Progress in Drug Research, Vol. 62
(Markus Rudin, Ed.)

The drug discovery process

By Paul L. Herrling

Novartis International AG
Corporate Research
CH-4002 Basel, Switzerland
<paul.herrling@novartis.com>

1 Introduction

In recent years drug discovery science has evolved into a distinct branch of science. It is highly multidisciplinary including among others, the disciplines of chemistry, multiple branches of biology (from molecular to behavioral biology), biophysics, computer sciences, mathematics and engineering. It distinguishes itself from academic biomedical sciences by having as its goal and measure of success a pharmacological therapy, while the focus of the academic environment is the generation of new knowledge. Scientists in a drug discovery environment must, therefore, be able to work in multidisciplinary teams, often not of their choosing, and must be able to communicate their specialist knowledge to scientists in other disciplines. They must equally be able to understand the contributions of other specialists towards their common goal. Drug discovery scientists adapt their scientific activities to the requirements of the project to which they contribute, and are often required to abandon one of their own ideas to contribute to somebody else's. This is distinctly different from the academic environment where scientists typically follow their own ideas and their interests, generated usually by the results of their previous research or occasional scientific 'hot topics'.

However, the interaction between academic and drug discovery sciences is essential. The life sciences (including chemistry) are absolutely central to drug discovery because they are needed to improve the knowledge about disease processes to enable progress in pharmacological (and biological) therapies. The life sciences are currently in an exponential phase of knowledge generation, which occurs primarily in the academic environment; therefore, drug discovery scientists need to have very close and frequent interactions with their colleagues in academia.

The tremendous progress in biomedical knowledge and technology in the last 10 years necessitated a complete redesign of the drug discovery process. Some of the key factors mandating change were: (1) an exponential increase in the number of therapeutic targets (a therapeutic target is the precise molecular entity in the human body that interacts with a therapeutic compound to achieve a biological effect in the context of disease), and (2) the discovery of very high levels of complexity in terms of interactions among genes (gene networks) and their products, as exemplified by the combinatorial interaction of proteins in signaling pathways. In 1996, all existing therapeutic

agents interacted with an estimated 500 drug targets [1], but the sequencing of the human genome revealed about 25000–30000 protein-coding genes [2, 3]. If one takes into account splicing and post-translational modifications, it can be estimated that there must be more than 100000 functionally different proteins assuming 25000 protein-coding genes [4], and a conservative average of five splice variants per protein. It is estimated that 57% of the human protein-coding genes display alternative splicing, and that they contain an average of 9 (8.94) exons [5], this would result in about 125000 proteins. This number does not take into account post-translational modifications such as proteolytic processing of larger proteins into smaller active ones or RNA editing [6, 7]. Some estimates indicate that only 5000–10000 of these proteins might be useful drug targets (or 'drugable') [8]. However, this was based on an estimate of 'disease' genes, and there might be many more proteins involved in disease processes than the number of 'disease' genes. Whatever the correct number is, it is orders of magnitudes larger than the past number of available targets, necessitating a high throughput strategy to validate and screen them.

This chapter summarizes some of the key steps in the drug discovery process, and describes some of the main activities at the different stages of the process. It aims at helping to understand the contributions of imaging described in the following chapters in the context of the whole drug discovery process.

2 The drug discovery (and development) phases: overview

The drug discovery community distinguishes four main discovery phases and four clinical phases (Fig. 1)

2.1 The D0 phase

Before the drug discovery process can begin, the strategic selection of therapeutic areas of interest to the company must be made, as no company will address all areas of medicine.

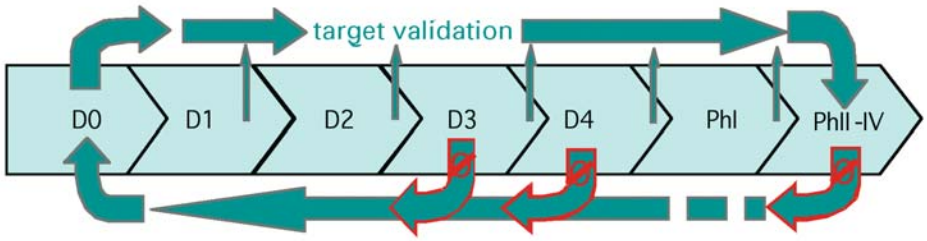

Figure 1.
The phases of drug discovery and development. D0: Basic sciences, target selection. D1: Assay development for high-throughput screening *in vitro*. D2: High-throughput screening of public and proprietary compound libraries, ligand finding (hits). D3: Lead optimization by medicinal chemistry, *in vitro* and *in vivo* models, initial pharmacokinetics and safety. D4: Preparation for human studies: biomarkers, extensive pharmacokinetics, safety, metabolism in animals, formulation, chemical up-scaling. PhI: Proof of concept/mechanism in human, tolerance. PhII: Dose finding. PhIII: Efficacy, registration studies. PhIV: Post-marketing studies.

