins act by knocking them out chemically. This would be most useful for gene where knockout mice (missing one or several genes) can not be made because the protein is required for development, growth or otherwise lethal when it is not there. Chemical knock out could address how and where these genes work. Further the technology is used in combination with RNAi to identify sets of genes involved in specific mechanisms, for example cell division. Here, libraries of RNAis, covering a whole set of predicted genes inside the target organisms genome can be used to identify relevant subsets, facilitating the annotation of genes for which no clear role has been established beforehand. The large datasets produced by automated cell biology contain spatially resolved, quantitative data which can be used for building for systems level models and simulations of how cells and organisms function. Systems biology models of cell function would permit prediction of why, where and how the cell responds to external changes, growth and disease.

High-content screening technology

High-content screening technology is mainly based on automated digital microscopy and flow cytometry, in combination with IT-systems for the analysis and storage of the data. “High-content” or visual biology technology has two purposes, first to acquire spatially or temporally resolved information on an event and second to automatically quantify it. Spatially resolved instruments are typically automated microscopes, and temporal resolution

still requires some form of fluorescence measurement in most cases.This means that a lot of HCS instruments are (fluorescence) microscopes that are connected to some form of image analysis package. These take care of all the steps in taking fluorescent images of cells and provide rapid, automated and unbiased assessment of experiments [22].

Technology providers

The instruments on the market can be divided on the basis of price, footprint and the ethereal design qualities of the box they come in - but the most incisive difference is whether the instruments are optical confocal or not. Confocal imaging summarizes as imaging/resolving a thin slice through an object and rejecting out of focus light that comes from outside this slide. This gives higher image signal to noise and higher resolution than the more commonly applied epi-fluorescence microscopy. For many biological assays, confocal imaging is not ideal (e.g. phototoxicity issues or the need for a larger focal depth etc.). What all instruments share is the ability to take, store and interpret images automatically and most integrate into large robotic cell/medium handling platforms.

- The current confocal platforms are the point scanning 4 color ImageXpress ULTRA (Molecular Devices, Union City, USA), the spinning disk (nipkow disk) Pathway 855 and 435 from BD Biosciences (formerly Atto Biosciences, Rockville, Maryland), Opera (PerkinElmer Inc., Waltham, MA) and the slit scanning IN Cell 3000 (GE/Amersham Biosciences, Cardiff, UK).

- The current widefield platforms are: Arrayscan VTI (Cellomics (Cellomics)), IN Cell Analyzer 2000 (GE Healthcare Piscataway, New Jersey, USA), Acumen eX3 (TTP LabTech Ltd, Cambridge, UK), Scanalyzer (Scanalyzer LemnaTec, Aachen Germany) and "ImageXpress MICRO" (Molecular Devices, Union City, USA). Several instruments are ready for live cell temperature controlled imaging - such as the "IN Cell 1000" (GE/Amersham Biosciences Piscataway, New Jersey, USA), the "Pathway HT" (Becton Dickinson Biosciences) and the "ImageXpress MICRO" (Molecular Devices, Union City, USA), "Scan^R" (Olympus Soft Imaging Solutions, Germany), which may be of use to biologists interested in following events over relatively long time courses in vivo.

Alternatively more generic microscope instrumentation such as the Cell Observer by Carl Zeiss can be used for high content screening. While such instruments are less specialized they can be more appealing to academic settings, where tasks and experiments change more rapidly than in industrial research [19,20].

Additionally, many of these HCS platforms have been extended by (Caliper life sciences (Caliper Life Science))'s (Hopkinton, MA) integration of liquid handlers Sciclone G3 Liquid Handling Workstation and Zephyr, as well as plate moving robot Twister II to further automate the screening process.

Dedicated software for image analysis is available from these vendors (IN Cell Investigator, Perkin Elmer Columbus) or from specialised firms such as DCILabs (Belgium), LemnaTec (Germany) and from Accerlys Pipeline Pilot, as well as from some groups which provide open-source software for image analysis, such as BioImageXD, CellProfiler, and ImageJ. Dedicated open-source software for data management and data mining: Open source web based database system to publish image based screening experiments HCPB.

Kits for high-content screening of various target proteins (e.g. p53, c-jun and NFkB) have recently become available from commercial suppliers.

Drug development is a blanket term used to define the process of bringing a new drug to the market once a lead compound has been identified through the process of drug discovery. It includes pre-clinical research (microorganisms/animals) and clinical trials (on humans) and may include the step of obtaining regulatory approval to market the drug.

New Chemical Entity (NCE) development

Broadly the process can be divided into pre-clinical and clinical work.

Pre-clinical. New Chemical Entities (NCEs)(also known as New Molecular Entities [NMEs]) are compounds which emerge from the process of drug discovery. These will have promising activity against a particular biological target thought to be important in disease; however, little will be known about the safety, toxicity, pharmacokinetics and metabolism of this NCE in humans. It is the function of drug development to assess all of these parameters prior to human clinical trials. A further major objective of drug development is to make a recommendation of the dose and schedule to be used the first time an NCE is used in a human clinical trial ("first-in-man" [FIM] or First Human Dose [FHD]).

In addition, drug development is required to establish the physicochemical properties of the NCE: its chemical makeup, stability, solubility. The process by which the chemical is made will be optimized so that from being made at the bench on a milligram scale by a synthetic chemist, it can be manufactured on the kilogram and then on the ton scale. It will be further examined for its suitability to be made into capsules, tablets, aeresol, intramuscular injectable, subcuteneous injectable, or intravenous formulations. Together these processes are known in preclinical development as Chemistry, Manufacturing and Control (CMC).

