New Paradigms in Drug Design and Discovery

Nouri Neamati*[1] and Joseph J. Barchi, Jr.*[2]

[1]University of Southern California, School of Pharmacy, Department of Pharmaceutical Sciences, 1985 Zonal Avenue, PSC 304BA, Los Angeles, CA 90089-9121, USA

[2]Laboratory of Medicinal Chemistry, NCI/FCRDC, Bldg 376, Rm 209, 376 Boyles Street, Frederick, MD 21702-1201, USA

*Address correspondence to these authors at the University of Southern California, School of Pharmacy, Department of Pharmaceutical Sciences, 1985 Zonal Avenue, PSC 304BA, Los Angeles, CA 90089-9121, USA; e-mail: neamati@usc.edu or Laboratory of Medicinal Chemistry, NCI/FCRDC, Bldg 376, Rm 209, 376 Boyles Street, Frederick, MD 21702-1201, USA; e-mail: barchij@dc37a.nci.nih.gov

Abstract: The new millennium has ushered in an era of science that will revolutionize a great majority of our daily activities. That revolution is being experienced by a growing number of the population who are pushing the average life expectancy closer to the 80-year mark. The primary reason for this increase is the changes we have made in the last 2-3 decades both in how we live our lives as well as how we treat our maladies when they arise. The advent of new techniques in diagnostics and surgery have allowed many to survive debilitating illnesses when their chances would have been slim only a few years ago. In addition, several new therapeutic agents have been developed in the latter part of the 20th century that have improved our quality of life and increased our overall survival time. New medicines to treat cardiovascular, degenerative, infectious, and neoplastic diseases are rapidly being discovered in an effort to further lengthen our lifetimes. The processes used by academic and industrial scientist to discover new drugs has recently experienced a true renaissance with many new and exciting techniques being developed in only the past 5-10 years. In this review, we will attempt to outline these latest protocols that chemists and biomedical scientist are currently employing to rapidly bring new drugs to the clinic.

INTRODUCTION

Remarkable progress has been made during the past five years in almost all the areas concerns with drug design and discovery. A limited survey of the literature reveals no less than 140 review articles published to date with the phrase “drug discovery” in the title or abstract an increase of more than 150% from just five years ago. Hence, it may seem redundant to compile still another, but the pace at which this field has progressed justifies continuous updates of advancements in new techniques and therapeutic successes. Therefore, this review will concentrate on the literature of primarily the past five years, adding historical

information for appropriate perspective. We will try to answer three basic questions concerning the many disciplines related to the drug design and discovery process: 1) What is the state-of-the-art in drug discovery today? 2) What are the latest tools used in the drug discovery process, and 3) Where is drug discovery going in the new millenium? Many very recent reviews have also addressed these questions [1-3].

Before we dwell into nanotechnology, human genome chip technology, ultra-high throughput screening and the like, perhaps it is prudent to highlight some of our achievements from the past and mention specific shortcomings. No attempts will be made to cover any subject in depth, because such efforts are futile (*vide supra*). However, our goal is to make the reader truly aware of what it takes to design and market a novel drug entity.

At the onset it is important to know what features an "ideal" drug should have. First, all drugs must be safe and effective. Second, they should be orally well absorbed and bioavailable. Third, drugs need to be metabolically stable to maintain a reasonably long half-life. Fourth, an ideal drug should be non-toxic and cause minimal or no adverse effects. Fifth, an effective agent will distribute selectively to the target tissue(s). It is oversimplified to state that finding compounds that possess all of these desirable traits is a formidable task. It is no wonder that the average time it takes to introduce a new drug to market from inception to shelf is ca. 12 years. With such a consuming and labor intensive task, why do many drugs still fail as commercial products even after advanced clinical trials and safety testing? The answer is usually a result of an unfavorable absorption, distribution, metabolism, and excretion/toxicity (ADME/T) profile. In the past, it was difficult to almost impossible to predict these characteristics for a specific compound. Drug discovery in the new millennium is armed with not only new and efficient techniques for producing, purifying and screening new entities, but with computing power that was unimaginable a decade ago. Hence, with data compiled from other commercial agents, we can *a priori* predict absorption and distribution properties of lead molecules *in silico* (See for example [4]). In today's high tech world, the winners in war, the stock market or business rely primarily on one thing: information. The more information we can compile and review before a lead is selected for further development, the greater our chances of success. Modern computers give us the opportunity to organize and submit new compounds to a rigorous virtual screening to assess their "druggability" before resources are committed to additional research. This can potentially save pharmaceutical companies, government and academic labs alike from pursuing the "wrong" leads. The investment of

time and resources that can be directed to more promising new agents will allow the lead-to-market time line to shorten considerably in the coming years.

LESSONS FROM THE PAST

It may be useful to offer a very brief summary of some of the historical approaches to drug design and discovery to learn from whence this "art" has evolved. It is impossible to trace the roots of drug discovery to their true origin. Many ancient populations made reports the medicinal properties of various plant extracts and elixirs, all a result of a necessary trial and error search for remedies of specific ailments [5]. Nature has been and still is the single most important source of drugs or drug precursors [6]. Although natural products such as morphine, cocaine, salicylates, atropine, quinine and digitalis are all considered, so to speak, "ancient", in the 21st century, these natural products and their derivatives remain as useful therapeutics even today, in some cases, thousands of years after their original "discovery". So from early civilizations, man has used nature to heal or soothe specific ailments. Unwittingly, the use of extracts and whole plants as remedies amounted to the administration of several chemical entities at once, whose constitution and synergism was wholly unknown. It was not until the 19th century when techniques for partitioning some of these extracts into individual components did single entity drugs become available.

Drug discovery by what is excessively referred to today as "rational" means (as to suggest all other means are "irrational"), probably did not take flight until the first structures of receptors were solved. If we use some poetic license, we then may anoint the first rational drug discovery effort to be made by the discoverers of the receptor concept. These would arguably be John Langley and Paul Ehrlich. In 1897, Ehrlich suggested a theory based what he had called *side chains* or groups on cells that can combine with a particular toxin. Langley had postulated 20 years earlier that alkaloids that caused different salivary flow in cats interacted with specific groups or entities on the nerve endings of the gland cells. Ehrlich actually termed his side chains *receptors*. Without any structural knowledge of the entities transmitting the effect, these may have been the first instances of ligand-receptor interactions observed and partially defined.

Since the early 20th century, thoughts about drug action and mechanism expanded as the analytical techniques of biology, chemistry and pharmacology progressed. Discoveries of different families of therapeutics followed the seminal observations of Ehrlich, and after

1910, a new era in drug discovery emerged. Science saw the development of many drugs discovered hundreds of years earlier. Although quinine was found by early explorers to be used by Indians of South America, it was not isolated until 1823 and development of analogues as antimalarials began in the early 1900's [7]. New medicines such as antihistamines, trypanocidals and several important alkaloids, many extracted in the 19th century, were being synthesized and developed into commercial entities. The age of antibiotics began just prior to that famed serendipitous discovery of a crude penicillin culture by Fleming, with the discovery that a dyestuff (Prontosil) could cure gram-positive bacterial infections in man (see for example [8]) The active component, sulfanilamide, paved the way for the development of sulfa drugs. The intensive study of Fleming's original molds by Florey and Chain in the early 1940s showed that there was a mixture of several components in the penicillin preparation and these were separated, tested and more active constituents were found and developed into the first anti-infectious agent. Around this same time, more extensive development of antihisatmines, analgesics, barbituates, hormones (e.g., epinephrine), sedatives, hypnotics and antidepressants was seen in the 1940s-1950s. The improvement in chromatographic and diagnostic (detection) techniques as well as advances in synthesis and understanding of chemical principles accelerated the discovery of new drug entities in the second half of the 20th century. Another case of serendipity led to the discovery of Librium in 1957 [9] and later to the benzodiazepines class of antianxiety medications, which include Valium and Xanax. Valium was once the best-selling prescription drug in America. In addition to small molecule therapeutics, the 20th century saw the rise and success of vaccines to cure several bacterial diseases such as tetanus, diptheria, yellow fever, measles, mumps, rubella and polio. Diagnostic techniques such as X-rays, electrocardiograms, CT and PET scans, ultrasound and MRIs were all products of the last 40-50 years, and each technique played its own role in the design and development of new drug entities.

PRESENT STATE-OF-THE-ART

We have come a long way since the days of extracting roots and chewing leaves were our only medicinal preparations. But there is still a host of serious diseases for which we have no useful therapies available. Two notable examples, cancer and AIDS, have resisted therapeutic intervention save for certain select blood-borne tumors. In this section we will attempt to highlight the recent changes that have shifted the paradigms of drug discovery to what could be called the second "golden age" of therapeutic design. Our goal is to list and expand upon

the various steps that are being followed today when initiating a new drug discovery program. Along the way we will focus on the more novel and exciting techniques that we feel will have impacted the drug discovery field most powerfully in recent years.

The past decade has seen the pharmaceutical industry and related academic programs in drug discovery transform their strategies based on the influx of modern techniques to discover, screen, modify, and optimize potential drug entities. Due to the recent advances in molecular biology, robotics and microelectronics, and, more importantly, the complete analysis of the human genome, modern drug discovery will make everlasting impact on human diseases. Technologies such as genetic diagnostics, identification of novel therapeutic targets and new innovative therapies combined with latest methodologies in bioinformatics will allow biological information to be analyzed and managed on a very large scale and in an extremely short time. For example, the ability to measure global changes in gene expression before and after treatment using microarray technologies is important for genome wide mapping for assigning disease gene loci and monitoring kinetic expression of disease genes. As a consequence, rapid evolution of high-throughput techniques in genomics is leading to an enormous increase of information. New computer algorithms designed to rapidly analyze the wealth of data being generated by the millions of compounds produced by combinatorial and parallel synthetic techniques are assisting in the intellectual review of this data and guiding future directions of design [10].

We have bandied about an ominous list of new terms in the previous paragraph. With all due deference to those who need no further explanation, we thought it wise to present a glossary of the most recent terminology in drug design and discovery. The following sections will describe what we feel are the five major areas of modern drug discovery and design programs. They are, 1) Target Identification, 2) Target Validation, 3) Lead Identification, 4) Lead Optimization and 5) Preclinical Pharmacology and Toxicology. We will stop short of a drug progressing to clinical trials as we wish to limit our discussion to the discovery and design phases of the process. We include area 5 since this phase may reveal deficiencies in a lead that may force it to be recycled back into the design phase (lead optimization).

