Drug Discovery and Development: An Overview of Modern Methods and Principles

Over the course of the last two centuries, modern medicines have improved the lives of countless patients. Diseases and conditions that were once deemed incurable or fatal have been conquered with therapeutic agents designed to extend and improve quality of life. The most recent, and perhaps most notable, of these accomplishments is the transition seen in the consequences of infection with human immunodeficiency virus (HIV), the virus known to cause acquired immune deficiency syndrome (AIDS).\(^1\) When the virus was first identified by two research groups in 1983,\(^2\) there were few antiviral agents available, none provided effective treatment for HIV infection and infection progressed rapidly to AIDS, and finally death by opportunistic infection. By 1987, AZT\(^\text{®}\) (Retrovir\(^\text{®}\), azidothymidine, Figure 1.1), the first reverse transcriptase inhibitor, was approved for clinical application for the treatment of HIV infection,\(^3\) and additional treatment options were developed through the next three decades. The discovery of modern antiviral drugs such as Viread\(^\text{®}\) (Tenofovir),\(^4\) Zeffix\(^\text{®}\) (Lamivudine)\(^5\) (reverse transcriptase inhibitors), Viracept\(^\text{®}\) (Nelfinavir),\(^6\) Norvir\(^\text{®}\) (Ritonavir),\(^7\) and Crixivan\(^\text{®}\) (Indinavir)\(^8\) (HIV protease inhibitors) provided additional treatment alternatives (Figure 1.1). In the late 1990s, multidrug cocktail treatment regimens, also known as highly active antiretroviral therapy (HAART),\(^9\) were introduced, further enhancing treatment options, culminating in the development of all-in-one fixed combination medications such as Complera\(^\text{®}\)\(^10\) and Stribild\(^\text{®}\).\(^11\) While additional progress is still required, it is clear that modern drug discovery and development changed the course of the AIDS epidemic in less than three decades, allowing patients who were once given a death sentence to lead productive lives.\(^12\)

Cancer treatment has seen a similar transition, as survival rates for many types of cancer have dramatically improved as a result of the discovery and
development of novel therapeutic agents. In the United States, overall cancer death rates have declined 11.4% between 1950 and 2009, and progress against some specific cancer types has been substantial. Breast cancer, prostate cancer, and melanoma, for example, have seen significant increases in their 5-year survival rates over the same period. The 5-year survival rate for breast cancer increased from 60% to 91%, while the survival rate for prostate cancer increased from 43% to over 99%, and melanoma survival rose from 49% to 93%. A significant portion of the improved clinical outcomes in cancer can be attributed to improved chemotherapeutic agents. The identification of antitumor natural products and natural product analogs such as Taxol® (Paclitaxel), Velban® (Vinblastine), Adriamycin® (Doxorubicin), and Hycamtin® (Topotecan) has clearly demonstrated the importance of natural products in modern medicine, while the development of small molecule kinase inhibitors such as Gleevac® (Imatinib), Tasigna® (Nilotinib), and Tarceva® (Erlotinib) provide clear evidence of the power of modern drug discovery techniques (Figure 1.2).

The treatment of cardiovascular disease has also seen dramatic improvements in the wake of the discovery of a multitude of therapeutic agents designed to mitigate symptoms or prevent the underlying causes of the disease. A myriad of treatments have been developed to address hypertension, also known as “the silent killer” because of its asymptomatic nature, leading to improvements in both the quality of life and life expectancy of patients. Diuretics (e.g., Midamor® (Amiloride), Lozol® (Indapamide), β-blockers (e.g., Tenoretic® (Atenolol), Inderal® (Propranolol), and
The natural products Taxol® (Paclitaxel), Velban® (Vinblastine), Adriamycin® (Doxorubicin), and Hycamtin® (Topotecan) are exemplary natural products that have been developed for the treatment of cancer, while Gleevec® (Imatinib), Tasigna® (Nilotinib), and Tarceva® (Erlotinib) were developed for the treatment of cancer through the application of modern drug discovery programs.
angiotensin-converting enzyme (ACE) inhibitors (e.g., Capoten® (Captopril),25 Vasotec® (Enalapril)26) are just a few of the types of treatments currently available to lower blood pressure and keep cardiovascular disease at bay. Revolutionary changes occurred in the prevention of cardiovascular disease with the introduction of HMG-CoA reductase inhibitors, also known as statins.27 Lipitor® (Atorvastatin),28 Zocor® (Simvastatin),29 and a number of related compounds have demonstrated remarkable capacity to lower cholesterol levels, a major risk factor associated with cardiovascular disease (Figure 1.3).30

