Pure Appl. Chem., Vol. 73, No. 1, pp. 67–75, 2001.

Continuing evolution of the drug discovery process in the pharmaceutical industry*

Emiliangelo Ratti† and David Trist

Glaxo Wellcome SpA Medicines Research Centre, Verona, Italy

Abstract: Many early discoveries in the pharmaceutical industry were through serendipity. Later, targets were mainly identified in animals and systematically exploited through the identification of potent and selective molecules. A disease association was normally obtained through the clinical testing of candidate molecules in patients. The technological advances in the last few years offer the possibility of knowing more about the disease, and this is driving the industry toward a disease-based approach where understanding the disease becomes central to the process. This is now possible thanks to the recent explosion in molecular and cellular biology, together with the application of genetics and genomics. New screening technologies have also revolutionized the identification of chemical leads. Now, high-throughput screening allows a wide chemical diversity to be applied in order to obtain tractable leads, which can then be optimized by the medicinal chemist. It is envisaged that these trends of continuously searching for process improvement will continue, being driven by the need to find medicines that add value in treating unmet medical need.

INTRODUCTION

The pharmaceutical industry has consistently shown that it can discover and develop innovative medicines for a wide range of diseases. This has been, and is being, accomplished against a backdrop of continually more stringent conditions in the market place (e.g., pricing policies, regulatory environment) and against diseases that have been traditionally resistant, until now, to pharmacotherapy (e.g., solid tumors). These latter continue to be a higher and higher proportion of the unmet need as other easier to treat diseases have become better controlled.

The philosophy by which valuable drugs have been discovered during the last 50 years has evolved from one that was mostly based around chemistry through a more biological approach to one that is concentrating more on disease. These changes were not only driven by strategic imperative, but are enabled also by the significant changes in technology that has occurred during this half century.

This article briefly describes the background that led to today's drug discovery process and attempts to predict what future practices might look like.

HISTORICAL BACKGROUND

A large part of the pharmaceutical industries' success had its origins in serendipity. In fact, long before such an industry existed, medicines were being discovered by accident, and their use was passed down by written and verbal tradition. For example, both digitalis and aspirin are active principals of natural products, namely foxglove leaf for dropsy, brought about by congestive heart failure, and willow bark for arthritis. Both of these remedies were described and used some hundred(s) of years before the iso-

*Lecture presented at the XVI[th] International Symposium on Medicinal Chemistry, Bologna, Italy, 18–22 September 2000. Other presentations are published in this issue, pp. 55–75.

†Corresponding author

lation of their active components. Thus, William Withering was struck by cases of dropsy (edema) that had been improved by herbal remedies, and he came to the conclusion that the active ingredient was foxglove. He published his results in 1785, but it was not until the 20th century that the cardiac glycosides were structurally and pharmacologically described. In a similar way, willow bark (*Salix alba*) was known by the ancients as an antipyretic, and in the early 19th century a glucoside of salicylic acid was isolated from it, and this was followed shortly by the isolation of salicylic acid which was shown to be antipyretic. Later, widespread use followed the recognition of its analgesic activity.

Serendipity was sought in the 1950s and 1960s in the pharmaceutical industry by screening known compounds or randomly testing any molecules that were at hand, to a large part in *in vivo* models. This method was not without success with drugs such as chlorpromazine, meprobamate, and benzodiazepines (chlordiazepoxide, diazepam) being discovered, all of which have gone on to become successful medicines.

The process is described in Fig. 1a. Lead molecules were found by chance or from screening the chemical diversity available. These were then optimized by medicinal chemists to produce candidates, which were passed to development and eventually into the market.

However, this approach at that time suffered from lack of sufficient molecules with high enough structural diversity, and the common use of animal models meant that other factors such as absorption, metabolism, brain penetration, and pharmacokinetics had profound effects on the number of active molecules found. In addition, many molecules that showed activity in the models were of unknown mech-

