Article

Drug Discovery in the New Millennium

Steven G. Aitken

Inotek Pharmaceuticals, Christchurch (e-mail: SAitken@inotekcorp.com).

Introduction

Over the years there have been numerous milestones in the development of what is now regarded as the modern day drug discovery process. Medicinal chemists recognise that their roots can be traced back to the late 1920s when Erlich[1] pioneered the use of synthetic sulfonamides as antibacterial agents. In the late 1930s the discovery by Alexander Fleming[2] of the penicillins heralded what would become the wide-spread exploitation of natural products as therapeutic agents. In early 1960s Snyder pioneered the use of grind and bind receptor binding assays[3] while many also investigated the now familiar biochemical principle of enzyme inhibition. This article will concentrate on developments in the field of drug discovery since the mid 1980s.

Today drug discovery is a truly interdisciplinary undertaking. While the focus of this article is the multidisciplinary nature of the science behind modern day drug discovery, it is important first to consider the current economic and regulatory environment within which pharmaceutical companies operate. The key issue for financial success is a recognition that drugs cost billions of dollars to develop; it is estimated that $US 49.3 billion was invested in drug development in 2004 and it is important to appreciate that the release of every new drug is preceded by years of expensive (and time consuming) laboratory work and clinical testing, the latter representing the greatest cost. Thus pharmaceutical companies expect and require a significant payback from each new drug to compensate for the vast sums invested in its development.[4]

The Drug Discovery Process

In principle the drug discovery process is relatively simple with overall success dependent upon the correct choice of biological target, the correct disease association for the chosen target, and the correct choice of a drug development candidate for successful human clinical trials. In practice the process of drug discovery uses just five steps i) selection of the biological target of interest, ii) identification of a lead chemical compound (*hit*), iii) optimization of the lead structure, iv) animal trials to ensure the drug is efficacious in animal disease models and is safe for human clinical trials v) human clinical trials that prove the drug is efficacious and safe for use in the general patient population.

The record $US 49.3 billion invested in drug R&D in 2004 suggests that successful drug discovery is a much more complicated and difficult undertaking than appears from this simple analysis. In reality, research and development is a lengthy process requiring 10+ years from concept to clinic. Let us consider each of the drug discovery process steps in turn.

Selection of the biological target

The key question in selecting a biological target is: *Will therapeutic intervention have a positive effect on managing the disease(s) in question?* This forms the first phase of drug discovery, and improved success here would markedly reduce failure rates.

Nowadays this phase is absolutely critical to the drug discovery process. The human genome project has yielded *ca*. 35,000 potential targets for treating disease, a huge increase from the *ca*. 3,000 at the turn of the millennium. However, only a fraction of the 35,000 biological targets can currently be exploited because much of the human genome function is unknown. This vast amount of new information has provided, and continues to provide, new and exciting opportunities for novel therapeutic design.

Lead generation

Traditional medicinal chemistry identified a lead compound by screening collections of naturally occuring or laboratory synthesized compounds. Over time, computerization catalysed the development of *high throughput screens* (HTS) which allowed hundred of molecules to be tested in the time previously required for one. The development of HTS in turn drove medicinal chemists to develop methodology which would allow more compounds to be synthesized in a shorter time. As such the 1990s witnessed the widespread use of combinatorial chemistry to create *libraries* of compounds.

Library generation generally involves attaching a *skeleton* to an insoluble polymer support that is then subjected to sequential reactions (excess reagents drive the reactions to completion) and purification (excess reagents and by-products removed by washing). This methodology generates a large number of product molecules, or *library*, in which all possible combinations of a series of reactants have been obtained. The name *combinatorial chemistry* was coined to describe this process and today combinatorial chemistry is most easily defined as *the synthesis of chemical compounds as ensembles (libraries) in the form of mixtures or discrete compounds*.[5] Typically compounds are stored in a 96-well plate and the *library* screened by a high throughput procedure.