FIGURE 1.3 The diuretics Midamor® (Amiloride) and Lozol® (Indapamide), the β-blockers Tenoretic® (Atenolol) and Inderal® (Propranolol), the ACE inhibitors Capoten® (Captopril), Vasotec® (Enalapril), and the HMG-CoA reductase inhibitors Lipitor® (Atorvastatin) and Zocor®, (Simvastatin) have significantly improved the treatment of cardiovascular disease.

Similar improvements in disease management, symptomatic relief, and improvements in the quality of life through the development of novel chemotherapeutics could be described across a wide range of health issues. It is clear that the treatment of infectious disease, pain management, respiratory disease, and many other conditions has been profoundly and positively impacted by the identification of novel therapies.31 There are, however, many challenging areas of health care that remain in need of improved medicine and advances in current therapy. Alzheimer’s disease, for example, is the most common form of dementia, and was originally described by German psychiatrist and neuropathologist Alois Alzheimer in 1906. Over 100 years later, treatment options for this disease remain limited at best, despite the enormous amount of effort and research funding dedicated to identifying novel treatments. Potential drug targets such as β-secretase (BACE), γ-secretase, glycogen synthase kinase 3β (GSK3β), and cyclin-dependent kinase-5 (CDK5)32 have been extensively studied, but clinically effective, disease modifying agents have as yet to be identified.
Additional challenges also exist in areas once thought to have been conquered by modern science. In the 1960s and 1970s, for example, it was widely believed that modern medicine had all but conquered infectious disease and that the major classes of antibacterial agents, β-lactams, quinolone, tetracyclines, and macrolide antibiotics (Figure 1.4) would provide all of the tools necessary to protect humanity. The rise of methicillin-resistant *Staphylococcus aureus* (MRSA) in the 1980s and 1990s, however, has made it clear that additional tools will be required in order to maintain the upper hand in the war against bacterial infection. Methicillin (Staphcillin®) was introduced in 1959 as a means of treating penicillin-resistant infections, but less than two years later, resistant strains were identified in European hospitals. By the 1980s, MRSA had spread throughout the globe, and as of 2009, MRSA infections cost the US health system $3 billion to $4 billion annually.33

There is no doubt that the discovery of new therapeutic agents has a positive impact on society, but to the casual observer, it is not clear how this goal is achieved. On the surface, providing a drug necessary to solve a medical problem would seem to be a relatively simple task; identify the cause of the disease or malady and design a drug that will fix or eliminate the problem. In the case of infectious disease, eliminate the infectious agent, whether bacterial or viral, and the health problem is solved. This is, of course, a very simplistic view, as there are many factors to consider beyond killing the offending organism. There are millions of compounds that will kill an infectious organism, but how many of these compounds can do so without negatively impacting the host? How many of the remaining compounds

![FIGURE 1.4](image-url)
can be delivered as safe and effective therapeutic agents? How does one determine which of the nearly infinite possibilities are useful and which ones are not? Of the useful compounds, which ones will be of interest to the companies that manufacture drugs and which ones will not? These issues are exceptionally complex, and become even more so, when the health issue is something other than an invading organism. In considering chronic pain management, for example, a drug provided to a patient should alleviate the chronic pain without interfering with the pain associated with protective instincts, such as withdrawing one’s hand from a hot stove. This added complexity is a common feature of the vast majority of disease states and must be addressed in order to successfully develop any new medication.