2.1.1 Choice of therapeutic areas and indication

Discovery research departments need to understand their company priorities, which are usually defined by a group of internal and external discovery scientists, clinical and development scientists, as well as commercial experts from marketing. Key criteria to select the areas for research include:

- expected added medical benefit at the time of introduction in comparison to existing therapy and therapies expected to be in place at that time, i.e., medical need
- existence of a viable scientific hypothesis
- number of patients and expected commercial return
- synergy potential (i.e., will working in this area/indication also contribute to other fields addressed by the company?)
- company skills and history.

2.1.2 Choice of therapeutic target

Once the therapeutic areas are chosen, the drug discovery process begins by selecting the appropriate therapeutic target. Therapeutic targets are the

exact molecular site in the human body at which a proposed therapy is aimed to beneficially modify the course of a disease or even prevent it. They include:

- Cell membrane receptors and ion channels
- Intra- or extracellular enzymes
- Proteins of signaling pathways
- Nuclear receptors
- Genes or gene regulatory processes.

Except for the last target class all others are exclusively proteins. The choice of a particular target depends on the level of scientific knowledge concerning its involvement in the disease process to be addressed. Some targets are clinically validated, i.e., it has been demonstrated in patients that affecting this particular target is of therapeutic benefit. Yet, the most innovative targets have a much lesser degree of validation, such as a genetic linkage with disease, pure speculation based on approximate knowledge about the disease process, or some evidence from gene inactivation experiments. Transgenic animals expressing human disease mutations have become an invaluable tool for intermediary validation [9, 10].

Once a protein has been chosen as a target, it is important to begin efforts to determine its three-dimensional structure so that a structure-based medicinal chemistry effort can be begun as soon as possible and in parallel to high throughput screening.

2.2 The D1 phase

Following target selection, the target protein must be obtained in sufficient quantities and in pure form to allow the design of appropriate high-throughput screening assays [11]. The protein is usually produced by recombinant methods either in bacterial, insect or human cell line systems. It is then included in the appropriate assay for high-throughput screening of large compound libraries to allow measurement of its interaction with a therapeutic tool. At this point, the nature of the therapeutic tool to be developed is selected based on the target characteristics. Therapeutic tools are usually one of the following.

2.2.1 Low molecular weight compounds, synthetics

Synthetic low molecular weight (MW) compounds (usually MW <500) mimic nature's use of small molecules, such as hormones and neurotransmitters, to modulate biological processes. Their main advantages are:
- potential access to all compartments of the human body
- potentially low cost of manufacturing (exceptions are molecules requiring many and complex synthetic steps)
- amenable to a large number of synthetic variations to improve their "drugability", i.e., solubility, membrane permeability, specificity for the target, reduced side effects.

One of the main drawbacks is that, due to their small size, they may have difficulty to interfere with large surface protein-protein interactions.

2.2.2 Low molecular weight molecules, natural products

Such compounds are isolated from natural sources, often as secondary metabolites of organisms that use them in biological functions, e.g., toxins or antibiotics for defensive purposes. Their MW ranges from 100 to about 1000. Their main advantages are:
- they are the result of millions if not billions of years of combinatorial chemistry and selection, so that the probability that they will display biological activity is very high
- they can reach most compartments of the human body
- they can be synthetically modified for drugability.

These are counterbalanced by a principal disadvantage, i.e. natural compounds are, in general, of highly complex structure, including many chiral centers and are difficult to synthesize by methods of synthetic chemistry; often, they can only be obtained by biological processes such as fermentation.