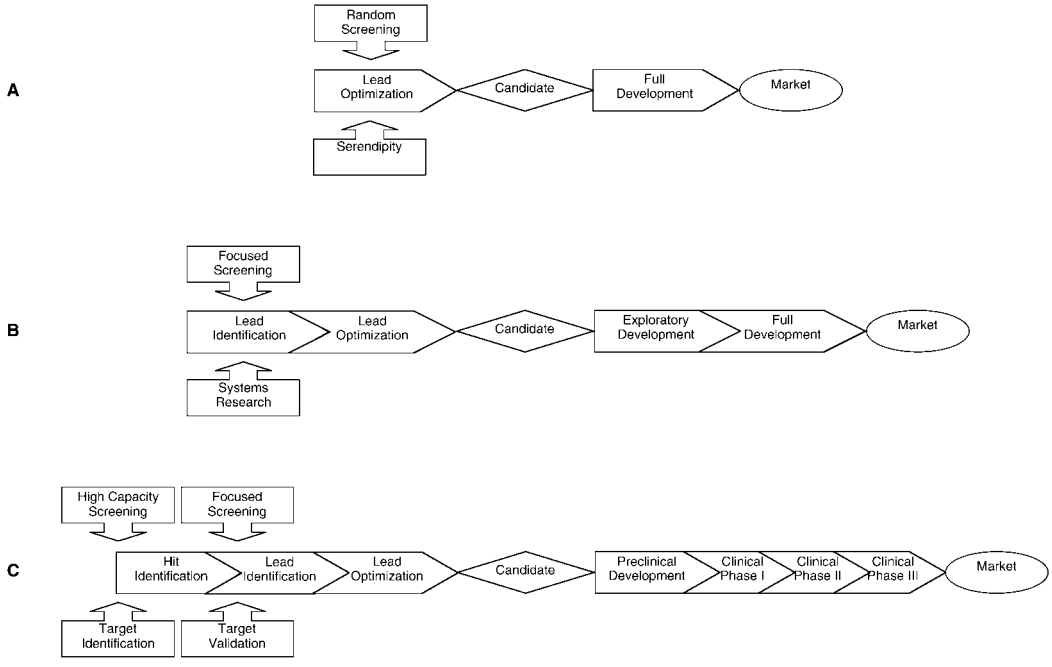


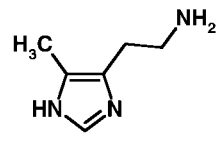
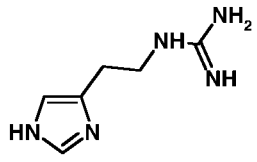
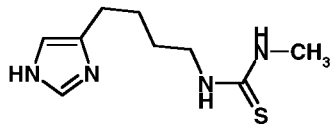
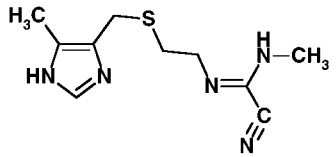
Fig. 1 Evolution of the drug discovery process during the second half of the 20th century. The process is centered on the crucial decision point that leads to the selection of a candidate for full development. The process used in the 1950s and 1960s (panel A), mainly driven by limited random screening and serendipity, was lengthened considerably during the 1980s (panel B) to allow for a more rational approach to structure–activity relationship studies and for improved safety of the molecules. Today's process (panel C) has been enriched by advances in technological developments in screening, synthetic chemistry, and by the increased number of possible targets due to the application of genomics and bioinformatics.

anism. This greatly impeded the development of back-ups when the lead failed due to toxicity or poor pharmacokinetics.

To combat these problems, a more rational approach was developed based around the structure of the agonist (i.e., hormones and neurotransmitters) and its receptor. This was set against a background of studying biological/physiological systems in animal tissues. Thus, knowledge around molecular determinants that contribute to affinity and efficacy enabled a generation of specific and potent agonists and antagonists to be developed. These included β-adrenergic receptor blockers, histamine H_2 receptor antagonists, β-adrenergic receptor partial agonists, and many more. For this approach, *in vitro* assays using animal tissues became central in the process for giving valuable information on structure–activity relationships and eventual pharmacophore construction. In this way, in theory, if the lead molecule fails there is sufficient information around structure and activity to allow whatever cause for failure to be built out of the molecule.

How this process works can be exemplified by the development of the H_2 receptor antagonist cimetidine. Black *et al.* [1] started with the observation that 2-methyl histamine selectively induces histamine responses that are blocked by mepyramine, whereas 4-methyl histamine activates mepyramine-insensitive responses. Mepyramine-sensitive effects had already been designated as characteristics of histamine H_1-receptors by Ash and Schild [2]. Thus, 4-methyl histamine was classified as an H_2-receptor agonist. The important key to moving from this observation to a selective antagonist was the finding that N^{α}-guanylhistamine was a weak partial agonist. By changing the basic group to thiourea and lengthening the side chain agonist activity (efficacy) is lost and affinity K_B is increased. This gave rise to burimamide, the first selective antagonist at 'H_2' receptors (Table 1). Cimetidine was then developed from this lead by chemical optimization to give a more potent, orally active analog of burimamide.