TARGET IDENTIFICATION

Traditional drug discovery began with a known pathological phenomenon in an organism and the development of a therapeutic theory to combat this process. A chemical concept would

follow to produce compounds for screening. Most of these processes originated with the understandings of some biological pathways and screening for an effect in tissues or cells. This may or may not eventually reveal a "target". Conventional approaches of identifying targets such as protein expression, protein biochemistry, structure function studies, knowledge of biochemical pathways, and genetic studies were instrumental in drug development (see the review by Kan in this issue). In the "omics" climate of today, genetic information is now guiding the identification of single molecular targets. These are derived from knowledge of the genes of specific cell phenotypes that encode proteins that may be involved in the pathogenesis of a particular disease state. The ability to sequence a genome and identify every expressed gene will lead to the identification of thousands of new targets, many of which will be relevant to the onset and persistence of disease. With the advent of proteomics and high throughput protein profiling information we will eventually reveal the role, function, structure, gene location, biochemical pathway, molecular interactions, and expression levels of each and every protein coded for by a particular genome. Therefore, the impact of recent progress in molecular biology and, in particular, of genomic sciences on drug discovery will change the course of this field remarkably. In fact, at present in most major pharmaceutical companies, 10% to 25% of new discovery projects are based on genomics [11].

There are several ways to use gene analysis to identify specific molecular targets [12]. Some of the new standard procedures for target discovery are high throughput sequencing analysis, positional cloning, the generation of cDNA libraries with expressed sequence tags (ESTs), database mining by sequence homology and mining by differential tissue expression. Profiling the protein products of these mined genes, the discipline known as proteomics, is also facilitating the identification of novel protein structures and families that may be relevant to a diseased state. As with gene profiling, protein profiles from diseased and healthy tissue can be fractionated and compared to examine the different proteins expressed in diseases phenotypes. Sequences can be aligned with known protein sequences and each protein can be subjected to mass spectral analysis to determine if those molecules expressed in diseased tissue correlate with known families of molecular targets. An undeniably powerful method for studying differential gene expression is the use of microarrays (biochip technology). These allow for the rapid analysis of the expression of thousands of genes. The genes can be either cloned or synthetic. Through the use of special chemical techniques and linkers are

used to attach the oligonucleotides to glass slides to form arrays. These are then hybridized with cDNAs from some particular tissue or cell type. In a typical example, a fluorescent detection system allows for the quantitation of interaction of the cloned gene with the cDNA. In this manner, gene expression patterns for many different animal tissues can be obtained under different experimental conditions. For example, the cDNA libraries can be derived from cells that are control or drug treated, and hence the gene profile of a disease tissue under the stress of a toxin or inhibitor can be directly compared and correlated. It follows (but is not always correct) to designate specific proteins encoded by those genes more highly expressed in the diseased state to be a potential target for therapy.

It is obvious then that with these technologies and the complete map of the human genome, there will be no shortage of new targets to be evaluated in a drug discovery setting. The problem arises in determining whether these novel targets are worthy of blocking: are they actually relevant to the physiology of the diseases state?

TARGET VALIDATION

Selection and validation of novel molecular targets have become of paramount importance in light of the plethora of new potential therapeutic drug targets that are continuously being discovered. The prospective targets identified in the previous section require confirmation that intervening at this step in a particular pathway will effect an appropriate biological response. There are several approaches to this, but there remain drawbacks (see below). The use of reliable animal models and the latest in gene targeting and expression techniques are all, essential to the validation process. The following constitute some of the more widely used methods of target validation. This process has also recently been reviewed (for a recent review see [13]).

Targeted gene disruption (TGD) can be considered a catch-all term for several different methods of target validation. TGD generally relates to the production of knockout or transgenic animals to study the effect of removing a particular gene coding for the putative molecular target. Conceptually, this approach seems worthwhile, since the system under study is an intact higher organism that may correlate well with a similar disruption in the intact human. However, there are compensatory mechanisms in any organism that may invalidate the intended result of the knockout. In addition, the time it takes to produce and analyze these animals is rate limiting in today's world of rapid turnover. Thus, results of

these experiments need to be critically reviewed. However, the power of knockout mice and gene targeting is undeniable for the *in vivo* validation of potential drug targets.

Other strategies to specifically deplete cells of a specific protein such as antisense [14,15] and ribozyme [16,17] technology offer a more specific and "targeted" approach to target validation. Both technologies utilize molecules to hybridize to mRNA and either prevents the expression of the protein product of that RNA through induced degradation via Rnase (antisense) or catalytic cleavage of the mRNA by the designed molecule itself (ribozyme). In particular, the ribozyme approach seems to be promising in cases where the protein product turnover is relatively rapid. This technology was shown to reduce the cellular content of HER-2/neu (a proto-oncogene found in many cancers, in particular tumors of the breast) mRNA and protein by greater than 90% whereas an antisense approach was only capable of reducing the expression by 50% [22]. Aptamer technology makes use of DNA molecules that bind specifically to proteins and can be employed to inhibit the function of a specific protein or as competitive binders of small molecule drugs to target proteins [18]. The synthetic demands of constructing these molecules can be limiting for their general use. Intrabodies, intracellular antibody constructs prepared by recombinant methods, are novel approach to neutralize specific molecular targets in the manner of monoclonal and polyclonal antibodies [19]. These constructs may be introduced into cells via cDNAs encoding the intrabody to directly neutralize intracellular targets. Intrabody therapy has proved successful when targeted to erbB-2 in breast cancer cells by reducing the surface expression of this protein and subsequent induction of apoptosis [20].

Recently, a new approach to validation using specific peptide binders to a potential pathogen target was reported. In this study, peptides were selected by phage display or combinatorial screening based on their binding to prolyl-tRNA synthetase, an essential enzyme in the bacterial life cycle of *E. coli*. This peptide was inducibly expressed in the pathogenic cells and injected into animals who were infected with a lethal dose of bacteria. The animals receiving the peptide inhibitor were rescued in five out of five cases. This approach to validation can be generalized and has the potential to become a valuable tool in the drug discovery process [21]

The novel approaches to drug discovery that have emerged in the past decade are part of the reason that the target validation step remains the bottleneck in the discovery process. The

progression of rapid, high throughput technologies in the areas of target discovery and lead identification (vide infra) will inundate the community with targets and compounds, the trick is still to prove the therapeutic value of modulating these targets in an animal system. Obviously, the advents of genomics and proteomics have created exciting new paradigms for drug discovery (see the review by Hatanaka and Sadakane in this issue). Two related disciplines derived from the genomics era that will contribute to the supply of available small molecule drug candidates are chemogenomics [11] and chemical genetic [22]. These approaches to drug discovery are becoming new alternatives to traditional methods discussed in this review. Chemogenomic has been defined as the discovery and description of all possible drugs to all possible drug targets and chemical genetics entails using defined chemical probes to dissect specific features biology and can be viewed as a subset of chemogenomics [11]. With families of gene products being assaulted on a medicinal chemistry front, it remains to "automate" the validation process to keep pace with the identification phase.

LEAD IDENTIFICATION

A lead is defined as a compound (usually a small organic molecule) that demonstrates a desired biological activity on a validated molecular target. To fulfill the criteria of what the industry considers a useful lead, the compound must exceed a specific potency threshold against the target (e.g., < 10 μM inhibition). The compounds used as potential leads could come from many sources. A majority of leads discovered in very recent programs are derived from a collection that is now referred to as a "library". These may take the form of natural product libraries, peptides libraries, carbohydrates libraries, and/or small molecule libraries based on a variety of different molecular scaffolds. Today, many libraries are commercially available or open to the public. Most pharmaceutical companies house their own compilation of compounds that have been synthesized over several years and screened against a variety of targets. These may be used for random screening of new targets in the discovery phase in an effort to define a scaffold to which that target (of unknown molecular structure) may bind. Many libraries have been synthesized *de novo*, either rationally, based on sequencing or structural knowledge of the active site or the catalytic domain of the target or in a more random manner. Today, the identification of a useful lead relies heavily on the sampling of an appropriate amount of conformational space when screening against a new target. This issue of exploring the correct mix of diversity space while maintaining "drug-like" qualities is

critical to the development stage of any new drug and is a central thrust of large drug discovery programs today [23]. The applications of combinatorial chemistry to produce small molecule libraries [24,25] and the growing ability to master more complex chemical conversions on the solid phase have placed increased attention on the generation of more drug-like libraries. Thus, some of the early limitations of the chemistry and concerns about structural diversity have given way to design principles and predictive algorithms that should increase the drug-likeness of large "virtual" libraries (see below). This should serve to improve the probability that members in the final construct will have desirable biophysical properties (e.g., solubility, oral absorption or CNS penetrability).

Each lead must be screened by an appropriate assay against the molecular target. This stage of drug discovery, depending on the target, can take as little as one day to as long as several years. An unfortunate characteristic of this phase is its high rate of failure with the percent hit rate routinely between 0.01-1%. Any one program could screen as few as 10 or as many as 10^6 compounds before the activity threshold is met. On the plus side, in the ultrahigh-throughput screening circles that we now move in, screening 10^6 compounds may only take a few days to weeks as opposed to 10^6 days. Incredibly, the days of screening compounds in 96-well plates are already behind us- thanks to combinatorial chemistry technologies for creating mega-libraries of compounds and the fully automated robotics instrumentation with the capacity to screen 0.2-1 million compounds per day. Different manufacturers have been developing instrumentation capable of handling multiple micro titer plate formats on the same platform using 384 and 1536-well plates. High and ultra-high throughput screening techniques have gone through a major revolution in optics and instrumentation that allow for faster and more accurate assay data. The variety of fluorescent probes available that allow detection of substrates in the picomolar range, assures that most positive hits will be observed. Technologies used for the bioassay endpoints such as ELISA, fluorescent-based calorimetric assays and scintillation proximity assay (SPA) have all been reviewed and will not be discussed here (see the review by Shoemaker et al in this issue). Advances in small volume liquid dispensing and pipetting, reliable handling of standardized plates and simplified assay formats all have made an impact on the reliability of the HTS process (see the review by Shoemaker et al in this issue).

