Journal of Cellular Biochemistry Supplement 37:13–21 (2001)

Current Practices in Generation of Small Molecule New Leads

Robert A. Goodnow, Jr.*

New Leads Chemistry Initiative, Roche Research Center, Nutley, New Jersey 07110-1199

Abstract The current drug discovery processes in many pharmaceutical companies require large and growing collections of high quality lead structures for use in high throughput screening assays. Collections of small molecules with diverse structures and ''drug-like'' properties have, in the past, been acquired by several means: by archive of previous internal lead optimization efforts, by purchase from compound vendors, and by union of separate collections following company mergers. More recently, many drug discovery companies have established dedicated efforts to effect synthesis by internal and/or outsourcing efforts of targeted compound libraries for new lead generation. Although high throughput/combinatorial chemistry is an important component in the process of new lead generation, the selection of library designs for synthesis and the subsequent design of library members has evolved to a new level of challenge and importance. The potential benefits of screening multiple small molecule compound library designs against multiple biological targets offers substantial opportunity to discover new lead structures. Subsequent optimization of such compounds is often accelerated because of the structure-activity relationship (SAR) information encoded in these lead generation libraries. Lead optimization is often facilitated due to the ready applicability of high-throughput chemistry (HTC) methods for follow-up synthesis. Some of the strategies, trends, and critical issues central to the success of lead generation processes are discussed below. J. Cell. Biochem. Suppl. 37: 13–21, 2001.

Key words: lead generation; high-throughput chemistry; drug-likeness; lead-likeness; chemical genomics

Although the value of a "good lead" in drug discovery programs cannot be overstated, it is a challenging undertaking to create a diverse and sizable collection of quality molecules. In meeting the challenge of synthesizing hundreds of thousands of novel and high quality molecules, the opportunities and challenges of several strategic directions must be realistically balanced.

MODERN DRUG DISCOVERY PROCESS REQUIRES MORE HIGH QUALITY COMPOUNDS ASSAYED AGAINST MULTIPLE BIOLOGICAL TARGETS

One can break the traditional processes of drug discovery into several distinct yet interdependent steps. These steps include biological target identification, target assay development, high throughput screening, and assay hit identification. Once hits have been identified, a process known as "hit-to-lead" is put into place. This progress refines hit structures before beginning medicinal chemistry-driven optimization for possible identification of a clinical candidate. (Fig. 1). Thereafter, entry-into-human enabling studies are required prior to clinical development and finally market launch. Unfortunately, the majority of molecules moving along this pathway fail for a variety of reasons, including lack of potency and/or efficacy, metabolic or toxicity liability, or invalidation of the biological target-disease indication. Partially due to this failure rate, current estimates indicate that bringing a drug to market often requires at least 10 years and greater than 350–500 million US dollars. In part due to the difficulties of this sequence of hurdles, the pharmaceutical industry currently registers on average only 0.5 new drugs per annum each [Arlington, 2000]. Despite these risks and hurdles, the pharmaceutical industry continues to be a competitive field in which it is thought that only the most innovative will survive to prosper. A recent

Robert A. Goodnow is a Senior Research Leader.

*Correspondence to: Robert A. Goodnow, Jr., New Leads Chemistry Initiative, Roche Research Center, Nutley, New Jersey 07110-1199. E-mail: robert.goodnow@roche.com