Table 1 Structures and pK_B estimates for histamine H_2-receptor antagonists.

Compound	Structure	pK_B
4-methylhistamine (H_2-receptor agonist)		
Nα-Guanylhistamine (partial agonist)		3.89
Burimamide (antagonist)		5.10
Cimetidine (antagonist)		6.10

Thus, the model described in Fig. 1a became modified to that shown in Fig. 1b. Now, leads can be identified from knowledge around the system, and, as before, they were optimized through the application of medicinal chemistry. However, by this time greater emphasis was being put on safety of molecules such that the development phase is now split into two parts—the part that is concerned with preclinical and early clinical safety (exploratory development) and the part that is required to show clinical efficacy (full development).

A consequence of this approach was the use of selective antagonists and agonists as aids to receptor classification. These were major tools for discovering new receptor subtypes in an age prior to the application of molecular biology.

The biggest problems that this approach did not address were how to find small molecules for receptors that bound peptides or proteins. In this case, it is more difficult to understand binding and efficacy determinants, and closely related receptor subtypes were often indistinguishable by antagonists when starting from known agonists that are not specific for the subtypes.

The use of animal tissues to develop selective drugs has its pitfalls. Activity in animals does not always translate into efficacy in humans. Thus, species differences in receptors between humans and animals or differences in the contribution of different pathways/receptors in animal and human pathologies can lead to no effects at safe doses in clinical trial. Thus, when molecular and cell biology grew to the point that it was possible to work with human receptors, and enzymes this was hailed as a significant advance.

THE DRUG DISCOVERY PROCESS TODAY

The advent of molecular biology, coupled with advances in screening and synthetic chemistry technologies, has allowed a combination of both knowledge around the receptor and random screening to be used for drug discovery. Probably, nearly all pharmaceutical companies today follow common technology process for discovering drugs. These include cloning and expressing human receptors and enzymes in formats that allow high-throughput, automated screening and the application of combinatorial chemistry. Thus, random screening can now be done with libraries sufficiently large and diverse to have a relatively high probability to find a novel molecule. These libraries are possible because they can be generated by the techniques of combinatorial chemistry (combichem). Importantly, human receptors are now routinely used for finding new molecules expressed in isolated recombinant systems and frequently in transgenic animals. This allows not only chemical optimization to be done against the human receptor, but also the possibility to tease out the physiological relevance of these receptors. For example, by controlling the expression level of receptors, particularly by overexpression, the property of inverse agonism has been disclosed and a similar overexpression of human β_2 adrenergic receptors in hearts of mice leads to the development of heart failure in old animals [3]. This is because β_1 adrenergic receptors normally mediate only cardiac stimulation to adrenergic agonists, but β_2 receptors can also inhibit contraction. Thus, when these are overexpressed, the potential for adrenergic agonists to cause inhibition of contractility becomes evident.

Backing all of this are important strides in the application of genetics and genomics to understand associations between diseases and gene products. Importantly, bioinformatics are beginning to identify putative targets for a number of diseases.

Thus, the process can now be represented as in Fig. 1c. This consists of target identification, target validation, hit identification, lead identification, lead optimization, preclinical development, clinical Phase I, clinical Phase II, Phase III, and launch into the market.

An early example of the application of this approach to obtain novel leads can be seen in the area of cholecystokinin (CCK) antagonists. CCK is a neuropeptide of 33 amino acids that was postulated to bind to two sub-types of receptors (CCK_A and CCK_B). This classification had been done on the basis of agonist potency around fragments of CCK. Basically, the sulphated octapeptide, CCK-8, has high

affinity for CCK_A receptors and low affinity for CCK_B receptors while the tetrapeptide, CCK-4, has a high affinity for CCK_B receptors and a lower affinity for CCK_A receptors. Such CCK receptor antagonists were known to be nonselective and of poor affinity. In 1985 scientists at Merck [4] discovered a potent (μM), selective CCK_A antagonist, asperlicin, by screening products from microbiological fermentation. A tractable lead was found from this hit, from the observation that structurally asperlicin can be considered a 1,4-benzodiazepine with a large 3-substituent derived from tryptophan. Thus, Evans *et al*. [4] suggested that diazepam combined with D-tryptophan might mimic asperlicin. In accord with this hypothesis, a series of 1,4-benzodiazepines were synthesized which maintained selectivity toward the CCK_A receptor and had improved affinity. The optimal antagonist found was L-364,718 (devazepide) which has a sub-nM affinity and greater than 1,000-fold selectivity for the CCK_A receptor [5].