Many aspects of drug development are focused on satisfying the regulatory requirements of drug licensing authorities. These generally constitute a number of tests designed to determine the major toxicities of a novel compound prior to first use in man. It is a legal requirement that an assessment of major organ toxicity be performed (effects on the heart and lungs, brain, kidney, liver and digestive system), as well as effects on other parts of the body that might be affected by the drug (e.g. the skin if the new drug is to be delivered through the skin). While, increasingly, these tests can be made using in vitro methods (e.g. with isolated cells), many tests can only be made by using experimental animals, since it is only in an intact organism that the complex interplay of metabolism and drug exposure on toxicity can be examined.

The information gathered from this pre-clinical testing, as well as information on CMC, and is submitted to regulatory authorities (in the US, to the FDA), as an Investigational New Drug application or IND. If the IND is approved, development moves to the clinical phase [23].

Clinical phase

Clinical trials involves three steps: Phase I trials, usually in healthy patients, determine safety and dosing; Phase II trials are used to get an initial reading of efficacy and further explore safety in small numbers of sick patients; Phase III trials a large, pivotal trials to determine safety and efficacy in sufficiently large numbers of patients.

The process of drug development does not stop once an NCE begins human clinical trials. In addition to the tests required to move a novel drug into the clinic for the first time it is also important to ensure that long-term or chronic toxicities are determined, as well as effects on systems not previously monitored (fertility, reproduction, immune system, etc.). The compound will also be tested for its capability to cause cancer (carcinogenicity testing).

If a compound emerges from these tests with an acceptable toxicity and safety profile, and it can further be demonstrated to have the desired effect in clinical trials, then it can be submitted for marketing approval in the various countries where it will be sold. In the US, this process is called a New Drug Application or NDA. Most NCEs, however, fail during drug development, either because they have some unacceptable toxicity, or because they simply do not work in clinical trials [24].

Cost

The full of cost of bringing a new drug (i.e. a drug that is a new chemical entity) to market - from discovery through clinical trials to approval - is complex and controversial. One element of the complexity is that the much-publicized final numbers often do not include just the simple out-of-pocket expenses, but also include "capital costs", which are included to take into account the long time period (often at least ten years) during which the out-of-pocket costs are expended; additionally it is often not stated whether a given figure includes the capitalized cost or comprises only out-of-pocket expenses. Another element of complexity is that all estimates are based on confidential information owned by drug companies, released by them voluntarily. There is currently no way to validate these numbers. The numbers are controversial, as drug companies use them to justify the prices of their drugs and various advocates for lower drug prices have challenged them. The controversy is not only between "high" and "low" -- the numbers also vary greatly at the high end.

A study published by Steve Paul et al in 2010 in Nature Reviews: Drug Discovery compares many of the studies, provides both capitalized and out-of-pocket costs for each, and lays out the assumptions each makes: see Supplemental Box 2. The authors offer their own estimate of the capitalized cost as being ~$1.8B, with out-of-pocket costs of ~$870M.

Studies published by diMasi *et al.* in 2003, report an average pre-tax, capitalized cost of approximately $800 million to bring one of the drugs from the study to market. Also, this $800 million dollar figure includes opportunity costs of $400 million. A study published in 2006 estimates that costs vary from around $500 million to $2 billion depending on the therapy or the developing firm. A study published in 2010 in the journal Health Economics, including an author from the US Federal Trade Commission, was critical of the methods used by diMasi *et al.* but came up with a higher estimate of ~$1.2B. Marcia Angell, M.D., a former editor of the New England Journal of Medicine, has called that number grossly inflated, and estimates that the total is closer to $100 million. A 2011 study also critical of the diMasi methods, puts average costs at $55 million [25].

Pre-clinical development

In drug development, pre-clinical development is a stage of research that begins before clinical trials (testing in humans) can begin, and during which important feasibility, iterative testing and drug safety data is collected.

The main goals of pre-clinical studies (also named preclinical studies and nonclinical studies) are to determine a product's ultimate safety profile. Products may include new or iterated or like-kind medical devices, drugs, gene therapy solutions, etc. Each class of product may undergo different types of preclinical research. For instance, drugs may undergo pharmacodynamics (PD), pharmacokinetics (PK), ADME, and toxicity testing through animal testing. This data allows researchers to allometrically estimate a safe starting dose of the drug for clinical trials in humans. Medical devices that do not have drug attached will not undergo these additional tests and may go directly to GLP

testing for safety of the device and its components. Some medical devices will also undergo biocompatibility testing which helps to show whether a component of the device or all components are sustainable in a living model. Most pre-clinical studies must adhere to Good Laboratory Practices (GLP) in ICH Guidelines to be acceptable for submission to regulatory agencies such as the Food & Drug Administration in the United States.

Typically, both *in vitro* and *in vivo* tests will be performed. Studies of a drug's toxicity include which organs are targeted by that drug, as well as if there are any long-term carcinogenic effects or toxic effects on mammalian reproduction.