Most biological assays are based on affinity screening of small molecules to the target and use physical properties such as UV or fluorescence as the endpoint detection of binding (see

the review by Hatanaka and Sadakane in this issue). However, there are alternative screening technologies that may be useful in specific contexts. Many exciting new models for lead discovery have recently emerged that facilitate more rapid identification and result in compounds whose physical and "druglike" properties have been initially optimized. Utilizing the information we have gleaned over the years of what a drug looks like and how it behaves in solution, most discovery programs will pass potential compounds through various "filters" in an attempt to discard those hits that may fail based on an undesirable "druggability " profile. In the next paragraphs we will briefly outline six major areas that have revolutionized the identification phase: 1) Virtual screening, 2) Informatics, 3) Advances in pharmacaphore mapping, (viz., database searching, modeling), 4) High throughput docking, 5) NMR-based screening and 6) Chemical genetics..

VIRTUAL SCREENING

A major breakthrough in lead identification in the recent years occurred with the availability of fast and cheap computers on one hand and commercially available databases of compounds with more than a million small molecules. This resulted in virtual screening technologies using high throughput docking, homology searching and pharmacophore searches of 3D databases. This is perhaps the cheapest way to identify a lead and several cases have already proven successful using this technology. Although this technique is essentially based on concepts that have been used for many years by those in molecular modeling, the introduction of more powerful computers has paved the way for the virtual screening of ever growing databases of compounds. As described by Walters [26], some important features to consider when developing a virtual screening system are, 1) knowledge about the compounds that you may screen against your receptor, 2) knowledge about the receptor structure and receptor-ligand interactions in general and 3) standard knowledge about drugs and drug characteristics.

A key requirement for a successful virtual screening is accession to a large and a diverse library or a database. It has been argued that a good database is one that is the most diverse when it comes to its sampling of chemical space. Among all possible libraries, natural products collections arguably represent the highest degree of chemical diversity. Natural products are often extolled as sources of drug leads however, frequently occurring natural product motifs are seldom found in drugs [27]. Therefore, although diversity is critical, it is

highly desireable to design a focused virtual library that contains synthesizable and drug-like compounds rather than one that maximally samples diversity space [26]. In fact, as recently reviewed, the world of drug-like compounds is limited in that there are currently only about 10,000 drug-like compounds, which are sparsely, rather than uniformly, distributed through chemistry space [27]. Moreover, when analyzing the properties of known drugs (drug database mining) based on molecular framework or shape and side chain data, it was found that the diversity of shapes and side chains in the set of known drugs is extremely low [28,29]. Another study by Wang employed the concept of multilevel chemical compatibility (MLCC) scoring to measure drug-like characteristics. This systematic comparison of the local environments within a compound and those within existing commercial drugs was applied to four test sets: top selling drugs, compounds under biological scrutiny prior to preclinical testing, anticancer drugs, and compounds known to have poor drug-like character. Most of the top drugs were predicted to be drug-like, about one-quarter of the compounds under testing were drug-like, and about one-fifth of the anticancer drugs were drug-like. The method also correctly predicted that none of the known problematic compounds were drug-like. It was also suggested that the current drug library contains about 80% of all the known viable drug types. The authors believe that this result argues that there is only a low probability that the discovery of a drug will expand the chemical space of that already sampled by existing drugs [30].

What is evident from the studies cited is that optimal virtual screening strategies rely on the availability of chemical libraries that are as diverse as possible yet constrained in favor of compounds possessing attributes that are normally associated with successful drug candidates. A useful algorithm describing an effective means to increase structural diversity in a chemical library while maintaining a bias toward compounds that retain the desirable properties of drugs has been previously reported by Koehler [31].

INFORMATICS

The unprecedented flood of information from genome sequences and functional genomics in one hand and combinatorial chemistry, HTS, and virtual screening on the other hand has given rise to new fields of bioinformatics and chemoinformatics, which combines elements of biology and chemistry with mathematics, statistics and computer sciences. Analyses in bioinformatics and chemoinformatics predominantly focus on several types of large datasets

available such as macromolecular structures, genome sequences, 3D chemical databases and compound libraries. Informatic methodolgies rely on a variety of computational techniques [10,32] including sequence and structural alignment, database design and data mining, macromolecular geometry, phylogenetic tree construction, prediction of protein structure and function, gene searching and expression data clustering, chemical-similarity clustering, diversity analysis, library design, virtual screening and QSAR (for a recent review see [32]). Combinatorial chemistry and HTS primarily depend on chemoinformatics to increase their effectiveness. Recent advances in chemoinformatics include new molecular descriptors and pharmacophore mapping techniques, statistical tools and novel visualization methods [33]. A major task of informatics in the future is to develop software tools that provide the means to store, extract, analyze, and display data in a way that chemists can easily understand and appreciate [34]. In attempts to decipher chemical/biological information, computers require the use of descriptors. Hence, hundreds of molecular descriptors have been reported in the literature, ranging from simple bulk properties to elaborate three-dimensional formulations and complex molecular fingerprints. A number of studies have been reported that investigate the performance of molecular descriptors in specific applications [35,36]. For example, a previous report used a combination of 65 preferred 1D/2D molecular descriptors and 143 single structural keys for their performance in compound classification based on biological activity. The analysis was based on principal component analysis of descriptor combinations and facilitated by the use of a genetic algorithm and different scoring functions. In these calculations, several descriptor combinations with greater than 95% prediction accuracy were identified. A set of 40 preferred structural keys was incorporated into a small binary fingerprint designed to search databases for compounds with biological activity similar to the query molecules. Thus, although the design of mini-fingerprints is conceptually simple, they perform well in activity-oriented similarity searching [35,36]. Moreover, similarity searches based on chemical descriptors have proven extremely useful in aiding large-scale drug screening [37,38].

THREE-DIMENSIONAL PHARMACOPHORE MAP-PING

The 3D pharmacophore search is an important, robust and a facile approach to rapidly identify lead compounds against a desired target. Traditionally, a pharmacophore is defined as the specific 3D arrangement of functional groups within a molecular framework that are necessary to bind to a macromolecule and/or an enzyme active site. The designation of a

pharmacophore is the first essential step towards understanding the interaction between a receptor and a ligand. Once a pharmacophore is established, the medicinal chemist has a host of 3D database search tools to retrieve novel compounds that fit the pharmacophore model. The search algorithms have evolved over the years to effectively identify and optimize leads, focus combinatorial libraries and assist in virtual high-throughput screening. Thus, this technology has been clearly established as one of the successful computational tools in modern drug design [39,40]

Numerous advances have been made in the computational perception and utilization of pharmacophores in drug discovery, database searching and compound libraries. For example, a hierarchical set of filtering calculations has emerged that can be used to efficiently partition a library into a trial set of pharmacophores. This sequential filtering permits large libraries to be efficiently processed, as well as analyze the compounds discovered as hits in great detail. Additionally, new and extended methods of QSAR analysis have evolved to translate pharmacophore information into QSAR models that, in turn can be used as virtual high-throughput screens for activity profiling of a library [41]. Moreover, a successful application of fingerprinting approach was previously employed to generate 10,549 three-point pharmacophores by enumerating several distance ranges and pharmacophoric features. Subsequently, the fingerprint was used as a descriptor for developing a QSAR model using partial least squares [42]. Recently, a more general concept of descriptor pharmacophore was introduced, which uses variable selection QSAR as a subset of molecular descriptors that afford the most statistically significant structure-activity correlation. These methods include partial least squares and K-nearest neighbors. Therefore, chemical similarity searches using descriptor pharmacophores yields efficient mining of chemical databases or virtual libraries to discover compounds with a desired biological activity [43].

The ever-expanding list of pharmacaphore search algorithms have been designed on a variety of platforms with diverse search criteria. One class of the so-called “genetic” algorithms mimics some of the major characteristics of Darwinian evolution such that small organic molecules could satisfy QSAR-based rules (fitness). The algorithm takes an initial set of fragments and iteratively improves them by means of crossover and mutation information that are related to those involved in Darwinian evolution [44,45]. Other pharmacophore algorithms are being designed for screening huge virtual combinatorial libraries of diverse compounds. A recently reported example of such an algorithm touts its ability to build and

screen libraries of ca. 10^{18} 3D molecular conformations within a reasonable time scale. The algorithm can potentially be used to design new molecules that display a desired pharmacophore on predefined sets of chemical scaffolds [46]. In yet another example, a 4-point pharmacophore method for molecular similarity and diversity was used for the design of combinatorial libraries for 7-transmembrane G-protein-coupled receptor targets. Up to 7 features and 15 distance ranges were considered, yielding up to 350 million potential 4-point 3D pharmacophores/molecule. The resultant pharmacophore fingerprint serves as a powerful measure for diversity or similarity and for the design of focused/biased combinatorial libraries [47]. For a recent application of 3-D pharmacophore fingerprints as molecular descriptors for similarity and diversity applications such as virtual screening, library design and QSAR see [48].

HIGH-THROUGHPUT DOCKING

Docking is simply referred to the ability to position a ligand in the active or a designated site of a protein and calculate specific binding affinities. Ligand-protein docking has evolved so remarkably throughout the past decade that docking single or multiple small molecules to a receptor site is now routinely used to identify ligands. Optimal docking procedures need to be fast, generate reliable ligand geometries, rank the ligand conformation correctly (scoring), and thereby, estimate the binding energy. A number of studies have shown that docking algorithms are capable of finding ligands and binding conformations at a receptor site close to experimentally determined structures (see below). These algorithms are equally applicable to the identification of multiple proteins to which a small molecule can bind. The application of this approach may facilitate the prediction of either unknown and secondary therapeutic target proteins or side effects and toxicity of particular drugs [49]. In computational structure-based drug design, the evaluations of scoring functions are the cornerstones to the success of design and discovery. Many approaches have been explored to improve their reliability and accuracy, leading to three families of scoring functions: 1) force-field-based, 2) knowledge-based, and 3) empirical [50]. For example, using different docking methods and various scoring functions it was shown that consensus scoring and free energy grids improved hit rates from docking databases of 3D structures into proteins [51-53]. Recently, a World Wide Web accessible database, the Ligand-Protein DataBase (LPDB) was designed to gather protein complexes with both high-resolution structure and known experimental binding affinity [50]. A multidimensional selection of ligand conformations and scoring of protein-

ligand binding affinities were used to dock a series of inhibitors on three matrix metalloproteinases. The selected ligand conformations were found to be very similar to the experimentally determined ligand conformation [54]. In a study evaluating different docking/scoring programs using protein-based virtual screening of chemical databases, a two-step protocol was proposed. First, screening of a reduced database containing a few known ligands is highly recommended for deriving the optimal docking/consensus scoring schemes. Second, these latter parameters are then used to screen the entire database [55]. A similarity-driven approach to flexible ligand docking has also been reported. Given a reference ligand or a pharmacophore positioned in the protein active site, the method allows inclusion of a similarity term during docking [56,57]. Yet another approach is the docking of molecular fragments to a rigid protein and evaluating the binding energy. For example, polar fragments are docked with at least one hydrogen bond with the protein while apolar fragments are positioned in the hydrophobic pockets [58].