An important consequence of today's approach is that more lead molecules are being discovered for diverse targets, giving the medicinal chemist more scope to find a candidate molecule. Because the human receptor/protein is being assayed it is much more likely that the eventual candidate will bind to its target in humans. In addition, problems related to understanding the complexities of large molecule interactions with proteins are eliminated, at least initially.

However, the probability of success of launching a candidate molecule into the market remains basically unchanged at around 10% [6], and the perception is that also time to market is not reducing as might be expected with greater automation. In fact, it might even be rising!

So what are the drivers that will cause the process to change, and, importantly, what will be the technological enablers that can/will be applied to achieve this transformation?

TOWARD THE FUTURE

The pharmaceutical industry is increasingly operating in a world where medicines have to add real value in an environment where costs are under constant pressure. This is the background to the drivers that are causing the evolution of the drug discovery process. In the future this process will have to be more efficient and quicker to deliver a higher percentage of pipeline molecules to the market.

Unmet medical need

A constant driver for developing new medicines has always been the unmet medical need. However, there are now strong pressures to treat the underlying pathology rather than give solely symptomatic relief. This is leading the industry toward a more disease-based approach where understanding better the human pathology should deliver targets that are involved in the causative processes of the disease. In some ways, this is reinventing the biological systems approach, but using humans rather than animals.

In order to accomplish this, the investment that has already been initiated in technologies such as noninvasive imaging, clinical genetics and genomics will increase. This is now assured with the publication of the human genome.

The lack of disease models in animals in some therapeutic areas is another major driver to understand the human pathology. This is particularly relevant in the CNS area. Thus, the complexity of diseases such as major depression, bipolar disorder and schizophrenia has so far not been modeled in animals. Therefore, in these diseases we are left with targeting components such as receptor or biochemical systems, with no simple ways to validate these targets in the complex intact system.

In these cases, the scientist is constrained to collecting a logical series of evidence that associates the target with the disease. This process will certainly be strengthened by the application of the technologies mentioned above. Already, imaging methods such as PET and fMRI are being used to understand the correlation between disease and specific receptors. Clinical genetics networks are being put into place to allow sufficient probands to be collected, such that associations between particular gene(s) and disease can be made and ultimately lead to target validation and eventually identification. The dis-

covery that the time to Alzheimer Disease onset is significantly associated with having the homozygous allele of APOE4 was the first robust finding of a disease-associated gene [7] and is pointing the way to more of such findings. Unfortunately, no drug has yet been developed based on this evidence.

The advent of the human genome's publication now offers a great opportunity for the understanding of the genetic make-up of disease and will furnish specific gene products and/or pathways as new targets that would not have been previously identified. Importantly, they will be born out of human data, so again adding to the level of confidence in the validity of the target.

Attrition

A driver of change that is becoming increasingly important is the overall success rate of the drug discovery process. As mentioned above, attrition has remained static despite the investment in the new technologies. This reflects the fact that good molecules need more than potency and selectivity to be successful, and it is in these areas where technology has been concentrating in the last few years. The challenges ahead lie in reducing the risk of not obtaining efficacy in humans, and in increasing the developability of the molecules.

Efficacy

Many new mechanisms fail when they get into humans through lack of efficacy. This is one of the risks that the industry takes when developing such molecules. One way to diminish risk is to get better validation in humans as soon as possible. The use of imaging, genetics, and genomics has already been discussed above as a way to help build early confidence in the target. Clinical readouts as early as possible are now being sought as part of making decisions as early as possible. It is now recognized that fast decision making saves money and allows resources to be more effectively used. In addition, killing compounds in Phase III is extremely costly. Thus, simple proof-of-concept studies (POCs) are being sought in Phase I or Phase II. The philosophy behind these is that sufficient evidence can be generated in humans that allows the molecule to go forward. In some cases, the POC might be very simple. New antibacterials are known to work from extensive *in vitro* and *in vivo* studies. Thus, the issue in humans is to find a safe molecule with the right pharmacokinetic profile. This can be accomplished in Phase I and could well constitute the POC. Whereas, a new stroke drug until now has only been shown to be active, or not, in Phase III. The application of a smaller Phase II study using MRI to follow structural damage progression post ischaemic event is one way in which some evidence of efficacy in humans can be obtained to allow decisions on progression to be made. This approach, using small numbers of patients or volunteers, will be actively followed in the future over a wide range of diseases.