The information collected from these studies is vital so that safe human testing can begin. Typically, in drug development studies animal testing involves two species. The most commonly used models are murine and canine, although primate and porcine are also used. The choice of species is based on which will give the best correlation to human trials. Differences in the gut, enzyme activity, circulatory system, or other considerations make certain models more appropriate based on the dosage form, site of activity, or noxious metabolites. For example, canines may not be good models for solid oral dosage forms because the characteristic carnivore intestine is underdeveloped compared to the omnivore's, and gastric emptying rates are increased. Also, rodents cannot act as models for antibiotic drugs because the resulting alteration to their intestinal flora causes significant adverse effects. Depending on drugs functional groups, it may be metabolized in similar or different ways between species, which will affect both efficacy and toxicology. Medical device studies also use this basic premise. Most studies are performed in larger species such as dogs, pigs and sheep which allow for testing in a similar sized model as that of a human. In addition, some species are used for similarity in specific organs or organ system physiology (swine for dermatological and coronary stent studies; goats for mammary implant studies; dogs for gastric studies; etc.).

Based on pre-clinical trials, No Observable Effect Levels (NOEL) on drugs are established, which are used to determine initial phase 1 clinical trial dosage levels on a mass API per mass patient basis. Generally a 1/100 uncertainty factor or "safety margin" is included to account for interspecies (1/10) and inter-individual (1/10) differences [26].

Animal testing in the research-based pharmaceutical industry has been reduced in recent years both for ethical and cost reasons. However, most research will still involve animal based testing for the need of similarity in anatomy and physiology that is required for diverse product development.

DRUG DESIGN

Drug design, sometimes referred to as rational drug design or more simply rational design, is the inventive process of finding new medications based on the knowledge of a biological target.[1] The drug is most commonly an organic small molecule that activates or inhibits the function of a biomolecule such as a protein, which in turn results in a therapeutic benefit to the patient. In the most basic sense, drug design involves the design of small molecules that are complementary in shape and charge to the biomolecular target with which they interact and therefore will bind to it. Drug design frequently but not necessarily relies on computer modeling techniques. This type of modeling is often referred to as computer-aided drug design. Finally, drug design that relies on the knowledge of the three-dimensional structure of the biomolecular target is known as structure-based drug design.

The phrase "drug design" is to some extent a misnomer. What is really meant by drug design is ligand design (i.e., design of a small molecule that will bind tightly to its target). Although modeling techniques for prediction of binding affinity are reasonably successful, there are many other properties, such as bioavailability, metabolic half-life, lack of side effects, etc., that first must be optimized before a ligand can become a safe and efficacious drug. These other characteristics are often difficult to optimize using rational drug design techniques.

Typically 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. In addition, these drugs should also be designed so as not to affect any other important "off-target" molecules or antitargets that may be similar in appearance to the target molecule, since drug interactions with off-target molecules may lead to undesirable side effects. Sequence homology is often used to identify such risks.

Most commonly, drugs are organic small molecules produced through chemical synthesis, but biopolymer-based drugs (also known as biologics) produced through biological processes are becoming increasingly more common. In addition, mRNA-based gene silencing technologies may have therapeutic applications [27].

Types

There are two major types of drug design. The first

is referred to as ligand-based drug design and the second, structure-based drug design.

Ligand-based

Ligand-based drug design (or indirect drug design) relies on knowledge of other molecules that bind to the biological target of interest. These other molecules may be used to derive a pharmacophore model that defines the minimum necessary structural characteristics a molecule must possess in order to bind to the target. In other words, a model of the biological target may be built based on the knowledge of what binds to it, and this model in turn may be used to design new molecular entities that interact with the target. Alternatively, a quantitative structure-activity relationship (QSAR), in which a correlation between calculated properties of molecules and their experimentally determined biological activity, may be derived. These QSAR relationships in turn may be used to predict the activity of new analogs.

Structure-based

Structure-based drug design (or direct drug design) relies on knowledge of the three dimensional structure of the biological target obtained through methods such as x-ray crystallography or NMR spectroscopy. If an experimental structure of a target is not available, it may be possible to create a homology model of the target based on the experimental structure of a related protein. Using the structure of the biological target, candidate drugs that are predicted to bind with high affinity and selectivity to the target may be designed using interactive graphics and the intuition of a medicinal chemist. Alternatively various automated computational procedures may be used to suggest new drug candidates.

As experimental methods such as X-ray crystallography and NMR develop, the amount of information concerning 3D structures of biomolecular targets has increased dramatically. In parallel, information about the structural dynamics and electronic properties about ligands has also increased. This has encouraged the rapid development of the structure-based drug design. Current methods for structure-based drug design can be divided roughly into two categories. The first category is about “finding” ligands for a given receptor, which is usually referred as database searching. In this case, a large number of potential ligand molecules are screened to find those fitting the binding pocket of the receptor. This method is usually referred as ligand-based drug design. The key advantage of database searching is that it saves synthetic effort to obtain new lead compounds. Another category of structure-based drug design methods is about “building” ligands, which is usually referred as receptor-based drug design. In this case, ligand molecules are built up within the constraints of the binding pocket by assembling small pieces in a stepwise manner. These pieces can be either individual atoms or molecular fragments. The key advantage of such a method is that novel structures, not contained in any database, can be suggested. These techniques are raising much excitement to the drug design community [28].

Active site identification

Active site identification is the first step in this program. It analyzes the protein to find the binding pocket, derives key interaction sites within the binding pocket, and then prepares the necessary data for Ligand fragment link. The basic inputs for this step are the 3D structure of the protein and a pre-docked ligand in PDB format, as well as their atomic properties. Both ligand and protein atoms need to be classified and their atomic properties should be defined, basically, into four atomic types:

- **hydrophobic atom**: All carbons in hydrocarbon chains or in aromatic groups.
- **H-bond donor**: Oxygen and nitrogen atoms bonded to hydrogen atom(s).
- **H-bond acceptor**: Oxygen and sp2 or sp hybridized nitrogen atoms with lone electron pair(s).
- **Polar atom**: Oxygen and nitrogen atoms that are neither H-bond donor nor H-bond acceptor, sulfur, phosphorus, halogen, metal, and carbon atoms bonded to hetero-atom(s).