During the 1990s combinatorial chemistry was perceived by many to be the ultimate *drug discovery factory* and its introduction was embraced with unprecedented enthusiasm. Initially, synthesis of combinatorial mixtures was achieved relatively easily and quickly using a process of *mix and split* (or *split and pool*) strategies. For example, the preparation of all possible tripeptides from the 20 naturally occurring amino acids using solid phase synthesis (Fig. 1) involves coupling an *N*-protected amino acid [A] to a solid phase resin, deprotection, addition of a second

N-protected amino acid [B], removal of excess reagents (repeated washing) followed by deprotection of function [B], and addition of a third *N*-protected amino acid [C]. Repeated washing of the resin, followed by cleavage of the peptide-resin bond then affords unprotected tripeptide [A] [B] [C] in solution. By using *split and mix* techniques, the method can be adapted give all possible combinations of the 20 amino acids, *i.e*. A-A-A, A-A-B, A-B-A, B-A-A, *etc*., leading to a 20 x 20 x 20 array of 8000 tripeptides.

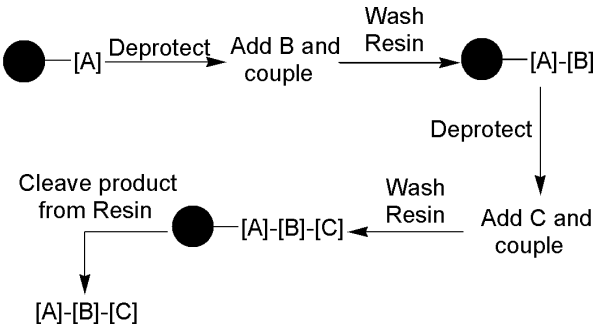


Fig. 1. Synthesis of a tripeptide by solid phase synthesis.

These early manual combinatorial strategies were very reaction- and vessel-efficient processes. However, the product library could not be rescued by purification when chemical conversion was incomplete or when side-products were generated. Moreover, a biologically active synthetic mixture required time consuming deconvolution to determine which compound was responsible for the activity. Even worse was the fact that combinatorial mixtures tended to give false positive results such that the labour- and time-consuming deconvolution process did not yield a single active compound responsible for the biological activity. These problems all but consigned this method of lead generation to the history books. Nonetheless, combinatorial chemistry is still an integral part of the lead generation process. However, only libraries of discrete compounds are now prepared for screening using modern combinatorial chemistry; the most successful applications of which pay great attention to the essentials of high throughput purification.

It is now approximately a decade since the large scale roll-out of combinatorial chemistry occurred and, given that that most drugs take approximately this time to reach the market, the significance of combinatorial chemistry should now be evident. This is clearly not the case as only 21 new drugs were approved by the FDA in 2004.

It should come as no surprise to most scientists that blind screening of large combinatorial libraries for drugs has had few successes. The libraries of the 1990s focused on increasing the number of compounds available for screening based on the naive assumption that this would result in more chemical leads being generated. Sadly, few chemicals make suitable drugs, and the ones that do are not uniformly distributed throughout *chemical space*. To counter this numerous models have been, and currently are being, developed to define the *drug-like* characteristics of a particular compound. The best-known, and most utilized,[6] model is the so-called *Lipinski rule of 5*. It derives its name from Pfizer employee Christopher A. Lipinski who, with others, decided that the relevant cut-offs for a *drug-like* compound were all multiples of 5. While the rule excludes most *undrug-like* compounds it does not guaranteed a *drug-like* one will succeed. The model assumes that most chemicals will not be suitable as oral drug candidates if they lack bioavailability. Accordingly, the rule states that for a compound to be absorbed through biological barriers it must not contravene more than two of the following:

- It should have no more than 10 hydrogen bond acceptors.
- It should have no more than 5 hydrogen bond donors.
- Its relative molecular mass (M_r) should be below 500.
- Its log P value[7] should be less than 5.

Despite exceptions (notably for antibiotics), approximately 90% of the top 200 drugs follow the rule. In light of this, scientists at GlaxoSmithKline studied the Lipinski properties of both the top 200 selling drugs and the lead compounds generated by high throughput screening of their combinatorial libraries from the late 1990s and early 2000s. From this analysis it was concluded that both the leads and the top selling 200 drugs follow the rule, especially with respect to M_r and the numbers of hydrogen bond donors and acceptors. However, and most significantly, leads generated by HTS tended to have higher log P values. Although the differences were small, it is widely acknowledged by medicinal chemists that in the process of transforming a chemical lead into a selective orally active drug, both the M_r and log P tend to increase; to undertake lead optimisation on a chemical that has higher than average log P is unwise.