The future trend in this field will be docking and virtual screening of multiple combinatorial libraries against a family of proteins. An example of such a method consists of three main stages: 1) Docking the scaffold, 2) selecting the best substituents at each site of diversity, and 3) comparing the resultant molecules within and between the libraries. Referred to as the "divide-and-conquer" algorithm for side-chain selection, this technique provides a way to explore large lists of substituents with linear rather than combinatorial time dependence [59]. An earlier example of the divide-and-conquer strategy in flexible ligand docking used a grid-based method to sample the conformation of an unbound ligand to select low-energy conformers. Rigid docking is then carried out to locate the low-energy binding orientations for these conformers. These docking structures are subsequently subjected to structure refinement including molecular mechanics minimization, conformational scanning at the binding site and a short period of molecular dynamics-based simulated annealing [60]. Multivariate relationships are observed in docking scores computed for a constant set of ligands in different binding sites of proteins that are dissimilar in structure and function. The structural basis for the correlations found among scores is analyzed in terms of size, shape and charge characteristics of the binding sites considered [61].

In brief, high-throughput docking for lead generation [62] if combined with rapid clustering analyses (see for example [63-65]) can greatly speed up the drug discovery processes. For a

review on different methods of virtual screening as a tool for lead structure discovery see [66]
.

NMR-BASED SCREENING

Nuclear Magnetic Resonance (NMR) spectroscopy, the technique that has been the chemist's gold standard for compound identification and conformation analysis, has a long history of involvement in drug discovery and design. Along with X-ray crystallography, NMR has been used to determine the preferred 3-dimensional disposition of potential drug candidates (small organic molecules) as well as to reveal the tertiary structures of the biomacromolecules (proteins and DNA) that interact or are inhibited by the drug entity. NMR-based methods have been successfully employed to screen small-molecule binding to proteins without any prior knowledge of the function of the target [67,68]. NMR can be used to screen very weak binders since global changes in the NMR events that are perturbed when a small molecule binds to a macromolecule can be detected by observing either the ligand or the receptor. Changes in properties such as molecular diffusion and relaxation can be detected when binding events occur even if the binding constants are in the millimolar range. Since small and large molecules have very different diffusion or relaxation properties, the use of special experiments allow the determination of these differences between the free ligand and the protein-ligand complex (bound form). Recently, the use of robotics, high capacity sample changers and flow probes for performing "tubeless" experiments in a high throughput mode has anointed NMR as a method of choice for the analysis and deconvolution of mixtures from split-and-mix combinatorial synthesis since the spectral parameters display separate features for different components in the mixture. NMR comes with the added bonus that structural data is automatically available from the spectrum observables (chemical shifts, coupling constants, etc.). The use of nanoprobes for detecting smaller samples of lower concentration and cryoprobes that add a 3-4 fold sensitivity enhancement are allowing NMR to enter realms of drug discovery that were not available with older technologies. A further boon to the structural analyses of combinatorial libraries has been the use of solid state NMR to screen synthetic samples prepared by solid phase synthesis while still attached to the resin [69].

A host of methods have recently been developed for use in NMR drug screening [70]. Some of these include diffusion ordered spectroscopy (DOSY[71]), saturation transfer difference [72], NOE pumping [73] and SAR by NMR [74]. The SAR by NMR technique was the

seminal method that proved that detection of chemical shift changes in 2D HSQC ^{15}N-^{1}H spectra could guide the design of small molecule with enhanced binding to the target protein [75,76]. This method has been extended through the use of CryoProbe technology to screen up to 200,000 compounds per month in mixtures of 100 entities per experiment [77]. Deconvolution of the mixtures requires more experiment time, but does not add considerably to the screening process. SAR by NMR is still widely used, but the need for NMR assignments of the labeled target proteins limits the initial robustness of the method. An intriguing and successful method developed by researchers at Vertex Pharmaceuticals known as the SHAPES strategy, employs a limited but diverse library of fragments from known drugs or compounds with drug-like properties along with those from protein binding molecules that are screened for binding a target (targets that are generally too large to analyze structurally by NMR) [78]. Weak binding "shapes" are screened by 1D line broadening or 2D transferred NOE measurements and used as lead scaffolds in library design and high throughput screening. Some more advanced techniques such as WaterLogsy (which uses the bulk water on a protein surface as the magnetization to be transferred to the binding ligand [79]) have also been employed to screen SHAPES libraries.

CHEMICAL GENETICS

Schreiber and colleagues at the Institute for Chemical and Cell Biology (ICCB) at Harvard University have described a method of perturbing biological systems with small molecules that they call "chemical genetics" [22,80], which has been defined as "the study of gene-product function in a cellular or organismal context using exogenous ligands" [81]. Several groups have used this approach to identify compounds that may "induce a specific cellular state" [81-85]. The ultimate goal is to discover compounds that may act as "knockouts", i.e. inactivate a gene product (protein) akin to using mutant mouse models, and be able to study the kinetic effects of the particular gene inactivation within the organism. The molecular targets initiative at the National Cancer Institute is already gearing up to create a repository of chemical entities with known protein binding data along with information about their effect on metabolic pathways and/or phenotypical changes in a cell [86]. The chemical genetics approach has worked with selected systems to identify small molecule modulators of cellular function. Foremost in the literature is the discovery of the cell cycle-arresting agent monastrol, an agent that halts cells in mitosis with monopolar spindles [87]. It was shown that this simple molecule inhibits the motility of the kinesin motor protein Eg5, a protein

necessary for spindle bipolarity. This novel finding was particularly significant since all previous mitotic-arresting compounds affect tubulin [88]. Other small molecule tools have been discovered for chemical genetic investigations, namely novel pyridopyrimidine inhibitors of the cell cycle in leukemic and breast cancer cells [89,90] and splitomicin, an inhibitor of the SirP2 histone deacetylase in yeast [91] (for recent reviews see for example [83,92]).

With the production of large libraries of small molecules and entire genomes coding thousands of proteins to screen them against, new techniques for high throughput screening for chemical genetics studies are in high demand [93,94]. The ICCB has also addressed this issue by developing a small molecule microarray system for analyzing protein-ligand interactions on glass slides. This method, called small molecule printing [95], uses robust coupling chemistry to attach small molecules from split-pool (one bead, one compound) synthesis onto glass slides in a microarray of spots separated by 300 mm. Each spot can be probed with different proteins that are tagged for fluorescence detection. As many as 10,800 spots on a single slide may be screened at one time. This elegant variation on cDNA microarrays was further extended to print proteins on microarrays to study protein-protein interactions, identify protein targets of small ligands and identify substrates of enzymes (kinases) [96] (see also [92,97]). This technique involves the plating of proteins on slides impregnated with an aldehyde-containing silyl reagent. Amines on the protein side chains and terminus form Schiff bases with the carbonyl group. In this way, different orientations of the proteins are displayed for interactions with binding partners. The test systems used proved that the proteins are properly folded in the context that they are displayed at each position on the slide. These very powerful techniques will have a lasting impact on the modern drug discovery process. This group of experiments constitutes a bridge between target identification and validation with lead identification and optimization in the discovery and design of novel chemical entities that modulate cellular functions.

Various techniques that are used in the lead discovery phase also play key roles in the optimization of the newly found lead. The next section will briefly describe some of the more pertinent of these approaches. See the reviews by Ohkanda *et al.* and DeJong *et al.* in this issue for additional examples of how target structural information and combinatorial methods have been successfully employed in the discovery of potent leads for preclinical development.

LEAD OPTIMIZATION

Once a lead compound is established in the identification process, the medicinal chemist will work closely with molecular pharmacologists to optimize the desirable traits of the lead. This process can be relatively fast since history has taught the medicinal chemistry community how to manipulate molecules to improve activity. Starting with intuitive structural modification to the development of structure-activity relationship (SAR) and quantitative SAR (QSAR) one can gain tremendous information. As explained throughout this review, these approaches have been modified remarkably and the chemist now has a plethora of resources under his/her disposal before the actual synthesis begins. In addition, computer-aided drug design (CADD) or structure-based drug design (SBDD) has made a considerable contribution to the field of drug candidate optimization, and has been the subject of numerous reviews and books (see for example [98,99]). It is also important to bear in mind that the synthesis of focused chemical libraries using parallel synthesis can facilitate lead optimization. Iterative optimization of lead compounds necessitates a broad knowledge in the general principles of *de novo* drug design. There are many tools for characterization of binding sites: Calculation of charge distribution, lipophilicity or pKa of side-chain functionalities and identification of H-bond donors and acceptors. In addition, docking programs are used in conjunction with large 3D databases of small molecule structures and the scoring algorithms that attempt to predict the binding affinity of designed ligands. To be considered for further development, lead structures should be amenable for chemistry optimization and have good ADME properties. The following properties can be easily estimated: Molecular weight, the calculated molecular refractivity, the number of rings, the number of rotatable bonds, the number of hydrogen bond donors and acceptors, the calculated logarithm of the n-octanol/water partition (ClogP) and the calculated logarithm of the distribution coefficient at pH 7.4 (LogD). In general, lead structures exhibit less molecular complexity (less MW, less number of rings and rotatable bonds) and are less hydrophobic (lower CLogP and LogD) [100]) than non-lead compounds. The activity of a drug is the result of a multitude of factors such as bioavailability, toxicity and metabolism [101].