Received 13 September 2001; Accepted 18 September 2001

DOI 10.1002/jcb.10061

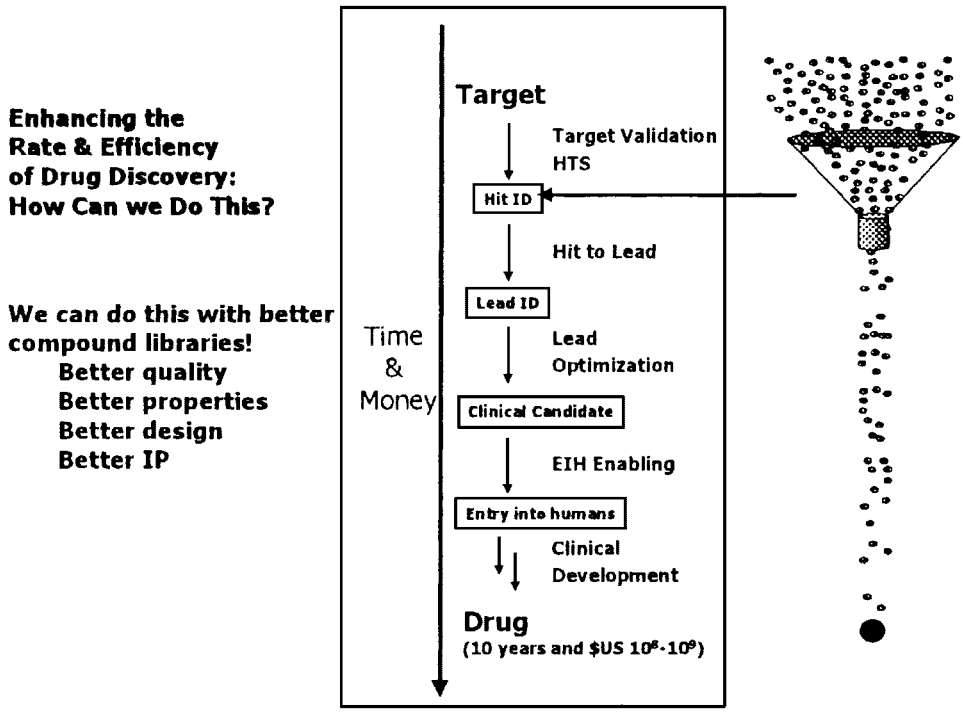

Fig. 1. Enhancing efficacy of the drug discovery process with better compounds.

study by Pricewaterhouse Coopers indicated that the growth rate of the pharmaceutical industry could not be maintained if the overall average of 0.5 new drugs registered per annum is not increased. The smooth function of this process requires among other things, a good compound collection to find high quality compounds to move through the discovery pathway. Conversely, a lack of a quality lead will often result in the premature termination of a drug discovery program.

While the challenges and competition in drug discovery persist, as a result of the human genome sequencing, there is a growing number of new targets of potential therapeutic value that are emerging. For example, it is estimated that in the human genome there are approximately 1,500 G-protein coupled receptors and approximately 860 kinases [Venter, 2001]. For each of these targets, classes no more than approximately 10% have been characterized to any extent useful for therapeutic targeting. Following the identification of many new genes is the growing field of proteomics. Proteomics is the study of gene products given the fact that for a single gene, there may be several protein products due to post-protein synthesis temporal modification processes [Fields, 2001]. With the current odds of success, the prospect of repeating and scaling the drug discovery process for such numbers of new protein targets seems inefficient.

The traditional strategy for lead discovery described above identifies drug leads based on the assumption of validated targets. However, target validation often requires a potent ligand, either endogenous or exogenous. Much time can pass during which efforts to validate a target of therapeutic interest followed by finding a potent molecule. Often, the efforts of target validation have been restricted to the realm of biology while lead generation remained solely an interest of chemists. Recently, many pharmaceutical organizations have begun to join the efforts of target validation and lead generation earlier in the discovery pathway. In this concept, protein targets, both characterized and orphan, are sorted into protein families; at the same time, databases of small molecules that are known to be active against the known targets are created. These small molecule databases form the basis for chemoinformatic attempts to summarize and understand common molecular descriptors. Subsequently, small molecule libraries which incorporate these chemoinformatic features can then be selected and synthesized. The basic assumption is that families of

similar protein targets are likely to interact with molecules that are similar to the families of known, target-associated ligands. This approach termed "chemical genomics" [Jacoby, 2001], requires a good compound collection. A lack of quality compounds will likely prevent the process from moving beyond the target validation phase.

In both approaches to drug discovery, the characteristics of molecules that populate the screening collection are important. The biological activity of a molecule must be complemented by other properties that make the molecule, a good drug substance. It is estimated that a large proportion of molecules in late stages of drug development fail due to reasons of drug–drug interaction action or poor ADME (absorption, distribution, metabolism, and excretion) properties. Not detecting these liabilities early in the drug discovery pathway can be extremely costly and time consuming. With the concept of detecting a molecule's liabilities early on, many drug discovery organizations have been experimenting with and implementing multiple strategies to enhance rates of success. Computational methods to detect such liabilities early are gaining acceptance particulary in the earliest stages of lead optimization studies. [Ekins and Wrighton, 2001]. Such strategies include improved computational methods for predictions of toxicity and physico-chemical properties relevant to ADME. The rationale is simple: with better compounds that are active with targets which are amenable to intervention with small molecules, the chances for successful drug discovery are increased. Thus, staring with a collection of good leads is likely to enhance the overall success of any drug discovery program.