As a solution to this problem, GSK, and many other companies now use computer-based virtual screening of virtual libraries (VSVL)[8] to identify the best compounds for combinatorial synthesis. Much of the detail is proprietary with few examples of VSVL in the open literature. However, computational chemists confirm that the methods work and a glut of start-up companies specializing in virtual screening now exist.

The potential and success of combining VSVL, combinatorial chemistry and HTS is best illustrated by GlaxoSmithKline. Table 1 shows the improvement in screening success rates, numbers of lead compounds generated per screen, and the potency of the lead from 1998-2004. The success is attributed largely to use of VSVL in the design of the combinatorial libraries for HTS screening.

Table 1. GlaxoSmithKline HTS scorecard.[a]

	1996	1999	2003	2004
Compounds screened	100,000	430,000	615,000	1,050,000
Average lead potency	3,000 nM	400 nM	10 nM	10 nM
Screen success	20%	50%	58%	65%
Leads per target	1.0	1.7	1.9	2.0

[a]Data taken from: *http://pubs.acs.org/cen/coverstory/8230/8230drugdiscovery.html*

Lead Optimisation

Perhaps the most challenging stage in drug development is to take to full development that compound with appropriate biological and chemical properties. Formally, the objective of lead optimisation is the design, synthesis and selection a drug development candidate that has the best balance of potency, selectivity and ADMET (Absorption, Distribution, Metabolism, Excretion, Toxicity) characteristics.

For most drugs the best route of administration is oral. Thus, even if a compound is very potent against its biological target *in vitro*, translation of this into *in vivo* activity in a human patient requires the understanding and optimisation of a number of biological processes. Firstly, the drug has to be taken into the blood stream and then carried to its effector site where it is then adsorbed by the target organ or cell. For example, if the target is in the central nervous system (CNS) the natural blood-brain barrier has to be crossed. Even when the drug reaches its target, its rate of metabolism has to be appropriate and it must be sufficiently stable to remain there long enough for the desired biological effect to be achieved. In some cases a prodrug approach is required. Here, chemical modifications taking place within the organism are pre-requisites for the biological activity of the chemical compound. Additionally, once metabolised, compounds have to be excreted to minimise bioaccumulation in organ or tissue.

A lead optimisation program needs to ensure that correct physicochemical properties are imparted to the potential molecule for the drug to be both efficacious and safe in humans. Lead optimisation still relies heavily on traditional medicinal chemistry techniques. Examples include quantitative structure activity relationships (QSAR), conformation constraint, pharmacophores, isosteres/bioisosteres, and heterocyclic similarity; the most commonly used strategies in lead optimisation, by far, are the last two. Isosteres are substituents or groups with similar molecular size and/or volume. A bioisostere extends this principle to groups with similar pK_a values, lipophilicity, and electrostatic properties. Each bioisostere within a group is expected to have similar interaction (and hence biological activity) with the target. Heterocyclic similarity can be thought of as a further extension of bioisosterism where one heterocycle is replaced by another with comparable pK_a, lipophilicity, and electrostatic properties.

The major change to lead optimisation over the last decade or so is the introduction of focused combinatorial chemistry. Synthesis of mini-libraries facilitates expeditious synthesis of closely related analogues. This form of combinatorial chemistry is more reliant on feedback and less on force of numbers in defining a quantitative structure-activity relationship (QSAR); it uses the principles of isoseteres/bioisosteres/heterocyclic similarity to define the synthetic targets.