STRUCTURE-BASED DRUG DESIG

Structure-based design is considered as one of the most innovative and powerful approaches in drug design and is most effective when the 3D structure of an existing inhibitor in complex with its target is known. This technique has played a major role in the design to number of drug candidates that have progressed to clinical trials. A prerequisite for this approach is an understanding of the principles of molecular recognition in protein-ligand complexes. SBDD is an iterative approach. It requires the 3D structure of the target protein, preferentially complexed with a ligand, where binding mode and affinity and conformation of a ligand binding can be discerned (see the reviews by Kan and Li and Roller in this issue). Subsequently, various methods are used to design a high affinity inhibitor either via virtual computer screening of large compound libraries or through design and synthesis of novel ligands. Designed compounds are then tested in appropriate assays and the information is further used to guide the SBDD. Recent advances in computational methods for lead discovery include various commercially available softwares for *de novo* drug design, iterative design, selectivity discrimination, and estimation of ligand binding affinities (see for example [98,102,103]).

Methods to accurately find binding sites for small molecules on the surface of a protein (see Hatanaka and Sadakane in this issue) and the use of the data emerging from structural genomics are paving the way to develop designer drugs [104,105]. Two approaches to SBDD, the docking of known compounds into a target protein and *de novo* drug design has been merging as a single robust and powerful tool [99]. In addition, dynamics simulation of multiple copies of molecular building blocks in the presence of a receptor molecule is also a useful strategy for drug design [106]. This technology was crucial in designing a series of antiviral and anticancer drugs that were designed from knowledge of the molecular structure of their target enzyme (see for example [107,108]).

In the future SBDD will merge with high throughput and informatic technologies to design drugs against multiple homologous targets simultaneously. For example, grouping potential drug targets into families based on common cross correlations of their SARs, provides a means to translate the information from genome-sequencing efforts into knowledge that will aid in the discovery of drugs [109]. Due to the abundant sequence information available from genome projects, an increasing number of structurally unknown proteins, homologous to examples of known 3D structures, will be discovered as new targets for drug design. When homology models do not provide sufficient accuracy to apply common drug design tools, a

recently developed approach called DragHome, may be used to dock ligands into such approximate protein models. DragHome combines information from homology modeling with ligand data, used by and derived from 3D QSAR [110]. However, linear stretches of sequences ("receptor-binding domains", RBDs) can be identified by analyzing hydrophobicity distributions in the absence of any structural information. In a recent study, RBDs were predicted from the 80,000 sequences of the Swissprot database. This procedure could detect residues involved in specific interaction sites such as specific DNA-binding or Cacium-binding domains. Therefore, this method is useful for predicting protein interaction sites from sequences and may guide experiments such as site-specific mutageneses or the synthesis of more potent inhibitors [111].

PREDICTING DRUG-LIKE PROPERTIES

The phrase “drug-like” is defined as those compounds that have sufficiently acceptable ADME and toxicity properties to survive through the completion of Phase I clinical trials [27]. To build a drug-like database, it is essential to apply a variety of filters to remove useless compounds such as those which contain reactive groups and exhibit false positive in a majority of assays (see for example [23,112-115]). Further filtering will depend on the type of target and project, where for example bioavailability, pharmacokinetics (PK), or CNS penetration may dictate the requirements for the target “drug-like” molecules (For recent reviews see [116-118]).

It is becoming clear that successful prediction of drug-like properties at the onset of drug discovery will payoff later in drug development. Therefore, there is increasing demand to design computer programs that can accurately predict physicochemical parameters [119]. Such parameters include oral absorption, blood-brain barrier penetration, toxicity, metabolism, aqueous solubility, logP, pKa, half-life, and plasma protein binding [23,27,120-123]. It is important to mention that the current level of automation using capillary electrophoresis techniques to experimentally determine pKa and log P coupled with flow injection analysis with UV detection to determine solubility and assess chemical stability of compounds at various pH’s supports the measurement of these properties for ~100 compounds per week [124].

A major impetus for a successful drug design strategies is to invest in *in silico* techniques with effective and reliable algorithms to predict oral bioavailability and avoid compounds

that do not meet safety requirements [125]. Therefore, *in silico* algorithms are based on "drug like" properties of known drugs such as a required molecular weight range, optimum H-bond donor and acceptor numbers and desirable log P values. As the information on the structure and function of numerous transporters becoming available and their importance in drug transport, efflux, and uptake becoming more thoroughly understood, *in silico* predictions of drug transport mechanisms, drug resistance and first-pass metabolism will be more reliable. Combine this with the computational methods being developed to predict the drug-likeness of compounds and it is clear that drug discovery is already on the road towards electronic R&D [126].

Another approach of classifying drug-like from nondrug-like entities is to use neural network methods. For example, the Bayesian neural network strategy could correctly predict over 90% of the Comprehensive Medicinal Chemistry (CMC) database and about 10% of the molecules in the Available Chemical Directory (ACD) as drug-like [127]. In an independent study using a scoring scheme neural network, a successful classification of 83% of the ACD as nondrugs and 77% of WDI (World Drug Index) as drug was achieved [128]. These studies provided fast automatic schemes to recognize molecules with drug-like properties (for reviews see [129-132]).

Other equally important techniques were developed with similar degrees of success in predicting drug-likeness features of a database [25,38,133-136]. A simple pharmacophore point filter has recently been developed that discriminates between drug-like and nondrug-like entities within a reasonable degree of accuracy. The application of this filter resulted in 66-69% of subsets of the MACCS-II Drug Data Report (MDDR), 61-68% of the CMC and 36% of the ACD as drug-like [137].

PRECLINICAL PHARMACOLOGY AND TOXICO-LOGY

Prior to clinical trials in human, each new chemical entity has to be tested in animals and in many cases, several species. Data concerning toxicity, PK and metabolism is necessary to determine the feasibility and safety of the drug in human. In some cases testing may include xenograft models and a complete toxicology profile should be clearly established at this stage. A careful study of ADME/T characteristics at this phase of design is extremely important since the majority of drug candidates fail clinical trials due to ADME/T deficiencies (see below). Clearly, the benefits of enhancing the ADME/T properties of

molecules through computational design in the discovery phase and actual validation of these properties in several species of animals in the preclinical phase are enormous.

PREDICTION OF ORAL BIOAVAILABILITY

Bioavailability of a compound depends on stability, absorption and transit through the GI tract, and the first pass effect of gut wall and liver metabolism [138]. The ability to predict the oral bioavailability of compounds from their physicochemical properties and structures using computational approaches has recently gained considerable attention. Computational methods are currently available to estimate solubility, metabolism, toxicity, pKa, blood-brain barrier permeability and other ADME and physicochemical parameters. Such information is saving time and money in drug discovery projects at all levels. Therefore, ADME /T information during the early stages of the drug design process will help to determine the ultimate fate of a valuable lead [139]. The need for high-throughput approaches in ADME prediction is driven by the impact of combinatorial chemistry and high throughput screening to the drug discovery process. It is highly suspect that the linking of *in silico* and *in vitro* methods will ultimately replace *in vivo* studies entirely, but recent applications of *in silico* approaches attest to their success [140]. For example, the recent QSAR model for predicting human oral bioavailability of 232 structurally diverse drugs was able to correctly classify 71% of the drugs and 97% were correct to within one class [138]. How computational approaches for ADME parameters have evolved and how they are likely to progress was recently reviewed [4].

The recent computational approaches are taking into account numerous factors to predict bioavailability. For example, it has been suggested that lipophilicity, molecular size, molecular shape, polar surface area, hydrogen bonding capacity, and similar parameters correlate to absorption or permeability [141]. Furthermore, prediction of specific binding to protein active sites and interaction with solvent systems are important features in ADME [142]. The role of partition coefficient, molecular weight, carrier-mediated transport and conformational flexibility for designing orally bioavailable drugs [143] and their physicochemical and delivery considerations was previously reviewed [144]. Additional multivariate data analyses are being employed to derive models that correlate passive intestinal permeability to physicochemical descriptors. A numerical molecular representation

called the molecular hashkey was developed to predict log P and intestinal absorption of a set of drugs [142].

Computational models for ADME prediction rely heavily on aqueous solubility, metabolic stability to microsomal incubation, and membrane permeability as measured in Caco-2 (human colon adenocarcinoma) cell culture systems (see below). There is a greater availability of *in vitro* and *in situ* approaches to screen compounds for intestinal permeability (as a surrogate for absorption) and metabolic stability (as a surrogate for clearance). There are now a variety of methods for predicting biopharmaceutical properties among which the intestinal permeability parameter is particularly important. Computational approaches in predicting the passive transcellular route have been straight forward, but the importance of active drug transport and efflux are also being appreciated [145]. Computational models based on partitioned molecular surface areas, that predict intestinal drug permeability with an accuracy comparable to that of quantum mechanical calculations has recently been presented [146].

More recent modifications of the *in vitro* and *in situ* approaches to assess the potential of absorption and metabolism have enabled a higher throughput and an ability to correlate better with *in vivo* PKs of compounds [147]. The best-described in vitro models of absorption are the Caco-2 and Madin-Darby canine kidney (MDCK) cell lines. The potential mathematical relationships between physicochemical properties and Caco-2 flux data have been investigated, alongside data from human studies. In all these studies it appears that the compound's molecular weight and H-bonding potential are the most crucial factors for absorption in contrast to earlier proposals that suggested solubility and permeability are the key variables. Enzymes of the cytochrome P450 3A (CYP3A) family constitute more than 70% of small intestinal cytochrome P450, and CYP3A is estimated to metabolize between 50-70% of currently used drugs. The major congener of CYP3A is CYP3A4, which has similar substrate specificities as P-glycoprotein (P-gp). This observation implies that CYP3A and P-gp may have a significant impact on the bioavailability of drugs.

Because of the recent progress in computational ADME, it is advised that combinatorial libraries should not only be designed based on diversity [136,148] and drug-like characteristics [25,30] but also on oral bioavailability [149,150]. It is becoming clear that the

drugs of future will be designed, optimized, and many of their features predicted prior to their actual synthesis and development.