VALUE OF A GOOD LEAD CANNOT BE OVERSTATED

Physical and calculated properties for known drugs have defined concepts for "drug-likeness" [Lipinski et al., 1997; Lipinski, 2001]. There are numerous calculated properties (so-called "descriptors") which are used to characterize molecules. Multiple efforts exist to sort drug-like from non-drug-like molecules by different combinations of these descriptors. None of the efforts has identified the critical descriptor or set of descriptors that makes the prediction of drug-likeness obvious. More recently, multiple lead optimization efforts have highlighted the value of creating small molecule libraries with "lead-like" properties [Teague et al., 1999] (e.g., molecular weights < 350, clogP < 3, solubilities > 10 μg/ml, etc.). Several studies have shown that compounds with lead-like properties are more easily elaborated during the lead optimization process of drug discovery. During lead optimization, in an effort to increase a compound's activity and selectivity, molecular complexities, molecular weights, and compound lipophilicities often increase; excessive increase in the latter two properties may result in less attractive ADME properties. Compounds with poor ADME properties are difficult to transform into useful drug substances. It stands to reason, therefore, that having more lead-like, less complex compounds is likely to facilitate their optimization.

As part of a general lead generation strategy, many companies have set about to use convenient, high throughput synthesis methods for the creation of high-quality lead generation libraries. Although the value of a "good lead" cannot be overstated, it is extremely challenging to create a diverse and sizable collection of "quality" molecules. In meeting the challenge of synthesizing hundreds of thousands of novel and high quality molecules, the opportunities and challenges of several strategic directions must be realistically balanced.

BALANCED STRATEGY TO OBTAIN COMPOUNDS BY DEDICATED HIGH THROUGHPUT SYNTHESIS TO COMPLEMENT OTHER SOURCES OF LEAD GENERATION COMPOUNDS

As high-throughput synthesis (HTS) technology has continued to evolve to the point where the screening of hundreds of thousands of compounds has become routine, greater and greater numbers of compounds are required. Simultaneous to the scale of screening, there has been a decrease in the quantities necessary for testing. These two changes have had an effect on the way in which pharmaceutical companies build their compound libraries.

Traditionally, the small molecule source for large pharmaceutical companies is composed of molecules primarily from internal lead optimization efforts. Although these molecules provide some of the best examples of drug-likeness and are often available in tens of milligram

quantities, their structures are usually focused around a particular lead structure and thus represent only limited advantages for discovery of diverse and novel lead molecules. Chemistry teams working with a lead molecule from a previous target's optimization program may find significant bias to the previous target's structure-activity relationship (SAR) requirements. One measure of the bias of a molecule is the calculated structural complexity score [Hann et al., 2001]. Methods exist for analyzing and quantifying an arbitrary measure of a molecule's complexity. The SAR for more complex leads can be difficult to disentangle to simpler structures.

During the past decade, many compound vendor companies have come into existence to meet the growing demand of the pharmaceutical industry. It is possible to acquire hundreds of thousands of non-exclusive compounds from commercial sources. These molecules may be obtained as purified solid samples. Such compounds collectively offer ready access to a diverse set of structures at relatively low cost. However, a large fraction of these molecules were synthesized before the concepts of good drug-like properties were well understood. As a result, many find that after filtration of these molecules for properties appropriate to a lead, many fewer molecules remain attractive for purchase. Further, since these molecules are often sold on a non-exclusive basis, an understanding of the likely limitations to intellectual property rights must be considered when using these molecules as new leads for optimization efforts.

Access to large numbers of exclusive, novel, and high-quality compounds designed according to the themes of drug-likeness and lead-likeness is a goal of many major pharmaceutical and biotechnology companies. One solution has been to establish dedicated groups of lead generation chemists. The need to dedicate these groups to the distinct task of lead generation is important as the tactical goals lead optimization teams are quite distinct. Whereas lead optimization chemists usually have quite specific biological and chemical targets, the lead generation chemistry must bring a balance of innovation, creativity, and risk to generate novel and useful structures.