In addition, diagnostics will play a greater role in helping to choose patient populations, at least initially to show that the mechanism works. This will see greater and greater use of imaging, proteomics and genetics in helping to identify the right patient group.

In the meantime, a better balance of novel molecules and those that are precedented will be seen in the drug discovery portfolio. This will mean that a higher proportion of molecules will not fail for efficacy. However, this strategy creates its own problems in that to be successful in the marketplace the molecule will need to be differentiated from those already present. To do this in the clinic will add to the cost and to the overall cycle time (see below), thus these problems will need to be addressed much earlier in the process.

Developability

A large proportion of all molecules that fail do so because of lack of developability. Prentis *et al.* [8] suggest that this proportion is as high as 69%, broken down as toxicity (22%), poor biopharmaceutical properties (41%) and market reasons (6%). This is not a new revelation, and efforts have been actively followed to automate and miniaturize methods to measure solubility, stability, pKa bioavailability, brain penetration, and hepatotoxicity, the main cause of toxicity. These methods (combinatorial lead optimization) are being applied to leads during optimization, but need to be developed further and applied

even earlier to maximize their impact. This is particularly true for toxicity screens, where it can be predicted that much effort will be done in the next few years.

Much work is being done in the field of predictive algorithms, and Pfizer has developed one known as the rule of 5 [7]. This is an awareness tool for medicinal chemists that suggests that there will be poor absorption if a molecule has two or more of the following: more than 5 H-bond donors; a molecular weight >500; c log P >5; the sum of Ns and Os (a rough measure of H-bond acceptors) >10.

While it is inherently costly to try to fix poor developability by formulation, pharmaceutical development will become more actively engaged in alternative formulations and delivery systems during the lead optimization phase. The trend toward higher potency compounds, that reduces cost of goods, also allows, due to the smaller dose, alternative delivery systems such as inhalation, nasal, buccal, and sublingual absorption.

Cycle times

Another driver of process evolution is to deliver molecules to the market quicker. The regulatory environment and the growing complexity of drug development have been potentially driving times within phases up. Screening automation and combichem have greatly reduced the time to candidate selection. This will almost certainly decrease again by further application of techniques like chemoinformatics to aid library design, both for those to be used for random screening and those within the process of lead optimization.

As mentioned above, continual automation of developability criteria will also speed up the process by selecting out compounds with a high probability of not succeeding. This raises the concept that speed in each phase should not always be the major driver. A candidate for development goes forward with all of its associated baggage. Fixing problems becomes costly and may lead to a suboptimal product that can not fulfill its medical and commercial potential. Thus, spending time choosing the right candidate will have major benefits downstream, both in terms of speed and value. The same concept applies to development candidates in Phase III. Differentiation may not be obvious if the mechanism is precedented with another marketed product. Thus, differentiation will become a challenge, which potentially will increase the time in Phase III. To aid in this process and help in choosing which differentiators to pursue, this problem will need to be addressed much earlier. This might stimulate automated assays for common side effects of drugs as part of the candidate selection criteria during the lead optimization stage.

Economic value

Value in a number of guises has become a key driver of the pharmaceutical industry.

There is growing internal pressure to increase productivity while controlling costs. This has led to the drive for high-value molecules in diseases with high unmet need. An extension of this concept is the "blockbuster" approach where projects that deliver medicines with potential peak sales greater than £1 billion are given the highest priority. This means that portfolio management will become more and more important with an associated greater interaction between R&D and the commercial functions. Thus, new portfolio tools will also be major contributors to the future process of drug development.

The real value of medicines to the health of society is only now beginning to be recognized. It has taken many years of persuasion that medicines can have profound economic benefit. One recent example is in the smoking cessation area. It has been known for many years that smoking can damage a person's health and passively damage those within his/her immediate environment. Even though one out of two people who smoke today will die from a smoking-related cause, smoking has been seen to be very much an individual's (bad?) choice with little thought to the costs involved, both socially and financially. However, it is now recognized that each year smoking costs about $130 billion in medical costs, and because tobacco often kills smokers in their prime working life, family income may be

reduced by years due to premature death. Importantly, due to the long-term nature of the disease, reducing smoking in adolescents (which is rising) will not have a short-term economic benefit, whereas reducing smoking in adults produces immediate effects on mortality and morbidity. Recently, reimbursement for smoking cessation aids has been granted or is under consideration, a position much different from the recent past.