METABOLISM

Metabolism is one of the most important determinants of the PK profile of a drug, thus metabolic analysis in drug discovery [151] and high throughput ADME prediction [152] will play an important role in early drug development [140]. Poor ADME /T results accounts for failure of nearly 60% of new clinical entities during development [153]. As a result, an increased effort has been applied to develop predictive computational methods to aid the optimization process during drug discovery and development [154].

Will computational ADME succeed in helping discover drugs faster and cheaper? There is a considerable interest in extrapolating *in vitro* data to the *in vivo* state for early ADME parameters such as absorption, clearance, drug-drug interactions and metabolic stability [155,156]. In addition, predictive computational algorithms, which are now being generated and validated in parallel with *in vitro* and *in vivo* methods have been reasonably successful, as noted in this review. As we increase the number of ADME parameters determined early, the overall successful prediction rate will increase [157]. For example, the enormous amount of information collected with the CYPs and transporter enzymes will undoubtedly allow ADME data mining and ADME informatics to prosper.

Successful application of any computational approach requires availability of a large and reliable database to build rational models. A recent approach to generate reliable and fast metabolism data is to use cassette dosing. Cassette dosing is a procedure to rapidly assess PKs of large numbers of compounds. In this procedure, multiple compounds are administered simultaneously to a single animal. Blood samples are collected and the plasma obtained is analyzed by LC/MS. Consequently, the PKs of multiple compounds can be assessed rapidly with a small number of experimental animals and with reduced assay times [158].

Finally, a new approach aimed at the design of safer drugs with increased therapeutic indices by integrating metabolism considerations into the drug design process has been developed. These so called soft drugs are new therapeutic agents that undergo predictable metabolism to inactive metabolites after exerting their therapeutic effect [159].

FUTURE PERSPECTIVES

The enormous progress in the development of new methods in the field of molecular biology and computer science is currently unprecedented. The drug discovery process is no longer limited to the organic chemist who tinkers with a known structure to fine tune an activity: Drug design and discovery is a multi-disciplinary field where the scientist may soon be able to construct a virtual drug with all the desired chemical, physical and biological properties to survive the rigors of clinical testing-all before doing a single chemical reaction. Drug design and discovery in the postgenomic era is shattering old paradigms and routinely reconstructing the drug discovery protocols by including the eons of information encoded in our genome. These data may be used to rationally construct a drug "blueprint" for each individual for tailored therapy based on our genetic makeup. One could have never imagined this only a few short years ago, but the future will be in proteomics and emerging fields like chemogenomics and metabonomics. Although DNA chips are becoming increasingly accessible and yield reproducible results, there is still much work ahead to construct protein chips and have the ability to perform high throughput structural genomics to unravel the conformations of all relevant proteins in specific disease processes. While various researches are working to generate protein microarray, other alternative strategies involving HPLC, 2-D gel electrophoresis and mass spectrometry are providing attractive alternative methods for protein analysis. We can only hope that this new era of therapeutics research will reduce the time and cost of discovering new drugs as well as help us to design better and more efficient clinical trials.

REFERENCES

[1] Ohlstein, E. H.; Ruffolo, R. R., Jr.; Elliott, J. D. Drug discovery in the next millennium. *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 177-191.

[2] Cunnigham, M. J. Genomics and proteomics The new millennium of drug discovery and development. *J. Pharmacol. Toxicol. Meth.* **2000**, *44*, 291-300.

[3] Debouck, C.; Metcalf, B. The impact of genomics on drug discovery. *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 193-207.

[4] Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A. *et al.* Progress in predicting human ADME parameters in silico. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 251-272.

[5] Sneader, W. Drug Discovery: The Evolution of Modern Medicines; Wiley: Chichester, **1985**.

[6] Verpoorte, R. Exploration of nature's chemodiversity: the role of secondary metabolites as leads in drug development. *Drug Discovery Today* **1998**, *3*, 232-238.

[7] Burger, A. *Medicinal Chemistry*; Wiley: New York, **1970**, 4-19.

[8] Silverman, R. B. *The organic chemistry of drug design and drug action*; Academic Press: San Diego, **1992**.

[9] Sternbach, L. H. The benzodiazepine story. *J. Med. Chem.* **1979***, 22*, 1-7.

[10] Manly, C. J.; Louise-May, S.; Hammer, J. D. The impact of informatics and computational chemistry on synthesis and screening. *Drug Discovery Today* **2001***, 6*, 1101-1110.

[11] Caron, P. R.; Mullican, M. D.; Mashal, R. D.; Wilson, K. P.; Su, M. S. *et al.* Chemogenomic approaches to drug discovery. *Curr. Opin. Chem. Biol.* **2001***, 5*, 464-470.

[12] Jones, D. A.; Fitzpatrick, F. A. Genomics and the discovery of new drug targets. *Curr. Opin. Chem. Biol.* **1999***, 3*, 71-76.

[13] Abuin, A.; Holt, K. H.; Platt, K. A.; Sands, A. T.; Zambrowicz, B. P. Full-speed mammalian genetics: in vivo target validation in the drug discovery process. *Trends Biotechnol.* **2002***, 20*, 36-42.

[14] Dean, N. M. Functional genomics and target validation approaches using antisense oligonucleotide technology. *Curr. Opin. Biotechnol.* **2001***, 12*, 622-625.

[15] Taylor, M. F. Emerging antisense technologies for gene functionalization and drug discovery. *Drug Discovery Today* **2001***, 6*, S97-S101.

[16] Christoffersen, R. E.; Marr, J. J. Ribozymes as human therapeutic agents. *J. Med. Chem.* **1995***, 38*, 2023-2037.

[17] Zhao, J. J.; Lemke, G. Rules for ribozymes. *Mol. Cell Neurosci.* **1998***, 11*, 92-97.

[18] Hartmann, G.; Bidlingmaier, M.; Eigler, A.; Hacker, U.; Endres, S. Cytokines and therapeutic oligonucleotides. *Cytokines Cell Mol. Ther.* **1997***, 3*, 247-256.

[19] Rondon, I. J.; Marasco, W. A. Intracellular antibodies (intrabodies) for gene therapy of infectious diseases. *Annu. Rev. Microbiol.* **1997***, 51*, 257-283.

[20] Grim, J.; Deshane, J.; Siegal, G. P.; Alvarez, R. D.; DiFiore, P. *et al.* The level of erbB2 expression predicts sensitivity to the cytotoxic effects of an intracellular anti-erbB2 sFv. *J. Mol. Med.* **1998***, 76*, 451-458.

[21] Tao, J.; Wendler, P.; Connelly, G.; Lim, A.; Zhang, J. *et al.* Drug target validation: lethal infection blocked by inducible peptide. *Proc. Natl. Acad. Sci. USA* **2000***, 97*, 783-786.

[22] Schreiber, S. L. Chemical genetics resulting from a passion for synthetic organic chemistry. *Bioorg. Med. Chem.* **1998***, 6*, 1127-1152.

[23] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Deliv. Rev.* **2001***, 46*, 3-26.

[24] Barry, C. E., 3rd; Slayden, R. A.; Sampson, A. E.; Lee, R. E. Use of genomics and combinatorial chemistry in the development of new antimycobacterial drugs. *Biochem. Pharmacol.* **2000***, 59*, 221-231.

[25] Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1999***, 1*, 55-68.

[26] Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual screening- an overview. *Drug Discovery Today* **1998***, 3*, 160-178.

[27] Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **2000***, 44*, 235-249.

[28] Bemis, G. W.; Murcko, M. A. The properties of known drugs. 1. Molecular frameworks. *J. Med. Chem.* **1996***, 39*, 2887-2893.

[29] Bemis, G. W.; Murcko, M. A. Properties of known drugs. 2. Side chains. *J. Med. Chem.* **1999***, 42*, 5095-5099.

[30] Wang, J.; Ramnarayan, K. Toward designing drug-like libraries: a novel computational approach for prediction of drug feasibility of compounds. *J. Comb. Chem.* **1999***, 1*, 524-533.

[31] Koehler, R. T.; Dixon, S. L.; Villar, H. O. LASSOO: a generalized directed diversity approach to the design and enrichment of chemical libraries. *J. Med. Chem.* **1999***, 42*, 4695-4704.

[32] Luscombe, N. M.; Greenbaum, D.; Gerstein, M. What is bioinformatics? A proposed definition and overview of the field. *Methods Inf. Med.* **2001***, 40*, 346-358.

[33] Hann, M.; Green, R. Chemoinformatics--a new name for an old problem? *Curr. Opin. Chem. Biol.* **1999***, 3*, 379-383.

[34] Heuer, K. I. The development cycle of a pharmaceutical discovery chemi-informatics system. *Med. Res. Rev.* **1999***, 19*, 209-221.

[35] Xue, L.; Bajorath, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Comb. Chem. High Throughput Screen.* **2000***, 3*, 363-372.

[36] Xue, L.; Godden, J. W.; Bajorath, J. Evaluation of descriptors and mini-fingerprints for the identification of molecules with similar activity. *J. Chem. Inf. Comput. Sci.* **2000***, 40*, 1227-1234.

[37] Hull, R. D.; Singh, S. B.; Nachbar, R. B.; Sheridan, R. P.; Kearsley, S. K. *et al.* Latent semantic structure indexing (LaSSI) for defining chemical similarity. *J. Med. Chem.* **2001***, 44*, 1177-1184.

[38] Hull, R. D.; Fluder, E. M.; Singh, S. B.; Nachbar, R. B.; Kearsley, S. K. *et al.* Chemical similarity searches using latent semantic structural indexing (LaSSI) and comparison to TOPOSIM. *J. Med. Chem.* **2001***, 44*, 1185-1191.

[39] Kurogi, Y.; Guner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr. Med. Chem.* **2001***, 8*, 1035-1055.

[40] Meyer, E. F.; Swanson, S. M.; Williams, J. A. Molecular modelling and drug design. *Pharmacol. Ther.* **2000**, *85*, 113-121.

[41] Hopfinger, A. J.; Duca, J. S. Extraction of pharmacophore information from high-throughput screens. *Curr. Opin. Biotechnol.* **2000**, *11*, 97-103.