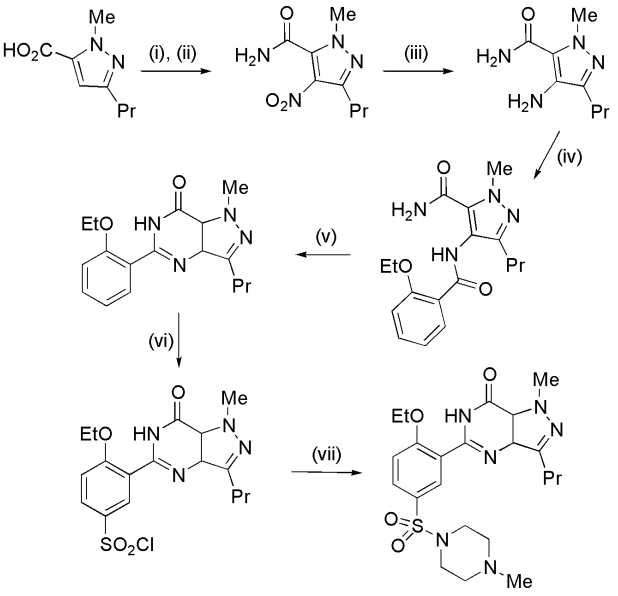
Chemical development

A very important stage of the drug discovery process that is often overlooked is that, after selection of a drug candidate for clinical trials, a commercial synthesis must be designed and implemented.[9] As can be seen from Table 2 the objectives and priorities for laboratory and chemical development syntheses are very different.

Table 2. Priorities for laboratory and commercial synthesis.

Laboratory synthesis	Commercial synthesis
Scale < 100 g	Scale > 100 kg
Cost of synthesis not relevant	Cost of synthesis vital in maximising potential profit
Overall yield not so important	Overall yield critical as above
Synthetic scheme designed to maximize diversity	Convergent synthetic scheme
Any reagent/reaction	Potentially dangerous/toxic reagents avoided if possible
Environmental impact not considered	Environment impact assessed

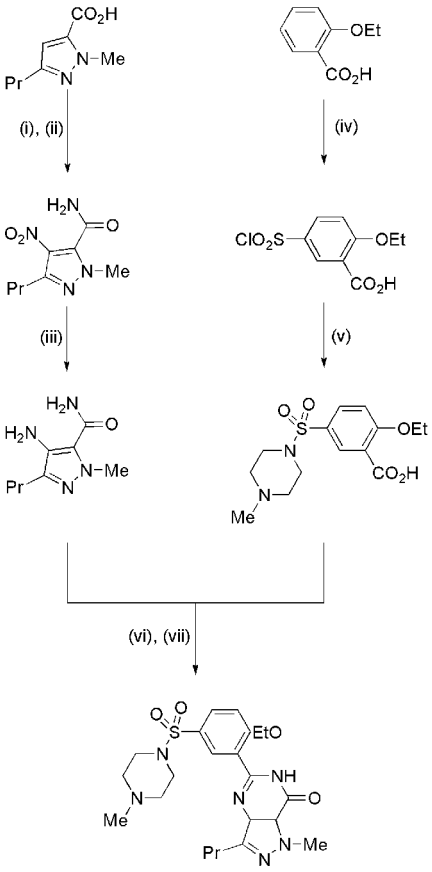
A classic example of this difference in priorities is illustrated by the development of a commercial route to Sildenafil (Viagra). The laboratory synthesis (Scheme 1) had several commercial disadvantages in that (i) it was linear with a low overall yield (7.5%), (ii) toxic materials were used in the final bond-forming step that was difficult to scale up, and (iii) the $SnCl_2$ used for nitro group reduction is too toxic for large scale use.



Reagents: (i) HNO_3, H_2SO_4; (ii) a) $SOCl_2$, b) NH_3; (iii) $SnCl_2$, HCl, EtOH; (iv) 2-ethoxybenzoyl chloride, Pyr, DCM; (v) NaOH; (vi) $ClSO_3H$; (vii) *N*-Methylpiperazine. Overall yield - 7.5%

Scheme 1. Laboratory synthesis of Viagra

In contrast, the commercial route (Scheme 2) is convergent and higher yielding (75.8%) with a *clean* cyclization as the final step and all potentially toxic materials used early in the synthesis. The $SnCl_2$ reduction has been replaced by catalytic hydrogenation. An added bonus is that the commercial route uses *ca*. 10% of the organic solvents required for the laboratory synthesis, making it both more environmentally friendly and cheaper; two solvents replace the previous six (Table 3).