The limited internal resources available to synthesize useful numbers of lead generation libraries often makes outsourcing part or all of the chemical synthesis for new lead generation an attractive strategy alternative. The internal versus external balance of efforts in this process is largely a function of internal goals, infrastructure, cost, research, and business goals.

DESIGN CONCEPTS FOR LEAD-GENERATION LIBRARIES: LIBRARY TARGETING AND SIZE

Early efforts in combinatorial chemistry promoted the concept that large numbers of compounds obviated the need for compound-by-compound design considerations. Although serendipity is an important advantage in the strategy of applying the high throughput chemistry method for lead generation, increasingly, libraries are designed with a particular targeting focus. The targeting concept may be quite focused around a specific pharmacophore hypothesis (Fig. 2). Alternatively, libraries have been targeted more generally around so-called "privileged structures" [Patchett and Nargund, 2000] or motifs frequently associated with protein families such as G-protein coupled receptors, kinases, and proteases. Several nice examples exist for such a targeted concept of lead generation [Maloney et al., 2000].

The computational association of particular library design core structures to targets can take at least two directions. First, where there is bio-structural information, it is possible to test library design ideas by in silico docking experiments. The structures of a virtual library, compounds yet to exist other than in a computer database, can be docked automatically into a particular target structure and evaluated for fit with that target. It is assumed that a good fit in a computational simulation will correlate the good fit in reality. Second, where the structure of the biological target is not known, then one must rely more heavily on similarity of ligands that are known to bind to the target in question. In the same manner that some have tried to sort drugs from non-drugs by specific computational metrics, one can attempt to test the fitness of a library design hypothesis to a set of ligands known to be active.

While it is important to balance the efforts of new lead generation with the potential advantages of starting library designs biasing privileged structures, it is equally important to consider the generation of de novo chemical structures. The concept of de novo ligand design is particularly important for "template hop-

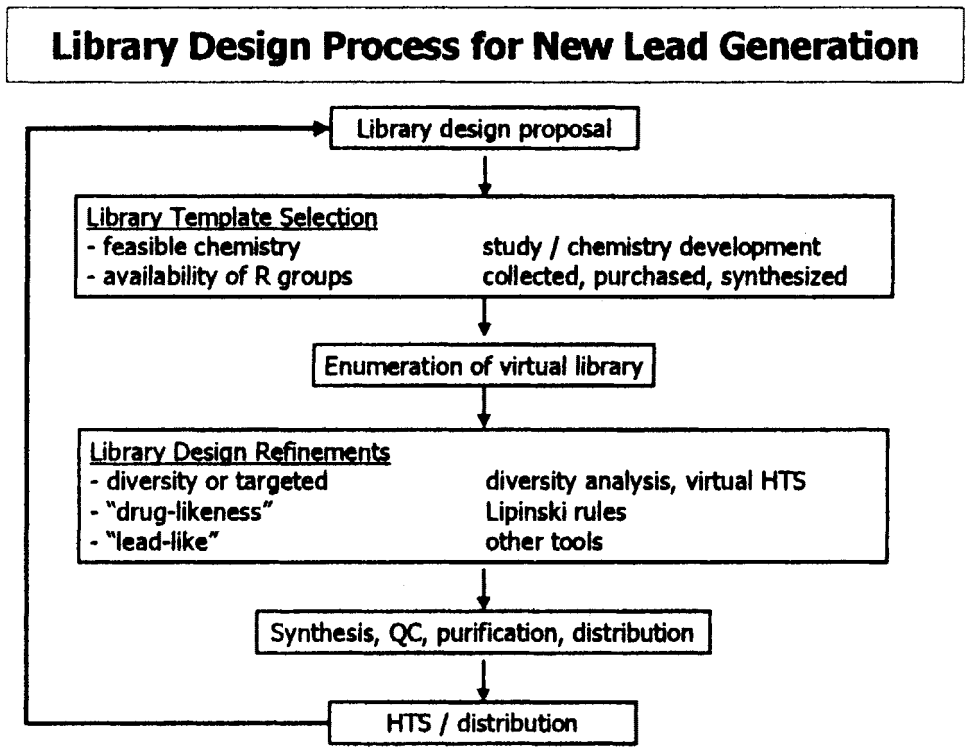

Fig. 2. Targeted library design process for the generation of new leads.