The push to raise health, economic, and quality-of-life issues has produced a counter response from some regulators that the industry demonstrates added value in its novel medicines. Thus, committees like National Institute for Clinical Excellence (NICE) in the UK will put pressure on the process to produce medicines that have significant value for society. This will mean that in the future more outcome studies will be needed to demonstrate quality-of-life and economic benefit.

CONCLUSIONS

During the last ten years, technology changes have enabled the process of drug discovery to evolve into a system where new lead molecules can be rapidly found against novel, and sometimes, difficult targets. This process will continue. In the near future, lessons learnt around automating pharmacological assays will be applied to assays in other parts of the process.

The future challenge will be how to identify disease targets in humans and then to validate them. It is in this realm that perhaps the greatest changes will be seen. In particular, the "right drug for the right patient" will become more of a reality. This will be driven by the application of diagnostics and prognostics. It may be that eventually the industry will make its money by selling diagnostics and the medicines become associated with smaller niche markets, defined by those diagnostics.

Another forecast is that the number of targets will increase dramatically as the human genome data becomes exploited. This means that the industry itself will not be able to take advantage of all of these. Thus, it can be envisaged that more and more strategic alliances will be formed between biotechnology and small pharmaceutical companies to make the most of all of the possible opportunities. In addition, future use of contract houses throughout the process may mean that the industry will move toward a model as shown in Fig. 2, where the core internal business is concentrated from lead generation to POC. The rest of the process is managed internally, but executed externally.

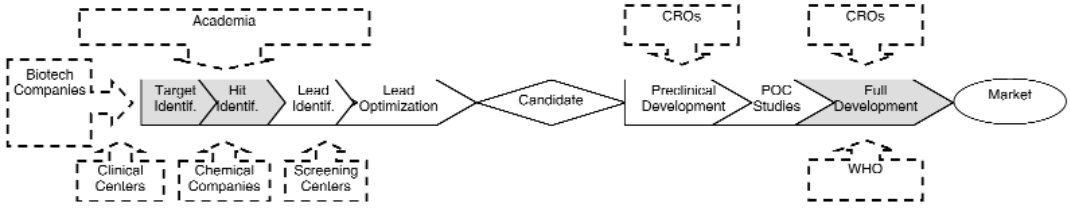


Fig. 2 Schematic description of the possible process of drug discovery in the future. The core internal business (solid-line blocks) is concentrated from lead generation to POC. The rest of the process (gray blocks) is managed internally, but executed externally by different external organizations (dotted-line blocks). This operational model will guarantee that all possible opportunities to cure diseases will be exploited (through strategic alliances with biotech and small pharmaceutical companies), that technological advances are fully exploited (by the use of contract houses both in the early and final phases), and that the real value of medicines to society is recognized by the close collaboration with organizations such as the World Health Organization (WHO) and clinical centers in all stages of development.

REFERENCES

1. J. W. Black, W. A. M. Duncan, L. J. Durant, J. C. Emmett, C. R. Ganellin, M. E. Parsons. *Nature* **236**, 385 (1972).
2. A. S. F. Ash and H. O. Schild. *Br. J. Pharmacol.* **27**, 427 (1966).
3. J. W. Black. In *The Pharmacology of Functional, Biochemical, and Recombinant Receptor Systems*, T. Kenakin and J. A. Angus (Eds.), pp. 1–11, Springer-Verlag, Berlin (2000).
4. B. E. Evans, M. G. Bock, K. E. Rittle, R. M. DiPardo, W. L. Whitter, D. F. Veber, P. S. Anderson, R. M. Freidinger. *Proc. Natl. Acad. Sci., USA* **83**, 4918 (1986).
5. B. E. Evans, K. E. Rittle, M. G. Bock, R. M. DiPardo, R. M. Freidinger, W. L. Whitter, G. F. Lundell, D. F. Veber, P. S. Anderson, R. S. L. Chang, V. J. Lotti, D. J. Cerino, T. B. Chen, P. J. Kling, K. A. Kunkel, J. P. Springer, J. Hirshfield. *J. Med. Chem.* **31**, 2235 (1988).
6. R. A. Lipper. *Modern Drug Discovery*, January/February, 55 (1999).
7. W. J. Strittmatter and A. D. Roses. *Proc. Natl. Acad. Sci., USA* **92**, 4725 (1995).
8. R. A. Prentis, Y. Lis, S. R. Walker. *Br. J. Clin. Pharmacol.* **25**, 387 (1988).