2.2.3 Proteins: antibodies and growth factors

Therapeutic antibodies can be made to interfere with specific molecular

processes and endogenous growth factors to sustain/rectify disease-related deficiencies. Their main advantages are:
- they can be mined from the human genome
- several efficient methods are available to make fully human antibodies
- very high specificity and affinity can be achieved;

on the other hand:
- they can usually only reach extracellular/cell surface targets (in particular antibodies)
- their synthesis, purification and refolding for activity can be expensive and difficult
- in nature, growth factors are very tightly controlled, both spatially and in time and in concentration, so that therapeutic systemic application can cause unwanted side effects.

2.2.4 Gene therapy

A conceptually very elegant method to substitute deficient gene functions found in many diseases such as cystic fibrosis, hemophilia, Gaucher's disease, ADA deficiency etc. [12] is gene therapy, where an engineered vector, often of viral origin, is used to convey an intact functional gene to deficient cells to restore their function. The main advantage of the approach is the direct causal reversal of disease generating malfunction.

However, this has to be balanced against the disadvantages (today) of:
- an often insufficient expression of the repair gene to achieve a therapeutic effect
- insufficient or absent regulation of the foreign gene causing unwanted effects
- imperfect genome integration control (retroviral vectors) can lead to oncogenicity [13]
- insufficient tissue specificity.

2.2.5 Organ transplantation, including xeno-transplantation

The repair of deficient organs can be achieved by transplanting organs from

a compatible donor. Due to the development of immunosuppressive regimes, human to human allo-transplantation today has become a routine procedure [14]. Its main advantage is that it is life saving.

However, today:

- there is an insufficient number of donor compared to medical need
- the immunosuppressive regimes are still imperfect, and have partly life-threatening side-effects.

To address the donor organ shortage, the strategy of transplanting organs from animals that have been genetically modified to inhibit the hyper-acute rejection usually observed in interspecies transplantation has been explored using pigs as donors (xeno-transplantation) [15].

Xeno-transplantation would offer the advantage of:

- 'unlimited' organ supply
- potential replacement of many damaged organs.

At present, the main disadvantage of the approach are:

- rejection mechanisms can not be sufficiently controlled to allow a sufficiently long-lasting donor organ survival in the host
- incompatible physiology between donor and recipient remains a prohibitive problem
- perceived safety concerns about reactivation of endogenous retroviruses [16].

2.2.6 Cell therapies: tissues as well as adult and embryonic stem cells derived

Blood transfusion is the oldest life-saving cell therapy. Newer versions of cell therapy aim at repairing damaged specialized tissue in the host by taking advantage of stem cells occurring in the human body that have the potential to regenerate specific cell types (adult stem cells) or all cell types (embryonic stem cells) [17]. Cell/stem cell therapy offers as main advantages:

- potential repair of many tissues
- no rejection is to be expected if autologous stem cell transplantation is performed.

Main disadvantages (today):
- to date, the ability to multiply of adult stem cells in sufficient quantities without differentiation has had only limited success
- because of their origin there is ethical concern to use human embryonic stem cells [17].

The scientific exploration of the potential of both embryonic and adult stem cells is at its very beginning, in particular the potential to repair complex organs.

2.2.7 Artificial organs

In some cases fully artificial organs can be considered to replace damaged ones [18]. Main advantages of using artificial organs are:
- their potentially unlimited supply
- the possibility to replace tissue/organ function without the drawbacks of biological transplantation and cell therapy methods.

Today, the approach is facing several hurdles:
- only a very limited number of functions can be artificially replaced, such as locomotory functions, cardiovascular pumping and acoustic deficiencies
- the size of artificial organs is technologically highly challenging and often prohibitive.

2.3 The D2 phase: ligand screening

During the D2 phase, the search for ligands for the selected target is performed. Ligands are obtained from a number of sources:
- diverse proprietary libraries, typically around 1 million compounds for a large pharmaceutical company. Compound handling is highly automated to allow for efficient screening operations.
- commercially available compound collections
- tailored combinatorial chemistry libraries [19]
- natural compound libraries from microorganisms or plants that can be preselected based on traditional medicinal knowledge [20]

- proteins and antibodies from genome mining [21].

Ligands that interact with the target are called 'hits' and are usually validated by repeating experiments and recording full dose-response curves. These original hits are then also selected for drugability (solubility, membrane permeability, *in vitro* genotoxicity, selectivity, etc.) before moving into the next phase. The compounds thus selected are then called 'leads', on which the medicinal chemists and pharmacologist perform optimization work.