Reagents (i) HNO_3, H_2SO_4; (ii) a) $SOCl_2$, toluene b) NH_3; (iii) Pd/C, H_2, EtOAc; (iv) $SOCl_2$, $ClSO_3H$; (v) *N*-Methylpiperazine, NaH, H_2O_2; (vi) DCC, EtOAc; (vii) *t*-BuOH, *t*-BuOK. Overall yield - 75.8%

Scheme 2. Commercial synthesis of Viagra

Table 3. Solvent volumes (L) required to produce 1000 kg of Viagra.

Solvent	Laboratory Route (L)	Commercial Route (L)
Toluene	39,000	2,800
Dichloromethane	37,000	
Ethyl acetate	19,000	10,700
Acetone	18,000	
Butan-2-one	10,000	
Pyridine	2,000	
Total	125,000	13,500

Clinical trials

Once a drug development candidate is identified clinical trials must be conducted so as to demonstrate to the regulatory authorities that the drug is safe, efficacious, and can be manufactured to acceptable quality standards. However, before such trials, a company performs careful cost/benefit exercises to ensure the potential for profit outweighs the risk (Fig. 2). Even at this stage development of a drug demonstrating promise as an effective therapeutic agent will be terminated if it is not seen to be profitable.

If the cost/benefit analysis indicates a profitable product full clinical trials will be instigated.[10] This phase of the discovery process is known to be comprised of four separate parts. However, in addition to these (and before candidate selection is made) animal studies are conducted—the pre-clinical phase. To illustrate this, the case history of Avandia follows.

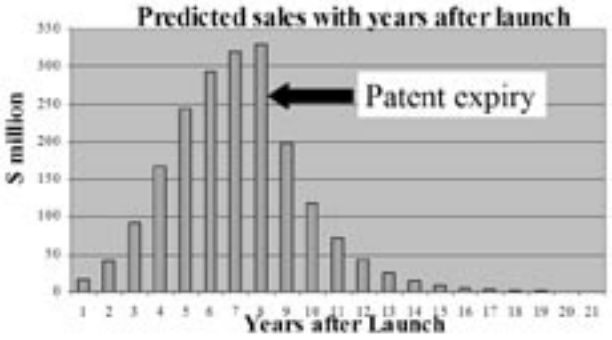


Fig. 2. Drug profitability time-line.

The pre-clinical phase

Typically taking about a year and involving several potential development candidates (*pre-candidates*) this is relatively cheap and can help to ensure that the best of a number of closely related leads is chosen. The phase is essential before the FDA (or other regulatory body) allows human trialling and involves at least two different animal species. The level at which toxicity develops and those organs most susceptible are established; the treated animals are sacrificed and autopsies performed. The studies involve a single high dose to define the gross toxic effect, and repeated dosage at various levels and durations to determine toxic effects. These provide the gross toxicity of the particular compound class or the specific candidate, and lead to an initial understanding of the drug metabolism in the animal species and whether such metabolism is likely to be harmful in human trials. The cost to this point is *ca*. $NZ 600,000, but the likelihood of the drug eventually reaching the market is still below 10%.

Phase I trials

Typically taking about a year and usually consisting of three separate studies, a *single dose* study in healthy male volunteers, a repeat *dose rising* study in healthy volunteers, and a *food interaction* study in healthy volunteers. The aim is to ensure that the drug is well tolerated in healthy humans with those safety issues identified from pre-clinical animal experiments appropriately assessed. The cost to this point is $NZ 5.4 M while the probability of the drug reaching the market is now 15%!

Phase II trials

Taking up to two years, this is much more expensive than the first studies and it is often separated into phases IIa and IIb. Phase IIa is a *proof of concept* stage where carefully selected patients are used to demonstrate that the compound is efficacious. Phase IIb is similar but uses larger patient groups, generally with multiple dosing arms. The data gathered from these studies are used to refine dosages and the parameters for the even more expensive phase III study. Several such pilot trials may be necessary to establish optimal dosage but these also help define the best way to administer the product and the best way to measure its benefits.

The aims of the phase II trials are to demonstrate efficacy and safety in patients with the medical condition for which the drug is designed, and to gain appropriate data to define the most efficacious way for drug administration and how to best measure the benefits obtained. This information allows for the design of a larger phase III trial with sufficient patients to determine if the benefits provided outweigh the risks of toxicity. The total costs to this point are *ca*. $NZ 80 M but the probability of success is close to 50%.