ping," an attempt to design a different core structure with similar chemical and biological properties. Template hopping is particular advantageous when looking to improve on a current chemical series with an entirely new system. Several reports exist of which describe the fragmentation of known drugs and incorporation of such fragments confer a greater likelihood of drug action in new structures [Bemis and Murcko, 1996]. Other efforts have reported on the use of chemistry under extreme conditions to generate structures from unusual chemical pathways [Johnson et al., 1998]. Despite their complexity, natural products have been one of the best sources for diverse structures. Natural products are a proven source of many active starting points for new lead generation. However, the practical difficulties concerning the available quantities, structural characterization, and resynthesis limit the scaling of this source of new lead generation. Perhaps ways to incorporate the diverse structures into synthetically convenient motifs will enhance natural products as a source of de novo ligand design.

Given limitations of time and expense for generating new lead structures, coupling library design with methods for virtual or in silico HTS are increasingly commonplace [Terstappen and Reggiani, 2001]. While it seems rational to design libraries around structural motifs previously shown to be active, it is important to maintain a balance for development of new chemistries and new motifs. A balance between focused or protein-family targeted library design and new chemistry-driven design is likely to assure the timely discovery of useful and novel chemical entities for drug discovery programs. As a result of these trends, many commercial compound library providers are increasing their efforts to design libraries around protein-family targeted themes, in addition to libraries often termed as diverse collections.

The necessary size of a compound collection presents an interesting theoretical and scientific problem. The size of a compound library is the result of balancing several features such as chemical tractability, diversity coverage, and design density. Obviously for an active series, design density will provide much SAR information. There are reports that the presence or absence of a single methyl group will make all the difference between an active and inactive hit structure. It is likely that a when balancing the density of diversity coverage by a particular library, one needs to focus on the correct fit of

innovation, technology, and proper chemistry meeting the practical reality of the challenges to finding an exploiting new lead structures for drug discovery. When searching for lead structures, however, it may be necessary to sacrifice some design density in favor of enhancing the diversity of a design to provide an initial hit which might be later optimized. The balance of this concern comes against the background of the enormous numbers of compounds that are theoretically possible to synthesize, numbers which range higher than are possible to synthesize. It would seem that a greater potential for providing coverage across the field of useful diversity exists by creating multiple chemotype library designs as opposed to a fewer number of library designs composed of many more compounds. Such a concept has been described as the "Fewer of many" as opposed to the "Many of fewer."

There are several other practical issues which come to bear on the question of appropriate library size. The capacity and cost of an organization's HTS technology will have some influence on the number of compounds that can be routinely screened. Although it may seem appealing to screen as many compounds as possible for every target, the cost-per-well consideration becomes a significant factor. Compound handling and tracking systems also influence the way in which an organization will determine the size of the libraries that it wishes to acquire. The distribution, archiving, and data management of screening compounds present significant challenges to process development and organization structures as the numbers of those compounds increase with each factor of ten. Many organizations see advantages of focused screening to accelerate the discovery of new lead structures at reduced cost. In effect, there is an increasing interest to shift from the brute force HTS approach to more focused, targeted and iterative screening practices.

It is in part for these reasons, that several companies continue the use of mixture screening. Although the trend in HTS has been towards the screening of high purity single compounds, the use of mixtures of compounds in conjunction with deconvolution routines continues as an integral part of the new lead generation in several companies.

We have observed a critical factor that determines any library's size is actually the chemistry itself. A diversity of reagents results in a diversity of reactivity. There are few reactions in organic chemistry that provide products in high yields for a wide diversity of chemical functionality. Further, the commercial availability of many reagents is quite limited except for a few key chemical classes. The response to these limitations has been to focus on the development of more robust library synthesis chemistry methodologies and to incorporate an increasing number of advanced building blocks.