Given the large number of complex issues associated with drug design and development, it should be abundantly clear that no one individual could possibly conquer all of the tasks required to discover, develop, and successfully bring to market a new therapeutic entity. The process is a multidimensional one and as such requires the coordinated effort of individuals with a wide array of expertise such as medicinal chemistry, in vitro biology, drug metabolism, animal pharmacology, formulations science, process chemistry, clinical research, intellectual property, and many other fields. Enabling technologies, such as high throughput screening, molecular modeling, pharmaceutical profiling, and biomarker studies, also play key roles in modern drug research. It is critical that anyone interested in pursuing a career in the development of pharmaceutically useful agents, whether in an industry setting or an academic institution, must be willing and able to participate in collaborative research effort over a significant period of time. In addition, it is important that any participant in this field understands the magnitude of the costs associated with the pursuit of new drugs. The rewards for those who are successful can be substantial, as indicated by the success of compounds such as Lipitor® (Atorvastatin), which had peak annual sales of over $13 billion, Prozac® (Fluoxetine, peak sales $2.8 billion), and Singulair® (Monteluclast, 2011 sales $5.5 billion), but the cost in time and resources is substantial (Figure 1.5). As indicated in Figure 1.6, it has been estimated that the identification of a single marketed drug can require an initial examination of over 100,000 candidate compounds, hundreds of preclinical animal studies, and numerous clinical trials involving thousands of patients. A recent analysis of clinical trial success rates has indicated that only 1 out of every 10 clinical candidates will successfully traverse clinical trials and reach the market. This represents a success rate of less than 0.001% if measured by the number of compounds examined at the outset of the process. If measured according to the number of programs required to advance a single drug to market, program attrition rates indicate that only 1 in 24 programs is successful.

The cost associated with the identification of useful and marketable therapeutic entities is also staggering. As of 2011, it is estimated that a
A single new drug costs over $1.75 billion to discover and develop. As a measure of comparison, the same amount of money could be used to buy 17 Boeing 737 jet aircrafts (based on 2012 prices on Boeing’s Web site), purchase approximately 7000 homes (assuming $250,000 per home), 70,000...
automobiles (average price $25,000), or provide for the raising 7000 children born in 2010 to the age of 18. The costs and complexity of drug discovery and development is staggering.

**DRUG DISCOVERY AND DEVELOPMENT FROM 20,000 FEET (FIGURE 1.7)**

![Diagram showing the drug discovery and development process](image)

**FIGURE 1.7** The drug discovery and development process viewed from “20,000 feet.”

Fortunately, like most complex processes, drug discovery and development can be broken down into many smaller tasks and functions. At the highest level, the process can be divided into two major stages. The first, referred to as drug discovery, includes all of the experimentation and studies designed to move a program from the initial identification of a biological target and associated disease state to the identification of single compound with the potential to be clinically relevant. The drug discovery stage may be further broken down into three distinct phases: target discovery, lead discovery, and lead optimization. Each phase of drug discovery is designed to establish a scientific link between a biological target (e.g., an enzyme, G-protein-coupled receptor, ion channel, etc.) and a disease state model designed to mimic the human disease state. This process, often referred to as target progression and target validation, is accomplished through the use of molecular probes designed to identify multiple series of compounds that will modulate the activity of the biological target of interest. In many cases, known compounds are employed to facilitate target selection, and are eventually transitioned into novel compounds through the processes of lead discovery and lead optimization. In lead discovery phase, sets of structurally related compounds with the desired biological activity are identified (lead discovery) through biological screening of large numbers of compounds. Once a candidate series has been identified, the lead optimization phase begins. In this phase, structural analogs within a lead series are studied to identify a single compound that...
may be progressed into the drug development stage. Typically, multiple lead series are identified in both the lead discovery and lead optimization phases through iterative rounds of experimentation. In many cases, the lead discovery and lead optimization phases overlap, as a typical drug discovery program will produce multiple sets of related compounds with the potential for identification of candidates that might progress into drug development. This approach is required for success, as it is often difficult, if not impossible, to identify the lead series that will contain the final lead candidate in the early phases of the drug discovery process. Parallel operations of this type mitigate the risk of failure of any one series of compounds. The lead discovery phase typically concludes with the successful demonstration of *in vivo* efficacy in an appropriate animal model employing a compound that possess physical and chemical properties consistent with eventual clinical study in the drug development stage.