The space inside the ligand binding region would be studied with virtual probe atoms of the four types above so the chemical environment of all spots in the ligand binding region can be known. Hence we are clear what kind of chemical fragments can be put into their corresponding spots in the ligand binding region of the receptor.

When we want to plant “seeds” into different regions defined by the previous section, we need a fragments database to choose fragments from. The term “fragment” is used here to describe the building blocks used in the construction process. The rationale of this algorithm lies in the fact that organic structures can be decomposed into basic chemical fragments. Although the diversity of organic structures is infinite, the number of basic fragments is rather limited.

Before the first fragment, i.e. the seed, is put into the binding pocket, and other fragments can be added one by one, it is useful to identify potential problems. First, the possibility for the fragment combinations is huge. A small perturbation of the previous fragment conformation would cause great difference in the following construction process. At the same time, in order to find the lowest binding energy on the Potential energy surface (PES) between planted fragments and receptor pocket, the scoring function calculation would be done for every step of conformation change of the fragments derived from every type of possible fragments combination. Since this requires a large amount

R. Ashok Raj. / Vol 2/ Issue 1/ 2012/ 22-39.

of computation, one may think using other possible strategies to let the program works more efficiently. When a ligand is inserted into the pocket site of a receptor, conformation favor for these groups on the ligand that can bind tightly with receptor should be taken priority. Therefore it allows us to put several seeds at the same time into the regions that have significant interactions with the seeds and adjust their favorite conformation first, and then connect those seeds into a continuous ligand in a manner that make the rest part of the ligand having the lowest energy. The conformations of the pre-placed seeds ensuring the binding affinity decide the manner that ligand would be grown. This strategy reduces calculation burden for the fragment construction efficiently. On the other hand, it reduces the possibility of the combination of fragments, which reduces the number of possible ligands that can be derived from the program. These two strategies above are well used in most structure-based drug design programs. They are described as “Grow” and “Link”. The two strategies are always combined in order to make the construction result more reliable [17,19].

Scoring method

Structure-based drug design attempts to use the structure of proteins as a basis for designing new ligands by applying accepted principles of molecular recognition. The basic assumption underlying structure-based drug design is that a good ligand molecule should bind tightly to its target. Thus, one of the most important principles for designing or obtaining potential new ligands is to predict the binding affinity of a certain ligand to its target and use it as a criterion for selection.

One early method was developed by Böhm to develop a general-purposed empirical scoring function in order to describe the binding energy. The following “Master Equation” was derived:

$$\Delta G_{\text{bind}} = -RT \ln K_{\text{d}}$$

$$K_{\text{d}} = \frac{[\text{Receptor}][\text{Acceptor}]}{[\text{Complex}]}$$

$$\Delta G_{\text{bind}} = \Delta G_{\text{desolvation}} + \Delta G_{\text{motion}} + \Delta G_{\text{configuration}} + \Delta G_{\text{interaction}}$$

where:

- desolvation – enthalpic penalty for removing the ligand from solvent
- motion – entropic penalty for reducing the degrees of freedom when a ligand binds to its receptor
- configuration – conformational strain energy required to put the ligand in its "active" conformation
- interaction – enthalpic gain for "resolvating" the ligand with its receptor

The basic idea is that the overall binding free energy can be decomposed into independent components that are known to be important for the binding process. Each component reflects a certain kind of free energy alteration during the binding process between a ligand and its target receptor. The Master Equation is the linear combination of these components. According to Gibbs free energy equation, the relation between dissociation equilibrium constant, K_d, and the components of free energy was built.

Various computational methods are used to estimate each of the components of the master equation. For example, the change in polar suface area upon ligand binding can be used to estimate the desolation energy. The number of rotatable bonds frozen upon ligand binding is proportional to the motion term. The configurational or strain energy can be estimated using molecular mechanics calculations. Finally the interaction energy can be estimated using methods such as the change in non polar surface, statistically derived potentials of mean force, the number of hydrogen bonds formed, etc. In practice, the components of the master equation are fit to experimental data using multiple linear regression. This can be done with a diverse training set including many types of ligands and receptors to produce a less accurate but more general "global" model or a more restricted set of ligands and receptors to produce a more accurate but less general "local" model.

Rational drug discovery

In contrast to traditional methods of drug discovery, which rely on trial-and-error testing of chemical substances on cultured cells or animals, and matching the apparent effects to treatments, rational drug design begins with a hypothesis that modulation of a specific biological target may have therapeutic value. In order for a biomolecule to be selected as a drug target, two essential pieces of information are required. The first is evidence that modulation of the target will have therapeutic value. This knowledge may come from, for example, disease linkage studies that show an association between mutations in the biological target and certain disease states. The second is that the target is "drugable". This means that it is capable of binding to a small molecule and that its activity can be modulated by the small molecule [15-20].

Once a suitable target has been identified, the target is normally cloned and expressed. The expressed target is then used to establish a screening assay. In addition, the three-dimensional structure of the target may be determined.