[42] McGregor, M. J.; Muskal, S. M. Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 569-574.

[43] Tropsha, A.; Zheng, W. Identification of the descriptor pharmacophores using variable selection QSAR: applications to database mining. *Curr. Pharm. Des.* **2001**, *7*, 599-612.

[44] Douguet, D.; Thoreau, E.; Grassy, G. A genetic algorithm for the automated generation of small organic molecules: drug design using an evolutionary algorithm. *J. Comput. Aided Mol. Des.* **2000**, *14*, 449-466.

[45] de Julian-Ortiz, J. V. Virtual darwinian drug design: QSAR inverse problem, virtual combinatorial chemistry, and computational screening. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 295-310.

[46] Olender, R.; Rosenfeld, R. A fast algorithm for searching for molecules containing a pharmacophore in very large virtual combinatorial libraries. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 731-738.

[47] Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C. *et al.* New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251-3264.

[48] Mason, J. S.; Good, A. C.; Martin, E. J. 3-D pharmacophores in drug discovery. *Curr. Pharm. Des.* **2001**, *7*, 567-597.

[49] Chen, Y. Z.; Zhi, D. G. Ligand-protein inverse docking and its potential use in the computer search of protein targets of a small molecule. *Proteins* **2001**, *43*, 217-226.

[50] Roche, O.; Kiyama, R.; Brooks, C. L., 3rd Ligand-protein database: linking protein-ligand complex structures to binding data. *J. Med. Chem.* **2001**, *44*, 3592-3598.

[51] Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100-5109.

[52] Pearlman, D. A. Free energy grids: a practical qualitative application of free energy perturbation to ligand design using the OWFEG method. *J. Med. Chem.* **1999**, *42*, 4313-4324.

[53] Pearlman, D. A.; Charifson, P. S. Improved scoring of ligand-protein interactions using OWFEG free energy grids. *J. Med. Chem.* **2001**, *44*, 502-511.

[54] Terp, G. E.; Johansen, B. N.; Christensen, I. T.; Jorgensen, F. S. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities. *J. Med. Chem.* **2001**, *44*, 2333-2343.

[55] Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43*, 4759-4767.

[56] Fradera, X.; Knegtel, R. M.; Mestres, J. Similarity-driven flexible ligand docking. *Proteins* **2000**, *40*, 623-636.

[57] Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213-218.

[58] Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* **1999**, *37*, 88-105.

[59] Lamb, M. L.; Burdick, K. W.; Toba, S.; Young, M. M.; Skillman, A. G. *et al.* Design, docking, and evaluation of multiple libraries against multiple targets. *Proteins* **2001**, *42*, 296-318.

[60] Wang, J.; Kollman, P. A.; Kuntz, I. D. Flexible ligand docking: a multistep strategy approach. *Proteins* **1999**, *36*, 1-19.

[61] Koehler, R. T.; Villar, H. O. Statistical relationships among docking scores for different protein binding sites. *J. Comput. Aided Mol. Des.* **2000**, *14*, 23-37.

[62] Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375-382.

[63] Godden, J. W.; Xue, L.; Bajorath, J. Combinatorial preferences affect molecular similarity/diversity calculations using binary fingerprints and Tanimoto coefficients. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 163-166.

[64] Allen, B. C.; Grant, G. H.; Richards, W. G. Similarity calculations using two-dimensional molecular representations. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 330-337.

[65] Stanton, D. T.; Morris, T. W.; Roychoudhury, S.; Parker, C. N. Application of nearest-neighbor and cluster analyses in pharmaceutical lead discovery. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 21-27.

[66] Langer, T.; Hoffmann, R. D. Virtual screening: an effective tool for lead structure discovery? *Curr. Pharm. Des.* **2001**, *7*, 509-527.

[67] Diercks, T.; Coles, M.; Kessler, H. Applications of NMR in drug discovery. *Curr. Opin. Chem. Biol.* **2001**, *5*, 285-291.

[68] Hicks, R. P. Recent advances in NMR: Expanding its role in rational drug design. *Curr. Med. Chem.* **2001**, *8*, 627-650.

[69] Gotfredsen, C. H.; Grotli, M.; Willert, M.; Meldal, M.; Duus, J. O. Single-bead structure elucidation. Requirements for analysis of combinatorial solid-phase libraries by Nanoprobe MAS-NMR spectroscopy. *J. Chem. Soc. Perk. Tran.* **2000**, *7*, 1167-1171.

[70] Keifer, P. A. High-resolution NMR techniques for solid-phase synthesis and combinatorial chemistry. *Drug Discovery Today* **1997**, *2*, 468-478.

[71] Johnson, C. S. Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications. *Progr. NMR Spectrosc.* **1999**, *34*, 203-256.

[72] Mayer, M.; Meyer, B. Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew Chem. Int. Ed. Engl.* **1999**, *38*, 1784-1788.

[73] Chen, A.; Shapiro, M. J. NOE pumping: a novel NMR technique for identification of compounds with binding affinity to macromolecules. *J. Am. Chem. Soc.* **1998**, *120*, 10258-10259.

[74] Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531-1534.

[75] Hajduk, P. J.; Sheppard, G.; Nettesheim, D.; Olejniczak, E. T.; Shuker, S. B. *et al.* Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* **1997**, *119*, 5818-5827.

[76] Medek, A.; Hajduk, P. J.; Mack, J.; Fesik, S. W. The use of differential chemical shifts for determining the binding site location and orientation of protein-bound ligands. *J. Am. Chem. Soc.* **2000**, *122*, 1241-1242.

[77] Hajduk, P. J.; Gerfin, T.; Boehlen, J. M.; Haberli, M.; Marek, D. *et al.* High-throughput nuclear magnetic resonance-based screening. *J. Med. Chem.* **1999**, *42*, 2315-2317.

[78] Fejzo, J.; Lepre, C. A.; Peng, J. W.; Bemis, G. W.; Ajay *et al.* The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* **1999**, *6*, 755-769.

[79] Dalvit, C.; Fogliatto, G.; Stewart, A.; Veronesi, M.; Stockman, B. WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J. Biomolecular NMR* **2001**, *21*, 349-359.

[80] Crews, C. M.; Splittgerber, U. Chemical genetics: exploring and controlling cellular processes with chemical probes. *Trends Biochem. Sci.* **1999**, *24*, 317-320.

[81] Stockwell, B. R. Chemical genetics: ligand-based discovery of gene function. *Nat. Rev. Genet.* **2000**, *1*, 116-125.

[82] Stockwell, B. R. Frontiers in chemical genetics. *Trends Biotechnol.* **2000**, *18*, 449-455.

[83] Alaimo, P. J.; Shogren-Knaak, M. A.; Shokat, K. M. Chemical genetic approaches for the elucidation of signaling pathways. *Curr. Opin. Chem. Biol.* **2001**, *5*, 360-367.

[84] Weiss, E. L.; Bishop, A. C.; Shokat, K. M.; Drubin, D. G. Chemical genetic analysis of the budding-yeast p21-activated kinase Cla4p. *Nat. Cell Biol.* **2000**, *2*, 677-685.

[85] Shogren-Knaak, M. A.; Alaimo, P. J.; Shokat, K. M. Recent advances in chemical approaches to the study of biological systems. *Annu. Rev. Cell Dev. Biol.* **2001**, *17*, 405-433.

[86] Adam, D. Database of molecular probes set to boost chemical genetics. *Nature* **2001**, *411*, 873.

[87] Mayer, T. U.; Kapoor, T. M.; Haggarty, S. J.; King, R. W.; Schreiber, S. L. *et al.* Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **1999**, *286*, 971-974.

[88] Haggarty, S. J.; Mayer, T. U.; Miyamoto, D. T.; Fathi, R.; King, R. W. *et al.* Dissecting cellular processes using small molecules: identification of colchicine-like, taxol-like and other small molecules that perturb mitosis. *Chem. Biol.* **2000***, 7*, 275-286.

[89] Soni, R.; O'Reilly, T.; Furet, P.; Muller, L.; Stephan, C. *et al.* Selective in vivo and in vitro effects of a small molecule inhibitor of cyclin-dependent kinase 4. *J. Natl. Cancer Inst.* **2001***, 93*, 436-446.

[90] Soni, R.; Chaudhuri, B. Cell cycle arrest mediated by a pyridopyrimidine is not abrogated by over-expression of Bcl-2 and cyclin D1. *Int. J. Oncol.* **2001***, 18*, 1035-1040.

[91] Bedalov, A.; Gatbonton, T.; Irvine, W. P.; Gottschling, D. E.; Simon, J. A. Identification of a small molecule inhibitor of Sir2p. *Proc. Natl. Acad. Sci. USA* **2001***, 98*, 15113-15118.

[92] Bishop, A. C.; Buzko, O.; Shokat, K. M. Magic bullets for protein kinases. *Trends Cell Biol.* **2001***, 11*, 167-172.

[93] Blackwell, H. E.; Perez, L.; Stavenger, R. A.; Tallarico, J. A.; Cope Eatough, E. *et al.* A one-bead, one-stock solution approach to chemical genetics: part 1. *Chem. Biol.* **2001***, 8*, 1167-1182.

[94] Clemons, P. A.; Koehler, A. N.; Wagner, B. K.; Sprigings, T. G.; Spring, D. R. *et al.* A one-bead, one-stock solution approach to chemical genetics: part 2. *Chem. Biol.* **2001***, 8*, 1183-1195.

[95] MacBeath, G.; Koehler, A. N.; Schreiber, S. L. Printing small molecules as microarrays and detecting protein-ligand interactions en masse. *J. Am. Chem. Soc.* **1999***, 121*, 7967-7968.

[96] MacBeath, G.; Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **2000***, 289*, 1760-1763.

[97] Bishop, A.; Buzko, O.; Heyeck-Dumas, S.; Jung, I.; Kraybill, B. *et al.* Unnatural ligands for engineered proteins: new tools for chemical genetics. *Annu. Rev. Biophys. Biomol. Struct.* **2000***, 29*, 577-606.

[98] Joseph-McCarthy, D. Computational approaches to structure-based ligand design. *Pharmacol. Ther.* **1999***, 84*, 179-191.

[99] Gane, P. J.; Dean, P. M. Recent advances in structure-based rational drug design. *Curr. Opin. Struct. Biol.* **2000***, 10*, 401-404.