The second major stage, drug development, typically begins once a single compound has been identified, which is then progressed through various studies designed to support its approval for sale by the appropriate regulatory bodies. The first step in this process is the submission of an Investigational New Drug (IND) Application that requests permission to move a clinical candidate into human study. This document provides regulatory agencies with detailed preclinical data describing animal pharmacology and toxicology studies, chemical manufacturing information (including formulation, stability studies, and quality control measures), and, of course, detailed clinical protocols that describe how the clinical compounds will be studied in human populations if the studies are approved.

While clinical trial designs can vary substantially from one candidate to another, the general goals of phase I, II, III, and IV are the same. Chapter 9 will provide a more detailed review of clinical trials, but the basic tenants of clinical trials are as follows. In phase I clinical trials, safety and tolerability of an investigational new drug is examined in a small number of healthy individuals, typically 20 to 100 people, with the goal of determining if safety margins are suitable for further progression in the clinical trial process. Pharmacokinetic and pharmacodynamic aspects of the candidate are closely monitored, and the drug candidate is typically administered first in a single ascending dose (SAD) study, followed by a multiple ascending dose (MAD) study. In the SAD study, the drug is given to a group of subjects once and they are monitored to determine the impact of the drug. If there are no adverse effects, then a second group is treated with a single higher dose of the drug candidate and monitored as before. The cycle is repeated until intolerable side effects appear in order to determine the maximum tolerated dose (MTD). MAD studies are similar, but each group of subjects is provided with multiple low doses of a candidate compound over a set time. As in the SAD studies, the manifestation of clinically intolerable side effects defines the MTD for the MAD studies. The
data developed through the course of the phase I studies are used to determine the doses that will be used in phase II and phase III clinical trials.

Phase II typically involves 100 to 300 patients and are designed to determine whether or not the clinical candidate provides the desired biological impact. Safety studies also continue through phase II trials. In the first part of phase II trials, referred to as phase IIA, the goal is to determine the dose required to provide the desired therapeutic impact or endpoint for the clinical candidate. Once the proper dose levels are determined, phase IIB studies can be initiated. The goal of phase IIB studies is to determine the overall efficacy of candidate compounds in a limited population of subjects. The majority of clinical drug candidates fail in phase II studies due to safety issues or lack of efficacy. As of 2011, only 34% of phase II clinical candidates successfully reach phase III studies.

The effectiveness of new drug candidates in larger patient population are determined in phase III clinical trial. These studies are typically randomized and involve hundreds to thousands of patients at multiple clinical trial sites and are designed to determine the efficacy of the candidate compound relative to the current standard of care. The cost and time associated with this phase of clinical study can vary dramatically depending on the clinical endpoint under investigation. Clinical trials for new, acute treatments, such as novel antibacterial agents, are shorter and involve far fewer patients than clinical trials for chronic conditions such as osteoarthritis. Patients are also closely monitored for adverse side effects, as the larger patient pools can identify safety issues that did not become apparent in smaller phase II trials. The number of subjects, time requirements, and complex design of phase III clinical trials (especially in chronic medical conditions) dictate that they are the most expensive aspect of drug discovery and development. Upon completion of phase III trials, a New Drug Application is submitted to the appropriate regulatory body. This document typically contains comprehensive details of both animal and human studies, all safety findings (adverse and side effects), manufacturing procedures (including methods of analysis to ensure drug quality), detailed formulation information for all dosing methods studied, and storage conditions. Regulatory reviews can lead to requests for additional information regarding the submission, or even additional clinical trials to further establish either safety or efficacy. Ideally, these reviews lead to regulatory approval, including labeling requirements, and approval to market the new drug.

Approval of regulatory bodies does not, however, signal the end of clinical trials. In many cases, regulatory agencies will require additional follow-up studies, often referred to as phase IV trials or postmarketing surveillance. In general, these studies are designed to detect rare adverse effects across a much larger population of patients than could be supported in phase III trials or long-term adverse effects that might be outside of the scope of phase
III trial durations. The impact of phase IV studies can include alterations to labeling based on safety results, contraindication for use of the new drug in combination with other medications, or even the withdrawal of marketing approval if the findings are severe enough. The COX-2 selective non-steroidal anti-inflammatory agent Vioxx® (Rofecoxib), for example, was removed from the market after phase IV studies indicated that it increased the risk of ischemic events in patients (Figure 1.8). In a similar fashion, Baycol® (cerivastatin), a 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitor marketed by Bayer AG for the treatment of high cholesterol and cardiovascular disease, was voluntarily removed from the market after reports of fatal rhabdomyolysis (Figure 1.8).