The search for small molecules that bind to the target is begun by screening libraries of potential drug compounds. This may be done by using the screening assay (a "wet screen"). In addition, if the structure of the target is available, a virtual screen may be performed of candidate drugs. Ideally the candidate drug compounds should be "drug-like", that is they should possess properties that are predicted to lead to oral bioavailability, adequate chemical and metabolic stability, and minimal toxic effects. Several

methods are available to estimate druglikeness such Lipinski's Rule of Five and a range of scoring methods such as Lipophilic efficiency. Several methods for predicting drug metabolism have been proposed in the scientific literature, and a recent example is SPORCalc. Due to the complexity of the drug design process, two terms of interest are still serendipity and bounded rationality. Those challenges are caused by the large chemical space describing potential new drugs without side-effects.

Computer-aided drug design

Computer-aided drug design uses computational chemistry to discover, enhance, or study drugs and related biologically active molecules. The most fundamental goal is to predict whether a given molecule will bind to a target and if so how strongly. Molecular mechanics or molecular dynamics are most often used to predict the conformation of the small molecule and to model conformational changes in the biological target that may occur when the small molecule binds to it. Semi-empirical, ab initio quantum chemistry methods, or density functional theory are often used to provide optimized parameters for the molecular mechanics calculations and also provide an estimate of the electronic properties (electrostatic potential, polarizability, etc.) of the drug candidate that will influence binding affinity.

Molecular mechanics methods may also be used to provide semi-quantitative prediction of the binding affinity. Also, knowledge-based scoring function may be used to provide binding affinity estimates. These methods use linear regression, machine learning, neural nets or other statistical techniques to derive predictive binding affinity equations by fitting experimental affinities to computationally derived interaction energies between the small molecule and the target.

Ideally the computational method should be able to predict affinity before a compound is synthesized and hence in theory only one compound needs to be synthesized. The reality however is that present computational methods are imperfect and provide at best only qualitatively accurate estimates of affinity. Therefore in practice it still takes several iterations of design, synthesis, and testing before an optimal molecule is discovered. On the other hand, computational methods have accelerated discovery by reducing the number of iterations required and in addition have often provided more novel small molecule structures [14-18].

Drug design with the help of computers may be used at any of the following stages of drug discovery:

1. hit identification using virtual screening (structure- or ligand-based design)
2. hit-to-lead optimization of affinity and selectivity (structure-based design, QSAR, etc.)
3. lead optimization optimization of other pharmaceutical properties while maintaining affinity

In order to overcome the insufficient prediction of binding affinity calculated by recent scoring functions, the protein-ligand interaction and compound 3D structure information are used to analysis. For structure-based drug design, several post-screening analysis focusing on protein-ligand interaction has been developed for improving enrichment and effectively mining potential candidates:

- Consensus scoring
  - Selecting candidates by voting of multiple scoring functions
  - May lose the relationship between protein-ligand structural information and scoring criterion
- Geometric analysis
  - Comparing protein-ligand interactions by visually inspecting individual structures
  - Becoming intractable when the number of complexes to be analyzed increasing
- Cluster analysis
  - Represent and cluster candidates according to protein-ligand 3D information
  - Needs meaningful representation of protein-ligand interactions.

Examples

A particular example of rational drug design involves the use of three-dimensional information about biomolecules obtained from such techniques as X-ray crystallography and NMR spectroscopy. Computer-aided drug design in particular becomes much more tractable when there's a high-resolution structure of a target protein bound to a potent ligand. This approach to drug discovery is sometimes referred to as structure-based drug design. The first unequivocal example of the application of structure-based drug design leading to an approved drug is the carbonic anhydrase inhibitor dorzolamide, which was approved in 1995.

Another important case study in rational drug design is imatinib, a tyrosine kinase inhibitor designed specifically for the *bcr-abl* fusion protein that is characteristic for Philadelphia chromosome-positive leukemias (chronic myelogenous leukemia and occasionally acute lymphocytic leukemia). Imatinib is substantially different from previous drugs for cancer, as most agents of chemotherapy simply target rapidly dividing cells, not differentiating between cancer cells and other tissues [27-29].

Additional examples include:

- Many of the atypical antipsychotics
- Cimetidine, the prototypical H_2-receptor antagonist from which the later members of the class were developed
- Selective COX-2 inhibitor NSAIDs

R. Ashok Raj. / Vol 2/ Issue 1/ 2012/ 22-39.

- Dorzolamide, a carbonic anhydrase inhibitor used to treat glaucoma
- Enfuvirtide, a peptide HIV entry inhibitor
- Nonbenzodiazepines like zolpidem and zopiclone
- Probenecid
- SSRIs (selective serotonin reuptake inhibitors), a class of antidepressants
- Zanamivir, an antiviral drug
- Isentress, HIV Integrase inhibitor