2.4 The D3 phase: lead optimization

During the D3 phase, the low MW leads obtained in D2 are modified and subjected to structure-activity evaluation to optimize their solubility, potency, selectivity, metabolic properties, as well as their side effect profile both in *in vitro* and in whole animal models. An example of such a lead optimization is shown in Figure 2 and Table 1.

The most promising compounds are then evaluated in at least two relevant animal species to obtain an indication of the species specificity of the target modulation. Longer term studies in intact animals are performed to evaluate the effect of repeated applications, including the occurrence of potential tachyphyllaxis (the attenuation of pharmacological effects after repeated applications). One of the most important aspects of this phase is to obtain data allowing judgment of the potential medical benefit for the patient, as compared to existing therapies or therapies believed to be available at the time of introduction. This is often done in extensive comparative studies with the competing therapeutic agents. The only relevant competitive advantages are advantages that bring a significant medical benefit compared to previous therapy in the patient's perception. A different molecule or mechanism of action as such is not sufficient unless it will plausibly translate into such a medical advantage for the patient.

The patenting of the new therapeutic principle occurs at the latest during the D3 phase. The optimized compound progresses to the next phase, but leads of different chemical structures are kept for potential backups in case the first candidate moving forward fails. However, backups are best selected when the nature of the limiting factors of the first candidate become apparent.

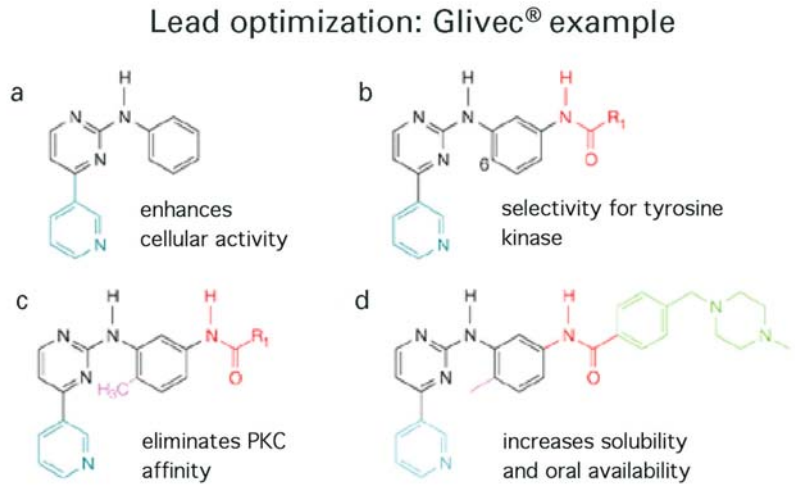

Figure 2.
Lead optimization: the Glivec® example. Glivec® is a new and revolutionary mechanism-directed therapy for chronic amyotrophic leukemia (CML). The addition of the colored substituents to the original lead structure allowed different desired properties of the compound to be improved as indicated. (from [22], reproduced with permission).

2.5 The D4 phase

This phase is the final preparation for the clinical evaluation of a potential drug candidate. It involves extensive pharmacokinetic, metabolic and safety studies in whole animals in at least two species. During D4, chemical up-scaling is carried out, from milligram to kilogram quantities, and an appropriate formulation for compound administration is developed. The clinical research strategy is defined, in recent times with a strong emphasis on biomarkers, that should indicate already during the early clinical testing phase whether the scientific therapy concept is likely to be achieved with the selected therapeutic approach (proof-of-concept studies).

Throughout the process, methods that provide temporo-spatial information on the distribution of potential drug targets, the drug candidate, the drug-target interaction and consequences thereof are of high relevance. Among such techniques, methods such as imaging, which provide non-invasive readouts, are highly attractive as they frequently allow a one-to-one

Table 1.
Selectivity of Glivec® against a panel of kinases. The optimized compound was selected for maximal selectivity towards the Abl kinase (red) but a residual affinity to the platlet-derived growth factor (PDGF, green) and c-kit receptors(blue) remained, which turned out to confer therapeutic benefit for cancers other than CML (from [22]).