It is should be noted that safety studies are not the only reason for phase IV clinical trials. Companies often use the data provided in postmarketing surveillance and additional clinical studies to identify competitive advantages, new markets, and new indication for their products. There is some level of risk associated with conducting trials designed to identify clinical superiority, as the results of competitive trials are often difficult to predict. In some cases, a company’s plan to demonstrate that their compound is superior to a competitor’s drugs backfires, and they prove the opposite.

TARGET SELECTION: THE FIRST STEP FORWARD

The process of identifying a new drug candidate begins with identifying a disease state or condition that can be addressed or modified through the application of a suitable chemotherapeutic intervention. In theory, the most pressing medical needs would have the highest priority in order to ensure improvement of the overall quality of life for patients. In practice, however, there are many factors that contribute to the decision of which disease or condition to attempt to address through drug discovery programs. First, the pathway to develop a therapeutic intervention may not be clear for a particular disease or condition even though the medical
need is urgent. For example, while it is clear that Alzheimer’s disease is pressing medical need,\textsuperscript{41} to date there are no disease modifying therapies available, despite the extraordinary amount of capital expended in an effort to identify useful therapies. This is in part due to the lack of proven targets for Alzheimer’s disease. Similarly, while there is a clear and pressing need for additional treatments for schizophrenia,\textsuperscript{42} the current level of understanding of the disease state and lack of sufficient animal models\textsuperscript{43} is a hindrance to progress in this important area.

There are also drug targets that have a theoretical connection to a particular disease state, but have as yet to be proven relevant to the human condition through the application of an appropriate chemotherapeutic agent. Cholesteryl ester transfer protein (CETP), for example, plays a key role in the interconversion of high density lipoproteins (HDL) and low density lipoproteins (LDL), and it has been suggested that inhibitors of this enzyme would have a positive impact on patients suffering from hypercholesterolemia.\textsuperscript{44} While potent CETP inhibitors have been identified, such as Torcetrapib (CP-529,414)\textsuperscript{45} and Dalcetrapib (JTT-705)\textsuperscript{46} (Figure 1.9), none have been approved for marketing as the clinical candidates examined to date failed to demonstrate statistically significant beneficial effects in patients. It is possible that these results are an indication that CETP inhibition is not a viable drug target for the treatment of cardiovascular disease. It is also possible, however, that the clinical candidates examined to date are flawed in ways unrelated to the CETP that prevented them from functioning in the desired manner (e.g., off-target effects, pharmacokinetic issues).

In the case of Torcetrapib (CP-529,414), clinical trial data demonstrated that the drug candidate increased HDL levels and decreased LDL levels, and did not negatively impact mortality rates.

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure1.9}
\caption{Torcetrapib (CP-529,414), Dalcetrapib (JTT-705), and Anacetrapib (MK-0859) are cholesteryl ester transfer protein (CETP) inhibitors that have been clinically studied as potential treatments for hypercholesterolemia. Torcetrapib increased HDL levels and decreased LDL levels, but increased mortality rates, while Dalcetrapib was not efficacious in clinical trials. Anacetrapib increased HDL levels and decreased LDL levels, and did not negatively impact mortality rates.}
\end{figure}
indicating that clinical efficacy could be achieved.\textsuperscript{47} Unfortunately, the clinical candidate also caused increase in blood pressure and mortality rates, leading to the termination of clinical development in 2006.\textsuperscript{48} Dalcetrapib (JTT-705) clinical studies were terminated by Roche in 2012 due to lack of efficacy.\textsuperscript{49} On the other hand, clinical trials with Anacetrapib (MK-0859, Figure 1.9),\textsuperscript{50} which also targets CETP, have successfully demonstrated that this compound can increase HDL levels and decrease LDL levels without increase in blood pressure or increased risk of cardiovascular disease-related deaths or events.\textsuperscript{51} As of the writing of this text, the value of CETP as a drug target remains an unanswered question.