Fig 1. An automated confocal image reader

Fig 2. Flow charts of two strategies of structure-based drug design

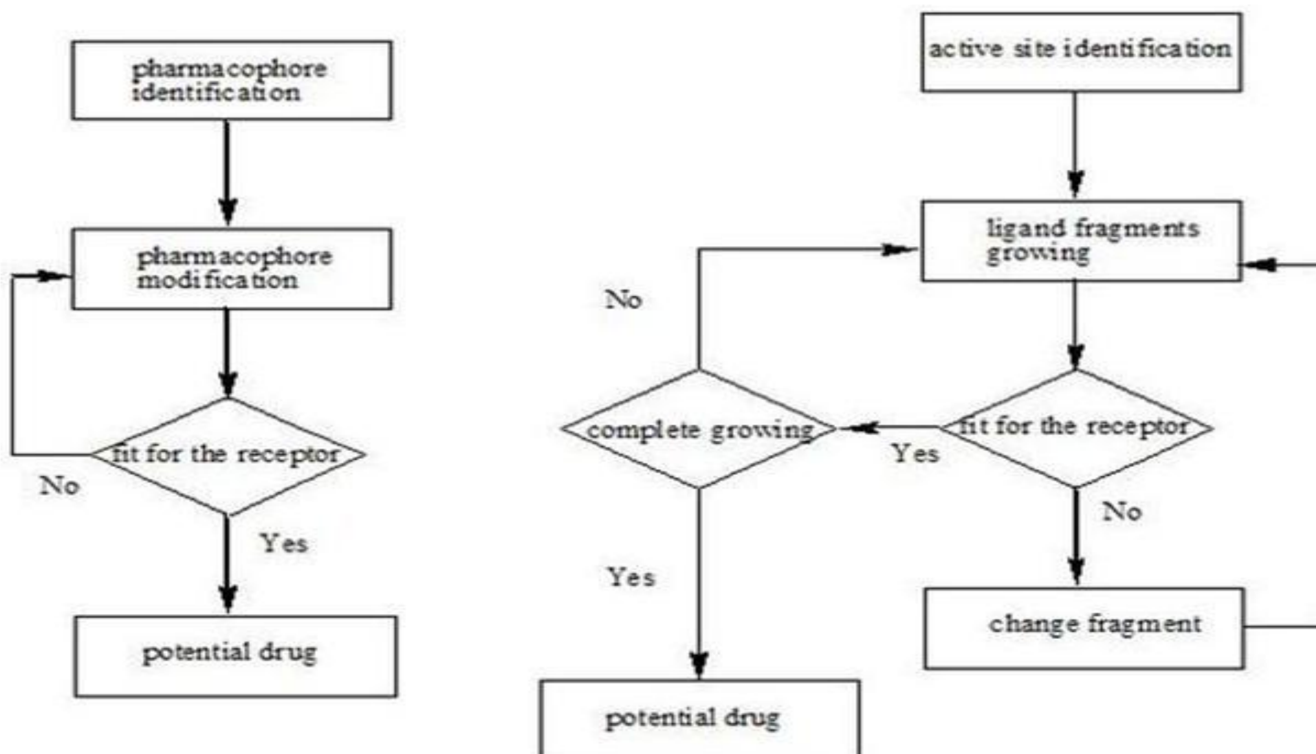


Fig 3. Structure-based drug design

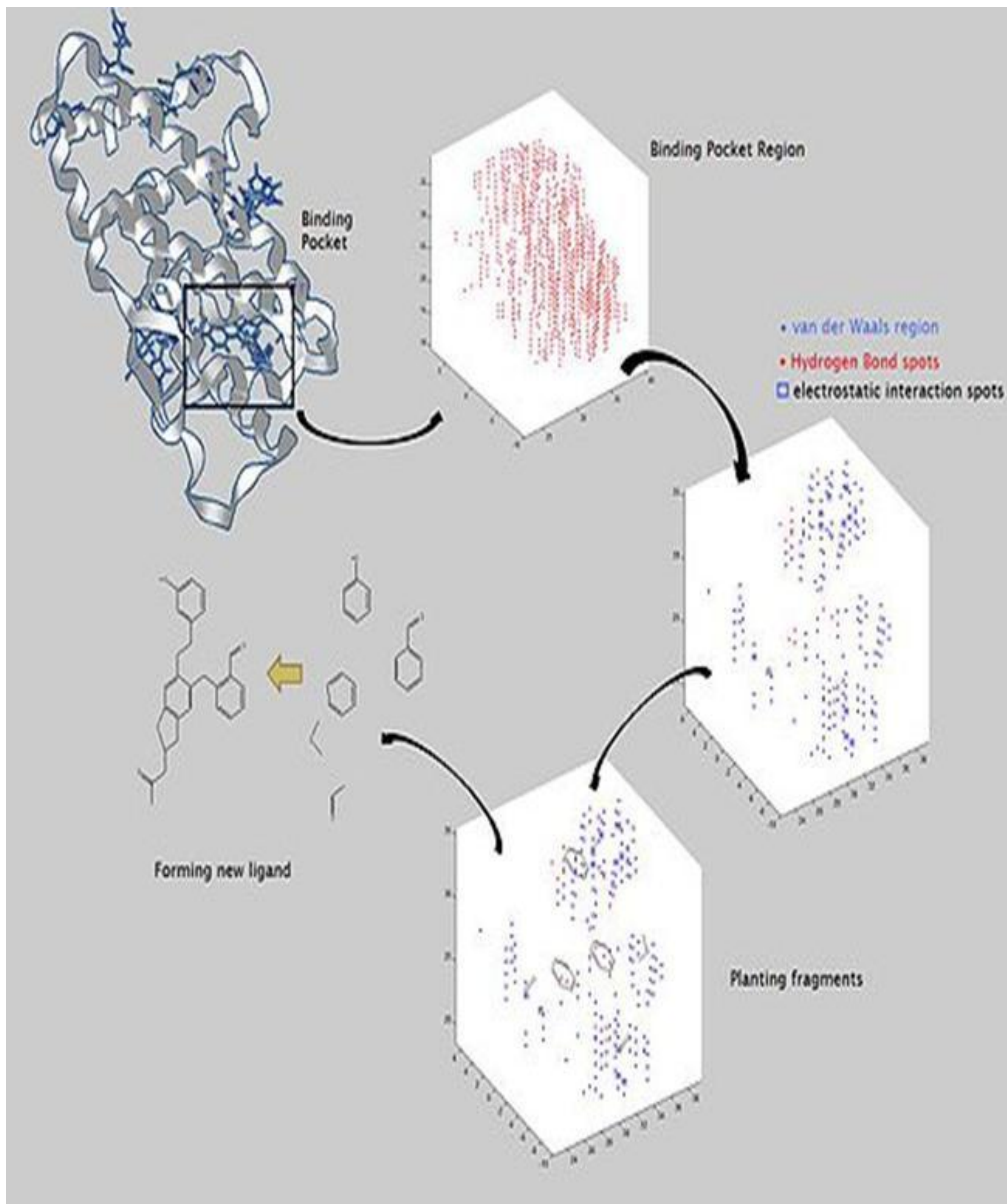
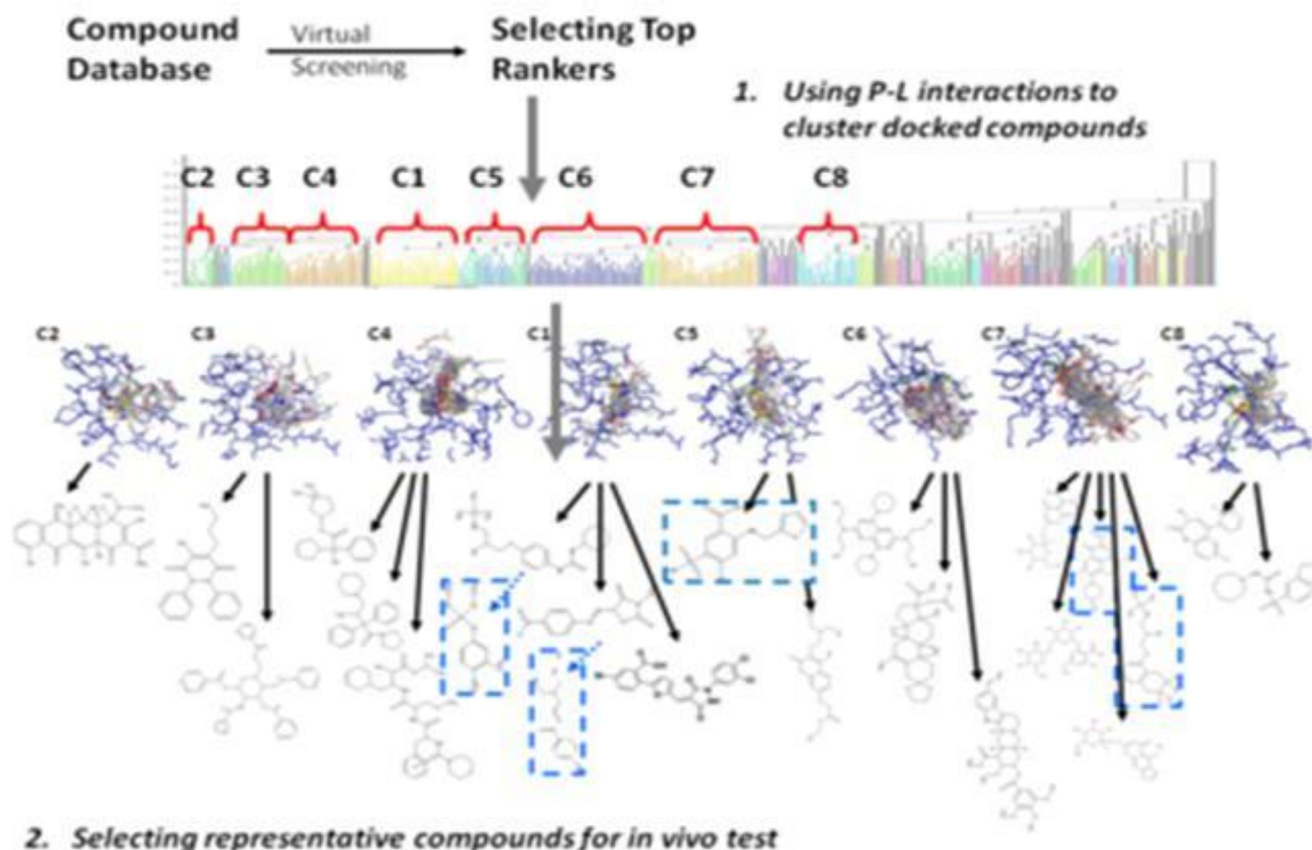


Fig 4. Usual Clustering Analysis for Structure-Based Drug Design



CONCLUSION

Candidates for a new drug to treat a disease might theoretically include from 5,000 to 10,000 chemical compounds. On average about 250 of these will show sufficient promise for further evaluation using laboratory tests, mice and other test animals. Typically, about ten of these will qualify for tests on humans. A study conducted by the Tufts Center for the Study of Drug Development covering the 1980s and 1990s found that only 21.5 percent of drugs that start phase I trials are eventually approved for marketing.

The discovery and development of new medicines is a long, complicated process. Each success is built on many, many prior failures. Advances in understanding human biology and disease are opening up exciting new possibilities for breakthrough medicines. At the same time, researchers face great challenges in understanding and applying these advances to the treatment of disease. These possibilities will grow as our scientific knowledge expands and becomes increasingly complex. Research-based pharmaceutical companies are committed to advancing science and bringing new medicines to patients.