LEVERAGING THE LIMITATIONS OF HIGH THROUGHPUT CHEMISTRY FOR BUILDING A LEAD GENERATION COLLECTION

The capabilities of high throughput chemistry are sometimes limiting with respect to synthesizing large, structurally diverse collections of high substance quality small molecules with drug-like and/or lead-like properties. Although astounding progress in the sophistication and complexity of structures possible by high throughput chemical methods is regularly reported [Dolle, 2000], expediency often requires shorter high throughput syntheses using advanced reagents and building blocks. The availability of appropriate building blocks facilitates faster library chemistry development and faster synthesis times of compounds in higher quantity and purity. Advanced building blocks also allow for flexibility in selection of the appropriate high throughput synthesis methods, such as solution phase parallel synthesis, solid phase combinatorial or a combination of solution phase synthesis and solid phase reagents. Frequently, building block acquisition has been determined by availability and supposed diversity; great collections of these building blocks have been stockpiled without extensive consideration of library designs for which they would be useful. More recently, the strategy to synthesize protein-family targeted libraries dictates the synthesis and acquisition of building blocks, according to particular design themes and motifs. In short, library design strategies should determine which building blocks to make and/or acquire.

The outsourcing of building block preparation for internal assembly into lead generation libraries provides many advantages. With appropriate planning, synthetically advanced building blocks may be readily diversified using convenient high throughput chemistry

methods. Advanced building blocks allow for the chemical diversity that may be inconvenient or inaccessible by a general high throughput chemistry method. The external acquisition of building blocks allows an internal lead generation group to focus on the core activity of production of lead generation of high throughput chemistry libraries. Outsourcing building block preparation may also ease general concerns with respect to retention of intellectual property rights associated with externalizing part of the lead generation process to library synthesis providers.

HIGH THROUGHPUT CHEMISTRY AS A MEANS TO BUILD A QUALITY LEAD GENERATION COLLECTION

The principal appeal of high throughput synthesis is the production of multiple, known small molecule structures in a short period of time. Compound characteristics for compound synthesis are often stated as diverse, drug-like, small molecule compound libraries. Strategies differ as to the number of useful compounds per library design from hundreds to tens of thousands. Practical realities of compound handling capabilities and HTS costs as well as the limitations of a particular chemistry scheme often come to bear on the question of appropriate number of compounds per library design. The necessary quantity of each small molecule is also determined by the in-house drug discovery processes; necessary minimum quantities vary from less than 1 mg/compound to more than 20 mg/compound. Synthesizing greater quantities of each compound allows for archiving of solid sample after distribution of the compounds to the HTS format. Should a particular compound prove to be active, follow-up activities are facilitated with the ready availability of that compound. The purity level of each compound is very important to most organizations. In order to reduce the false positive rate due to compound impurities (i.e., the rate that screening samples appears active during HTS, but later prove inactive in follow-up assays), many organizations define minimum purity criteria. Often purity levels are stated as $\geq 85\%$ pure. Unfortunately, many organizations do not specify explicitly the manner in which purity is defined. It is better to remove ambiguities resulting from different analytical methods and interpretations in order to determine the compound purity specifications more exactly (e.g., purity $\geq 85\%$ as measured according to $UV_{214\ nm}$ absorption in a validated LC/MS analytical system). The overall value of the library is increased if compounds which fail to meet these criteria are removed or purified before plating, distribution, and archiving of the library. The value of a library is maintained over time if the stability of the compounds contained therein is verified periodically.

In summary, there are increasingly stringent standards for lead generation libraries with respect to high substance quality as well as compound quantities and numbers. The explicit determination of these numbers is different for each lead generation operation. It is likely that the days of one lead generation library to fit all drug discovery operations is an aging concept.

DERIVING BENEFIT FROM LIBRARY COMPOUNDS FOUND TO BE ACTIVE DURING HTS

Once initial signs of activity have been detected for a particular design within a compound library, a great deal more information is readily available for "mining." Compound libraries from which a few hits are active provide an early glimpse of the SAR of that particular design. There is a compelling nature of a preliminary SAR based on a set of related hits as opposed to the activities for the same number of diverse, single compounds. Follow-up activities are particularly accelerated when working with a compound library design, since similar chemistry used to make the initial compounds is often readily amenable to synthesize related compounds to test follow-up questions. Further, the possibility to employ similar high throughput chemistry in the optimization phase adds to the appeal of finding and exploiting lead structures from high throughput chemistry derived libraries.