A similar scenario has surrounded $\gamma$-secretase inhibition. While it is known that $\gamma$-secretase plays a key role in the formation and deposition of amyloid plaques during the progression of Alzheimer’s,\textsuperscript{52} inhibitors of this key enzyme have failed to provide the clinically beneficial results expected. Semagacestat (LY450139, Figure 1.10), a compound developed by Eli Lilly inhibits $\gamma$-secretase, shows a dose-dependent lowering of amyloid plaque formation in humans, but did not improve cognitive function in patients. In fact, Semagacestat produced statistically significant declines in cognitive function compared to the placebo group in clinical trials.\textsuperscript{53} Once again, this raises the question as to whether the target pathway is a dead end for treatment of the disease in question or if the compound in question is flawed in some unforeseen manner. In the case of Semagacestat (LY450139), it is possible that unexpected off-target activity may be clouding the clinical results. Semagacestat (LY450139) also interferes with Notch signaling, which plays a key role in cognitive function,\textsuperscript{54} and it is not unreasonable to suggest that Notch signaling modulation masked potential positive effects that might have been observed if this off-target activity was absent.

The unanswered question raised by the failures of clinical candidates such as Torcetrapib (CP-529,414) and Semagacestat (LY450139) highlights the risks associated with choosing a target that is not clinically proven, as well as the potential for negative clinical results to be clouded by factors not related to the targeted mechanism. There are, however, substantial financial incentives to attempt to develop a “first-in-class” therapeutic agent. Prior to the introduction of the statins (or HMG-CoA reductase inhibitors),\textsuperscript{55} there

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{Figure110}
\caption{The $\gamma$-secretase inhibitor Semagacestat (LY450139) failed to improve cognitive function in Alzheimer’s patients, despite the fact that it lowered amyloid plaque formation.}
\end{figure}
was no clear pathway forward to inhibit cholesterol synthesis. The companies that took the leap of faith that inhibition of HMG-CoA reductase would provide therapeutic relief for the prevention of cardiovascular disease received significant financial rewards in the marketplace as indicated by the success of drugs like Mevacor® (Lovastatin)\textsuperscript{56} and Lipitor® (Atorvastatin)\textsuperscript{57}.

There are, of course, many biological targets with proven clinical utility that an organization might choose to focus on, such as phosphodiesterase-5 (PDE-5),\textsuperscript{58} β-adrenergic receptors,\textsuperscript{59} and 5-hydroxytryptamine (5-HT) receptors.\textsuperscript{60} In considering whether or not to pursue known drug targets, one must keep in mind both the benefits and potential pitfalls related to previously defined targets. On the positive sides, a wealth of research and development information will be available in the literature, as companies and research institutions (universities, non-profit research institutions, etc.) patent and publish their research in order to garner support for their marketed products and research programs. The availability of research tools such as biological assays, reference compounds, and clinical trial data can provide an excellent springboard for a drug discovery program. On the other hand, the availability of this kind of information presents a significant hurdle to the development of new therapeutic agents, as any new compound or biological agent will be required to demonstrate clinical superiority to the current standard of care. Also, scientific disclosures in the literature will be available as prior art and could prevent an organization from gaining patent protection for their research (this topic will be covered in more detail in Chapter 12). If, however, one is successful in developing a new therapeutic entity based on clinically proven targets, substantial benefits can be available. Sepracor, now a division of Sunovia, for example, took the risk of developing a new antihistamine at a time when the market was dominated by Seldane® (Terfenadine).\textsuperscript{61} They were able to demonstrate that Allegra® (Fexofenadine), a metabolite of Seldane® (Terfenadine), is significantly safer than its predecessor and quickly took over the antihistamine market (Figure 1.11).\textsuperscript{62}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{seldane_and_allegra}
\caption{Seldane® (Terfenadine), the first non-sedating antihistamine, dominated the market until serious safety issues were identified. It was replaced by a Allegra® (Fexofenadine), an active metabolite that is safer than the original.}
\end{figure}
Financial considerations also play a major factor in the determination of which diseases and potential drug targets are examined and which are not. Clearly, the amount of money and time available to pursue new therapeutic entities is limited, so not every target or disease state can be addressed. In the corporate world, disease state and target selection is generally driven by the ability to generate profitable products whose sale will support future research programs. On the surface, this would seem to dictate that only diseases or conditions with large numbers of patients would be of interest to corporate entities, but this is not the case. Chronic diseases such as osteoporosis, hypertension, hypercholesteremia, and arthritis clearly have a large patient population that creates opportunities for corporations. Rare diseases, however, also present significant opportunities and a pathway for growth. Amyotrophic lateral sclerosis (ALS), for example, is a disease with a small, but consistent patient population with significant unmet medical needs. At any given time, there are only 20,000–30,000 ALS patients in the United States whose life expectancy is only 3–5 years, and there are no life extending therapies currently available. This would appear to be a very small market that is unlikely to provide the kind of profitability required to sustain a corporation. However, it is important to realize that if a suitable treatment were available, this terminal condition would become a chronic condition wherein patients would be treated for the disease throughout the course of an otherwise normal life span. In addition, increased survival time for ALS patients would lead to a larger patient pool, providing additional revenue to a company that develops a life-extending treatment for ALS.