[100] Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is There a Difference between Leads and Drugs? A Historical Perspective. *J. Chem. Inf. Comput. Sci.* **2001***, 41*, 1308-1315.

[101] Ooms, F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr. Med. Chem.* **2000***, 7*, 141-158.

[102] Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med.* **2000***, 78*, 269-281.

[103] Hornak, V.; Dvorsky, R.; Sturdik, E. Receptor-ligand interaction and molecular modelling. *Gen. Physiol. Biophys.* **1999***, 18*, 231-248.

[104] Farber, G. K. New approaches to rational drug design. *Pharmacol. Ther.* **1999***, 84*, 327-332.

[105] Fernandes, P. B. The ATCG of drug discovery. *Curr. Opin. Mol. Ther.* **2000***, 2*, 624-632.

[106] Liu, H.; Duan, Z.; Luo, Q.; Shi, Y. Structure-based ligand design by dynamically assembling molecular building blocks at binding site. *Proteins* **1999**, *36*, 462-470.

[107] Jackson, R. C. Contributions of protein structure-based drug design to cancer chemotherapy. *Semin. Oncol.* **1997**, *24*, 164-172.

[108] Gibbs, J. B. Mechanism-based target identification and drug discovery in cancer research. *Science* **2000**, *287*, 1969-1973.

[109] Frye, S. V. Structure-activity relationship homology (SARAH): a conceptual framework for drug discovery in the genomic era. *Chem. Biol.* **1999**, *6*, R3-7.

[110] Schafferhans, A.; Klebe, G. Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* **2001**, *307*, 407-427.

[111] Gallet, X.; Charloteaux, B.; Thomas, A.; Brasseur, R. A fast method to predict protein interaction sites from sequences. *J. Mol. Biol.* **2000**, *302*, 917-926.

[112] Rishton, G. M. Reactive compounds and in vitro false positives in HTS. *Drug Discovery Today* **1997**, *2*, 382-385.

[113] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery. *Adv. Drug. Deliv. Rev.* **1997**, *23*, 3-25.

[114] Ajay; Bemis, G. W.; Murcko, M. A. Designing libraries with CNS activity. *J. Med. Chem.* **1999**, *42*, 4942-4951.

[115] Anzali, S.; Barnickel, G.; Cezanne, B.; Krug, M.; Filimonov, D. *et al.* Discriminating between drugs and nondrugs by prediction of activity spectra for substances (PASS). *J. Med. Chem.* **2001**, *44*, 2432-2437.

[116] Apostolakis, J.; Caflisch, A. Computational ligand design. *Comb. Chem. High Throughput Screen.* **1999**, *2*, 91-104.

[117] Clark, D. E. Prediction of intestinal absorption and blood-brain barrier penetration by computational methods. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 477-496.

[118] Hodgson, J. ADMET--turning chemicals into drugs. *Nat. Biotechnol.* **2001**, *19*, 722-726.

[119] Streng, W. H. Physical chemical characterization of drug substances. *Drug Discovery Today* **1997**, *2*, 415-426.

[120] van De Waterbeemd, H.; Smith, D. A.; Beaumont, K.; Walker, D. K. Property-based design: optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* **2001**, *44*, 1313-1333.

[121] van de Waterbeemd, H.; Smith, D. A.; Jones, B. C. Lipophilicity in PK design: methyl, ethyl, futile. *J. Comput. Aided Mol. Des.* **2001**, *15*, 273-286.

[122] Huuskonen, J. Estimation of aqueous solubility in drug design. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 311-316.

[123] Huuskonen, J. J.; Villa, A. E.; Tetko, I. V. Prediction of partition coefficient based on atom-type electrotopological state indices. *J. Pharm. Sci.* **1999**, *88*, 229-233.

[124] Kibbey, C. E.; Poole, S. K.; Robinson, B.; Jackson, J. D.; Durham, D. An integrated process for measuring the physicochemical properties of drug candidates in a preclinical discovery environment. *J. Pharm. Sci.* **2001**, *90*, 1164-1175.

[125] De Groot, A. S.; Rothman, F. G. In silico predictions; in vivo veritas. *Nat. Biotechnol.* **1999**, *17*, 533-534.

[126] Terstappen, G. C.; Reggiani, A. In silico research in drug discovery. *Trends Pharmacol. Sci.* **2001**, *22*, 23-26.

[127] Ajay, A.; Walters, W. P.; Murcko, M. A. Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? *J. Med. Chem.* **1998**, *41*, 3314-3324.

[128] Sadowski, J.; Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41*, 3325-3329.

[129] Walters, W. P.; Ajay; Murcko, M. A. Recognizing molecules with drug-like properties. *Curr. Opin. Chem. Biol.* **1999**, *3*, 384-387.

[130] Mitchell, T.; Showell, G. A. Design strategies for building drug-like chemical libraries. *Curr. Opin. Drug Discov. Devel.* **2001**, *4*, 314-318.

[131] Schneider, G. Neural networks are useful tools for drug design. *Neural Netw.* **2000**, *13*, 15-16.

[132] Kovesdi, I.; Dominguez-Rodriguez, M. F.; Orfi, L.; Naray-Szabo, G.; Varro, A. *et al.* Application of neural networks in structure-activity relationships. *Med. Res. Rev.* **1999**, *19*, 249-269.

[133] Gillet, V. J.; Willett, P.; Bradshaw, J. Identification of biological activity profiles using substructural analysis and genetic algorithms. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 165-179.

[134] Wagener, M.; vanGeerestein, V. J. Potential drugs and nondrugs: prediction and identification of important structural features. *J. Chem. Inf. Comput. Sci.* **2000**, *40*.

[135] Xu, J.; Stevenson, J. Drug-like index: a new approach to measure drug-like compounds and their diversity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1177-1187.

[136] Brown, R. D.; Hassan, M.; Waldman, M. Combinatorial library design for diversity, cost efficiency, and drug- like character. *J. Mol. Graph Model* **2000**, *18*, 427-437, 537.

[137] Muegge, I.; Heald, S. L.; Brittelli, D. Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* **2001**, *44*, 1841-1846.

[138] Yoshida, F.; Topliss, J. G. QSAR model for drug human oral bioavailability. *J. Med. Chem.* **2000**, *43*, 2575-2585.

[139] Podlogar, B. L.; Muegge, I.; Brice, L. J. Computational methods to estimate drug development parameters. *Curr. Opin. Drug Discov. Devel.* **2001**, *4*, 102-109.

[140] Smith, D. A.; van de Waterbeemd, H. Pharmacokinetics and metabolism in early drug discovery. *Curr. Opin. Chem. Biol.* **1999**, *3*, 373-378.

[141] Mannhold, R.; van de Waterbeemd, H. Substructure and whole molecule approaches for calculating log P. *J. Comput. Aided Mol. Des.* **2001**, *15*, 337-354.

[142] Ghuloum, A. M.; Sage, C. R.; Jain, A. N. Molecular hashkeys: a novel method for molecular characterization and its application for predicting important pharmaceutical properties of molecules. *J. Med. Chem.* **1999**, *42*, 1739-1748.

[143] Navia, M. A.; Chaturvedi, P. R. Design principales for orally bioavailable drugs. *Drug Discovery Today* **1996**, *1*, 179-189.

[144] Chan, O. H.; Stewart, B. H. Physicochemical and drug delivery considerations for oral drug bioavailability. *Drug Discovery Today* **1996**, *1*, 461-473.

[145] Stenberg, P.; Luthman, K.; Artursson, P. Virtual screening of intestinal drug permeability. *J. Control Release* **2000**, *65*, 231-243.

[146] Stenberg, P.; Norinder, U.; Luthman, K.; Artursson, P. Experimental and computational screening models for the prediction of intestinal drug absorption. *J. Med. Chem.* **2001**, *44*, 1927-1937.

[147] Chaturvedi, P. R.; Decker, C. J.; Odinecs, A. Prediction of pharmacokinetic properties using experimental approaches during early drug discovery. *Curr. Opin. Chem. Biol.* **2001**, *5*, 452-463.

[148] Martin, E. J.; Critchlow, R. E. Beyond mere diversity: tailoring combinatorial libraries for drug discovery. *J. Comb. Chem.* **1999**, *1*, 32-45.

[149] Darvas, F.; Dorman, G.; Papp, A. Diversity measures for enhancing ADME admissibility of combinatorial libraries. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 314-322.

[150] Fecik, R. A.; Frank, K. E.; Gentry, E. J.; Menon, S. R.; Mitscher, L. A. *et al.* The search for orally active medications through combinatorial chemistry. *Med. Res. Rev.* **1998**, *18*, 149-185.

[151] Cardenas, M. L.; Cornish-Bowden, A. Metabolic analysis in drug discovery. *Science* **2000**, *288*, 618-619.

[152] Tarbit, M. H.; Berman, J. High throughput approaches for evaluating absorption, distribution, metabolism, and excretion properties of lead compounds. *Curr. Opin. Chem. Biol.* **1998**, *2*, 411-416.

[153] Kennedy, T. Managing the discovery/development interface. *Drug Dis. Discov.* **1997**, *2*, 436-444.

[154] Kumar, G. N.; Surapaneni, S. Role of drug metabolism in drug discovery and development. *Med. Res. Rev.* **2001**, *21*, 397-411.

[155] Thompson, T. N. Optimization of metabolic stability as a goal of modern drug design. *Med. Res. Rev.* **2001**, *21*, 412-449.

[156] Thompson, T. N. Early ADME in support of drug discovery: the role of metabolic stability studies. *Curr. Drug Metab.* **2000**, *1*, 215-241.

[157] Ekins, S.; Ring, B. J.; Grace, J.; McRobie-Belle, D. J.; Wrighton, S. A. Present and future in vitro approaches for drug metabolism. *J. Pharmacol. Toxicol. Methods* **2000***, 44*, 313-324.

[158] White, R. E.; Manitpisitkul, P. Pharmacokinetic theory of cassette dosing in drug discovery screening. *Drug Metab. Dispos.* **2001***, 29*, 957-966.

[159] Bodor, N.; Buchwald, P. Soft drug design: general principles and recent applications. *Med. Res. Rev.* **2000***, 20*, 58-101.