The value of any library design is ultimately defined by HTS and bioassay results. Results from multiple HTS assays for a particular library design may provide validation of a design idea and/or privileged structural motifs. For example, the design of a kinase inhibitor library may be evaluated by assay against a portfolio of kinase assays. If a significant number of the compounds in a specific library design are active against one or more kinase targets, then indeed, that design reflects chemical features sufficient

to confer kinase inhibitor activity. The collective information concerning relating sets of targets with set of structurally similar molecules provides what is known as an affinity fingerprint [Kauvar et al., 1995]. Such fingerprints, part of the chemical genomics concept, allow for linking information about biological targets to information about small molecules. One goal of chemical genomics is to increase the chances of finding potent ligands rapidly because of such information. Although there is little information in the public domain concerning screening of multiple library designs against multiple protein family assays, it is reasonable to assume a validation of protein family designs with privileged structures will provide numerous successes from the efforts that are on-going in many drug discovery organizations today. It is worth repeating that in whatever assay a particular library design (targeted or diverse) is active, the amenability for further optimization of the individual structures is a common requirement. Thus, preeminent among design considerations and strategies is the design for the lead-like and/or drug-like molecule.

BALANCE OF EFFORTS AND INTERESTS FOR LONG TERM ADVANTAGE: LEAD GENERATION AND LEAD OPTIMIZATION

The synthesis of many thousands of compounds for lead generation is a detailed, technology-driven effort requiring significant commitment of resources to implement the necessary processes that benefit from the high throughput synthesis approach. Methods for efficient HTC change rapidly, often requiring the incorporation of the latest technologies and their effective implementation. Particularly, it is the commitment of large numbers of chemistry personnel and expansive laboratory space to lead generation efforts (as opposed to medicinal chemistry efforts) which is perhaps the greatest challenge to pharmaceutical research organizations. Many pharmaceutical organizations may prefer to focus on their chemistry resources such as medicinal chemistry and lead optimization as opposed to extensive and expensive commitments to lead generation processes. Ultimately, the proper balancing of lead discovery and lead optimization chemistry efforts makes for a faster discovery process overall which may be easily tuned to frequently changing priorities. Significant effort and resources should be anticipated in dedicated lead generation efforts. Synthesis of compound libraries should also include plans for large-scale compound handling, compound identity and purity validation, registration, plating, and archiving. Proper planning and realistic anticipation of this work will allow for the appropriate time required to assimilate the compound libraries into the corporate discovery process and to benefit from the results of this work.

FUTURE PROSPECTS FOR DEVELOPING BETTER STRATEGIES OF LIBRARY SYNTHESIS IN NEW LEAD GENERATION

Areas for increasing the impact of lead generation efforts include better library design practices, in silico design methods, and methods for defining chemical diversity useful in drug discovery. Methods for design of new chemical systems, so called de novo design methods, to adapt pharmacophore information are particularly attractive to pharmaceutical organizations. The opportunity for directing library design according to virtual screening experiments will continue as common strategies [Langar and Hoffman, 2000]. Improved library designs guided by better predictive tools to direct synthesis according to ADME and solubility characteristics are becoming increasingly possible. Pharmaceutical organizations that are able to excel at these and other enhancements to the library design process will be in a position of advantage. There is a continuing need for novel, high throughput chemistries to provide lead-like molecules. Having populated corporate collections with molecules that are easily made, organizations now search for compound libraries that may be made only after non-trivial chemistry development efforts. It is unlikely that the demand for compound library fidelity and purity will decrease, and as a result, chemistry and analytical technologies which assure high quality compounds will continue to be regarded as value-adding. There remain large, developing possibilities for integrating lead generation with down stream activities such as hit-to-lead and lead optimization efforts in the most rapid and efficient ways.