The selection of targets and disease states of interest sets the course for all future aspect of a research program, so the importance of this decision cannot be understated. Once the biological target is selected, the process of identifying a clinical candidate can begin.

**HIT IDENTIFICATION: FINDING A STARTING POINT**

Once a target of interest has been identified, the remainder of the research program is essentially a quest to identify a single compound that is suitable for use in a clinical setting. Of course, this relatively simple statement is actually a representation of an exceptionally complex and multifaceted problem. Currently, there are over 70 million compounds registered in the Chemical Abstract Service database, and the total number of possible compounds to consider as drug candidates is nearly infinite, so the question of where to start the process is significant. Fortunately, there are some guidelines that have been developed in order to provide some guidance as to where one might begin to look for biologically useful molecules. Lipinski’s rule of 5, for example, suggests that the majority of druglike
compounds exist within a limited portion of chemical space. This concept will be discussed in greater detail in Chapter 5, but for the purposes of this discussion, Lipinski’s rules suggest that “druglike” compounds will have (1) a molecular weight lower than 500, (2) a logP below 5, (3) less than 5 hydrogen bond donors, (4) less than 10 hydrogen bond acceptors, and (5) less than 10 rotatable bonds. While there are exceptions to these rules (most notably in the natural products arena), their application to chemical space can be useful in that it provides a framework for further movement towards a manageable number of compounds for consideration.

However, these limitations still leave an enormous expanse of chemical space that could be mined in an effort to identify compounds that interact with a biological target of interest. This issue is further complicated by the fact that drugs interacting at the same target may have very little structural overlap. There are, for example, clear similarities between the HMG-CoA reductase inhibitors Lipitor® (Atorvastatin), Lescol® (Fluvastatin), and Crestor® (Rosuvastatin) (Figure 1.12). They each contain a para-fluorobenzene ring and 1,3-diol-carboxylic acid side chain displayed in a similar orientation, but the remainder of the three compounds are substantially different from each other. Mevacor® (Lovastatin, Figure 1.12), which also inhibits HMG-CoA reductase, is from an entirely separate structural class, and it is not clear to the naked eye how this compound is related to the previously mentioned drugs. Similarly, Viagra® (Sildenafil) and Cialis®

![Image of chemical structures](image-url)
(Tadalafil)

are both PDE-5 inhibitors, but structurally, they are quite different (Figure 1.13). It is not immediately clear how these two compounds

are related or why they would serve the same biological function. The same is true of morphine,

Demerol® (Meperidine) and Duragesic® (Fentanyl) (Figure 1.14). While all of these compounds are μ-opioid receptor agonists,

it would not be obvious to a casual observer that they share a common biological target. The selective serotonin reuptake inhibitors (SSRIs) Zoloft® (Sertraline), Zelmid® (Zimeldine), Celexa® (Citalopram), Prozac® (Fluoxetine), and Paxil® (Paroxetine) represent distinct chemical classes that are not clearly related to each other (Figure 1.15). Given the breadth of

Zoloft® (Sertraline), Zelmid® (Zimeldine), Celexa® (Citalopram), Prozac® (Fluoxetine), and Paxil® (Paroxetine) are all selective serotonin reuptake inhibitors (SSRIs) useful for the treatment of depression, but structural similarities are limited.
structural diversity that can be employed for any single molecular target, it is clear that even the process of finding an initial chemical lead for a drug discovery program can be challenging.