Kinases	**IC_{50} [μM]**
v-Abl	0.1–0.3
p210$^{bcr\text{-}abl}$	0.25
p190$^{bcr\text{-}abl}$	0.25
TEL-Abl	0.35
TEL-Arg	0.5
PDGF receptor	0.1
TEL-PDGF receptor	0.15
c-Kit (stem cell factor receptor)	0.1
Flt-3	>10
c-Fms and v-Fms	>10
KDR	>10
EGF receptor	>100
c-erbB2	>100
Insulin receptor	>100
IGF-I receptor	>100
v-Src	>10
Jak-2	>100

translation from preclinical to clinical drug evaluation. The remainder of this book addresses the many parts of the drug discovery process, where imaging techniques can make major contributions.

References

1 Drews J (2000) Drug discovery: a historical perspective. *Science* 287: 1960–1964

2 International Human genome sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860–921

3 Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA et al (2001) The sequence of the human genome. *Science* 291: 1304–1351

4 Orchard S, Hermjakob H, Apweiler R (2005) Annotating the human proteome. *Mol Cell Proteomics* 4: 435–440

5 Modrek B, Lee CJ (2003) Alternative splicing in the human, mouse and rat genomes is

associated with an increased frequency of exon creation and.or loss. *Nat Genet* 34: 177–180

6 Eisenberg E, Nemzer S, Kinar Y, Sorek R, Rechavi G, Levanon EY (2005) Is abundant A-to-I RNA editing primate specific? *Trends Genet* 21: 77–81

7 Sommer B, Kohler M, Seeburg PH (1991) RNA editing in brain controls a determinant of ion flow in glutamate gated channels. *Cell* 67: 11–19

8 Drews J (2000) Drug discovery: a historical perspective. *Science* 287: 1960–1964

9 Stuerchler-Pierrat C, Abramwoski D, Duke M, Wiederhold KH, Mistl C, Rothacher C, Ledermann B, Buerki K, Frey P, Paganetti PA et al (1997) Two amyloid precursor protein transgenic mouse models with Alzheimer disease-like pathology. *Proc Natl Acad Sci USA* 94: 13287–13292

10 Stuerchler-Pierrat C, Sommer B (1999) Transgenic animals in Alzheimer disease research. *Rev Neurosci* 10: 15–24

11 Bleicher KH, Boehm H-J, Mueller K, Alanine AJ (2003) Hit and lead generation: beyond high-throughput screening. *Nat Rev Drug Discov* 2: 369–378

12 Victor A. McKusick et al: OMIM™ Database – Online Mendelian Inheritance in Man™: *http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM* (Accessed May 2005)

13 Fischer A, Hacein-Bey-Abina S, Cavazzana-Calvo M (2004) Gene therapy of children with X-linked severe combined immune deficiency: efficiency and complications. *Medicine/Sciences* 20: 115–117

14 Borel JF (1994) Cyclosporine. In: MM Dale, JC Foreman, T-P Fan (eds): *Textbook of Immunopharmacology*, 3rd edition. Blackwell Scientific Publications, Oxford, Chapter 29

15 Fishman J, Sachs D, Shaikh R (eds) (1998) Xenotransplantation: Scientific Frontiers and Public Policy. Proceedings of a conference. New York, New York, USA. May 18–20, 1998. *Ann NY Acad Sci* 862: 1-251

16 Van der Laan LJW, Lockey C, Griffeth BC, Frasier F, Wilson CA, Onions CW, Hering BJ, Long Z, Otto E, Torbett BE, Salomon DR (2000) Infection by porcine endogenous retrovirus after islet transplantation in SCID mice. *Nature* 407: 90–94

17 Aldhous P (2000) Stem cells: panacea, or Pandora's box? *Nature* 408: 897–898

18 Pescovitz D (2004) Parts for vital organs. *Scientific American*, Special Edition Vol. 14/3: 78–83 (www.sciam.com)

19 Furka A (1995) History of combinatorial chemistry. *Drug Discovery Research* 36: 1–12

20 Yan X, Zhou J, Xie G (1999) *Traditional Chinese Medicines*. MPG Books, Ashgate, Aldershot UK

21 Ostendorn R, Frisch C, Urban M (2004) Generation, engineering and production of human antibodies using HuCAL®. In: G Subramanian (ed): *Antibodies*, Volume 2: Novel technologies and therapeutic use; Kluwer Academic/Plenum Publishers, New York, 13–52

22 Capdeville R, Buchdunger E, Zimmermann J, Matter A (2002) Glivec: CGP057148 (STI571, imatinib), a rationally developed, targeted anticancer drug. *Nat Rev Drug Discov* 1: 493–502