CONCLUSIONS

Continued growth in the pharmaceutical industry requires greater success in the genera-

tion of high quality leads for many therapeutic targets. The meaningful impact for small molecule library synthesis for new lead generation is possible by integration of multiple functions such as library synthesis chemistry, design innovation, and quality control efforts that emphasize compound quality and production volume. Several strategies for the design of high quality lead generation libraries currently guide the selection of molecules to make by HTC methods. Subsequent screening of multiple library designs against multiple targets followed by data mining provides significant potential to improve the rate of discovery of quality lead structures. Emphasis on the synthesis of quality leads will likely facilitate a rapid lead optimization process. Given the need to accelerate the drug discovery process with new, better compounds, it would seem that there are few alternatives to engaging in the various efforts necessory for chemical synthesis of lead generation compound libraries.

ACKNOWLEDGMENTS

Dr. Steve Ferguson and Dr. Sung-Sau So are acknowledged for helpful discussion and critical reading of this manuscript.

REFERENCES

Arlington S. 2000. Pharma 2005, An industrial revolution in R&D *Pharma R&D*, 26–28 June 2000, Barcelona Spain.

Bailey D, Brown D. 2001. High-throughput chemistry and structure-based design: survival of the smartest. Drug Discovery Today 6:57–59.

Bemis GW, Murcko MA. 1996. The properties of known drugs. 1. Molecular frameworks. J Med Chem 39:2887–9283.

Dolle RE. 2001. Comprehensive survey of combinatorial library synthesis: 2000. J Comb Chem 3:477–517.

Dolle RE. 2000. Comprehensive survey of combinatorial library synthesis: 1999. J Comb Chem 3:383–384.

Dolle RE. 1999. Comprehensive survey of combinatorial library synthesis: 1998. J Comb Chem 1:235–282.

Ekins SI, Wrighton SA. 2001. Application of in silico approaches to predicting drug-drug interactions. J Pharmacol Toxicol Methods 45:65–69.

Fields S. 2001. Proteomics: Proteomics in genomeland. Science 291:1223–1224.

Hann MM, Leach AR, Harper G. 2001. Molecular complexity and its impact on the prbability of finding leads for drug discovery. J Chem Inf Comput Sci 41:856–864.

Jacoby E. 2001. A novel chemogenomics knowledge-based ligand design strategy-application to G protein-coupled receptors. Quant Struct-Act Relat 20:115–123.

Johnson BM, Babcock WC, West JB, Friesen DT. 1998. Preparation of bio-active compounds by plasma synthesis. U.S. (1998), 22 p. CODEN: USXXAM US 5772855 A 19980630 CAN 129:53942 AN 1998:430001.

Kauvar LM, Higgins DL, Villar HO, Sportsman JR, Engqvist-Goldstein A, Bukar R, Bauer KE, Dilley H, Rocke DM. 1995. Predicting ligand binding to proteins by affinity fingerprinting. Chemistry and Biology 2:107–118.

Langer T, Hoffman RD. 2000. Virtual screening: An effective tool for lead structure discovery? Curr Pharm Des 7:509–527.

Lipinski CA. 2001. Drug-like properties and the causes of poos solubility and poor permeability. J Pharmacol Toxicol Methods 44:235–249.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced Drug Delivery Reviews 23: 3–25.

Maloney PR, Parks DJ, Haffner CD, Fivush AM, Chandra G, Plunket KD, Creech KL, Moore LB, Wilson JG, Lewis MC, Jones SA, Willson TM. 2000. Identification of a chemical tool for the orphan nuclear receptor FXR. J Med Chem 43:2971–2974.

Patchett AA, Nargund RP. 2000. Chapter 26. Priveldge structures—an update. Annu Rep Med Chem 35:289–298.

Teague S, Davis A, Leeson P, Oprea T. 1999. The design of leadlike combinatorial libraries. Angew Chem Int Ed 38:3743–3748.

Terstappen GC, Reggiani A. 2001. In silico research in drug discovery. Trends in Pharmacol Sciences 22:23–26.

Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, Durkin AS, Gwinn M, Kolonay JF, Nelson WC, Peterson JD, Umayam LA, White O, Salzberg SL, Lewis MR, Radune D, Holtzapple E, Khouri H, Wolf AM, Utterback TR, Hansen CL, McDonald LA, Feldblyum TV, Angiuoli S, Dickinson T, Hickey EK, Holt IE, Loftus BJ, Yang F, Smith HO, Venter J, Dougherty BA, Morrison DA, Hollingshead SK, Fraser CM. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 291:1304–1351.