Phase III

Before commencing this, the most expensive phase, a further cost/benefit exercise is performed using updated information to ensure that the potential for profit still outweighs the risk. Even though vast amounts of money have been spent and the probability of success is now high, the development of a potentially non-profitable drug can be terminated at this point. Phase III trials typically take a further one-to-two years and are conducted on large patient populations. Usually the format consists of at least two global trials at two dose strengths with approximately two thousand patients per trial. They are almost always double-blind, placebo-controlled and are the *gold standard* for substantiating product effectiveness; these trials may also contain a drug that is already on market so as to demonstrate that the new drug is better, in someway, than the one currently on the market.

The ultimate objective here is to obtain sufficient data that the regulatory authorities accept that the drug works and is safe. There is no guarantee that marketing approval will be granted and the FDA review may take 18 months. By now a total of $NZ 1.1 billion has been spent with a 90+% probability of the drug reaching the market.

Phase IV

In this final phase the risk-to-benefit ratio is evaluated in the marketplace to ensure that the drug performs as expected, after all the marketplace is the ultimate test of product value. The system is not perfect and the FDA has withdrawn several drugs because their demonstrated toxicity could not counterbalance their benefits; the withdrawal of Vioxx last year provides a well published example.

Is drug discovery financially worthwhile?

One may wonder whether the huge financial costs involved can be justified. The R&D costs of Avandia was $NZ 1.1 billion, it took some 12 years to get to the market, and, with exclusive patent rights of 20 years, approximately 8 years remained for the drug to recover its R&D costs and return a profit. For a blockbuster drug such as Avandia with sales of NZ 2.5 billion in 2003 it has to be worthwhile. However, it needs to be remembered that only 3% of compounds that make it to clinical trials reach the market place.

Summary

Drug discovery appears to have moved full circle. In the 1960-1970s molecules were painstakingly synthesized and only slowly screened with structure-activity relationships guiding further synthesis. In the late 1980s and early 1990s high throughput screening and combinatorial chemistry made possible the testing of tens of thousands of compounds per year. This led to enormous volumes of biological data, but left large numbers of inactive compounds. Lacking an effective way to analyse these data and provide the necessary structure-activity relationships for further syntheses, the anticipated increase in new drug approvals did not eventuate. However, the turn of the millennium saw chemists focus on structure-activity relationships by employing computational and artificial intelligence techniques. VSVL is showing an impact in the drive for intelligent selection of the compounds that will form the targets of the vast combinatorial libraries of the future. It remains to be seen whether or not this promising technique to new drugs will be faster and more effective than before.

References

1. For a review of Erlich's work see: Winau, F. *et al. Microbes and Infection,* **2004**, *6*(8), 786-789.

2. For a review of the discovery of penicillin focusing on the contributions by Alexander Fleming see: Bennett, J.W *et al. Adv. Appl. Microbiol*., **2001**, *49*, 163-184.

3. Snyder, S. *et al. Biochem. Pharmacol*. **1964**, *13*, 805-806.

4 For background on the pharmaceutical industry and annual reports see the PhRMA website: *http://www.phrma.org*.

5. For background and further information on combinatorial chemistry see: *The world of combinatorial chemistry: http://combichem.net/home/index.aspx.*

6 For background and further information on the *Lipinski rule of 5* and VSVL see: Lipinksi, C.A. *et al. Adv. Drug Del. Rev.* **1997**, *23*, 3-25.

7. Log P is a measure of the partition coefficient of a compound between octanol and water and may be regarded as a measure of its hydrophobicity/hydrophilicity. Hydrophobicity relates to key drug properties such as absorption, bioavailability, hydrophobic drug-receptor interactions, metabolism, and toxicity. Log P is one of the most critical properties to be optimised to develop a suitable drug candidate.

8. For an excellent review on virtual screening of virtual libraries see: Green, D.V.S. *Prog. Med. Chem*. **2003**, *41*, 61-97.

9. For an excellent review on property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry see Feher, M.; Schmidt, J.M. *J. Chem. Inf. Comp. Sci*. **2003**, *43*, 218-227.

10 For further information on clinical trials see the US and Canadian clinical trials registry: *http://www.clinicaltrials.com*.