REFERENCES

1. Anson Blake D, Ma Junyi, He Jia-Qiang. Identifying Cardiotoxic Compounds. *Genetic Engineering & Biotechnology News*, 29 (9), 2009, 34–35.
2. Steven M. Paul, Daniel S. Mytelka, Christopher T. Dunwiddie, Charles C. Persinger, Bernard H. Munos, Stacy R. Lindborg & Aaron L. Schacht. How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nature Reviews Drug Discovery,* 9 (3), 2010, 203–214.
3. Hopkins AL, Groom CR and Alexander A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today,* 9 (10), 2004, 430–431.
4. Ryckmans T. et al. Rapid assessment of a novel series of selective CB2 agonists using parallel synthesis protocols: A Lipophilic Efficiency (LipE) analysis. *Bioorg. Med. Chem. Lett.,* 19 (15), 2009, 4406–4409.

5. Leeson PD. et al. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat. Rev. Drug Disc.,* 6 (11), 2007, 881–890.
6. Feher M, Schmidt JM. Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J Chem Inf Comput Sci,* 43 (1), 2003, 218–27.
7. Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years". *J. Nat. Prod.,* 70 (3), 2007, 461–77.
8. Von Nussbaum F, Brands M, Hinzen B, Weigand S, Häbich D. Antibacterial natural products in medicinal chemistry--exodus or revival?. *Angew. Chem. Int. Ed. Engl.,* 45 (31), 2006, 5072–129.
9. John Faulkner D, Newman DJ, Cragg GM. Investigations of the marine flora and fauna of the Islands of Palau. *Nat Prod Rep,* 21 (1), 2004, 50–76.
10. Bleicher KH et al. *Nat Rev Drug Discov*, 2(5), 2003, 369-78.
11. Astle JM, Simpson LS, Huang Y, Reddy MM, Wilson R, Connell S, Wilson J, Kodadek T. Seamless Bead to Microarray Screening: Rapid Identification of the Highest Affinity Protein Ligands From Large Combinatorial Libraries. *Chemistry & Biology,* 17 (1), 2010, 38–45.
12. Hann MM, Oprea TI. Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin,* 8 (3), 2004, 255–63.
13. Anonymous 1. http://peds.oxfordjournals.org/content/20/7/327.abstract
14. Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, Hill DP, Kania R, Schaeffer M, Pierre SS, Twigger S, White O, Rhee SY. Big data: The future of biocuration. *Nature,* 455 (7209), 2008, 47–50.
15. Zhang XHD. *Optimal High-Throughput Screening: Practical Experimental Design and Data Analysis for Genome-scale RNAi Research*. Cambridge University Press, 2011.
16. Eisenstein M. Quality control. *Nature,* 442 (7106), 2006, 1067–70.
17. Zhang XHD, Espeseth AS, Johnson EN, Chin J, Gates A, Mitnaul LJ, Marine SD, Tian J, Stec EM, Kunapuli P, Holder DJ, Heyse JF, Strulocivi B, Ferrer M. Integrating experimental and analytic approaches to improve data quality in genome-scale RNAi screens. *Journal of Biomolecular Screening,* 13 (5), 2008, 378–89.
18. Zhang JH, Chung TDY, Oldenburg KR. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *Journal of Biomolecular Screening,* 4 (2), 1999, 67–73.
19. Zhang, XHD. A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays. *Genomics,* 89 (4), 2007, 552–61.
20. Zhang XHD. Novel analytic criteria and effective plate designs for quality control in genome-scale RNAi screens. *Journal of Biomolecular Screening,* 13 (5), 2008, 363–77.
21. Zhang XHD. A new method with flexible and balanced control of false negatives and false positives for hit selection in RNA interference high-throughput screening assays. *Journal of Biomolecular Screening,* 12 (5), 2007, 645–55.
22. Zhang XHD, Yang XC, Chung N, Gates A, Stec E, Kunapuli P, Holder DJ, Ferrer M, Espeseth AS. Robust statistical methods for hit selection in RNA interference high-throughput screening experiments. *Pharmacogenomics,* 7 (3), 2006, 299–09.
23. Brideau C, Gunter G, Pikounis B, Liaw A. Improved statistical methods for hit selection in high-throughput screening. *Journal of Biomolecular Screening,* 8 (6), 2003, 634–47.
24. Cohen J. The Earth Is Round (P-Less-Than.05). *American Psychologist,* 49 (12), 1994, 997–1003.
25. Zhang XHD. A method for effectively comparing gene effects in multiple conditions in RNAi and expression-profiling research. *Pharmacogenomics,* 10 (3), 2009, 345–58.
26. Zhang XHD. Strictly standardized mean difference, standardized mean difference and classical t-test for the comparison of two groups. *Statistics in Biopharmaceutical Research,* 2 (2), 2010, 292–99.
27. Zhang XHD. Assessing the size of gene or RNAi effects in multifactor high-throughput experiments. *Pharmacogenomics,* 11 (2), 2010, 199–213.
28. Agrestia JJ, Antipovc E, Abatea AR, Ahna K, Rowata AC, Barete JC, Marquezf M, Klibanovc AM, Griffiths AD, Weitz DA. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proceedings of the National Academy of Sciences* 107 (9), 2010, 4004–4009.
29. Ethan Schonbrun, Adam R. Abate, Paul E. Steinvurzel, David. A. Weitz and Kenneth B. Crozier. High-throughput fluorescence detection using an integrated zone-plate array. *Lab on a Chip,* 10 (7), 2010, 852–856.