Fortunately, a number of tools and methods have been developed to address the simple and yet very complex question of identifying a molecular starting point for a drug discovery program. Essentially, there are two general methods utilized in modern drug discovery programs, physical high throughput screening (HTS) methods and virtual high throughput screening methods. There is some degree of overlap between the two categories, and the use of one set of tools does not preclude the use of the other. In point of fact, they are often employed in tandem in order to improve the likelihood of success. Physical high throughput screening approaches depend on the ability to screen large compound libraries containing hundreds, thousands, or even millions of samples. Large libraries are often designed to be chemically diverse in order to cover as much of the “drug-like” chemical space as possible, but focused libraries designed to target specific types of biological targets (e.g., kinases, phosphatases) have also been employed. Physical samples for screening are available from commercial sources (e.g., Maybridge, Enamine, Aldrich, etc.), and pharmaceutical companies generally maintain an internal compound collection of proprietary compounds.

Physical HTS techniques also require sophisticated, fully automated systems capable of manipulating reagents and 96-, 384-, or even 1496-well microtiter plates, as well as reagent distribution, data acquisition, and waste disposal for thousands of samples per hour. Automated data analysis is also required in order to handle the volume of information generated in a typical high throughput screening run.

There are some key points that one must consider in evaluating the data provided by an HTS screen. First and foremost is the possibility of false positives and false negative results. In physical screening methods, the sheer number of manipulations involved leaves open the possibility that an error may occur during some facet of reagent handling (such as a clogged pipette tip). There is also the possibility that the screening sample may have degraded over time, creating “ghost samples” within the chemical library (in other words, a sample whose structure no longer matches the material originally entered into the library). In order to ensure that programs are directed towards authentically active compounds, the chemical integrity of “hit” samples is generally assessed using High-performance liquid chromatography/Mass spectroscopy (HPLC/MS) methods. In addition, biological screening is often repeated with the “hit” compounds in order to validate the HTS results.

As an alternative to physical HTS methods, it is also possible to perform virtual high throughput screening (also referred to as in silico screening). In this scenario, advanced molecular modeling techniques are combined with virtual chemical libraries (data files containing detailed structural information on millions of compounds) and structural data on the biological target...
in order to assess a compound’s ability to interact with the target of interest. Virtual chemical libraries are often freely available from commercial vendors (the largest of which is the ZINC database; http://zinc.docking.org/) and, as with physical samples, pharmaceutical companies generally maintain virtual libraries of their proprietary compounds for internal use. Structural information on biological targets may be available through X-ray crystallography, as a large number of protein crystal structures are available through the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (http://www.rcsb.org/pdb/home/home.do). If a structure is not readily available, it may be possible to create a homology model of the biological target using crystal structure data of a closely related macromolecules.81 In either case, the individual compounds of the chemical libraries can then be “docked” in a hypothetical binding site in the target of interest to determine a relative rank order for the entire set of compounds. Automated data analysis tools are then employed that organize the predictions provided by the “docking” of the chemical libraries to the hypothetical binding sites of the biological targets. The predictions can then be used to select a smaller subset of a large library for physical biological screening as potential starting points.

Much like physical HTS, there are some important limitations that must be considered in evaluating virtual screening data. First and foremost, virtual screening results are predictions based on model system and not actual data on physical compounds. As such, the quality of the results will depend on the quality of the model. In silico models based on X-ray crystal structures tend to be stronger models than homology models built on related biological structures, but it is important to realize that there are limitations associated with X-ray crystal-based models as well. Crystal structures can provide exceptionally detailed structural information, with resolution as low as 1.5 Å, but by definition, X-ray crystal structures are solid state version of the desired target. It is possible that the structure provided by X-ray crystallography matches the biologically active form of the biological target of interest, but it may not. In “real-life” situations, biological targets are either dissolved in water or membrane bound, and it is possible that they may have a different configuration in these situations as compared to the close-packed structure of a crystal form. Also, in many cases, sections or a macromolecule must be altered or removed in order to generate a crystallizable form of the biological target (with or without a ligand). Given these limitations, virtual “hits” should also be physically validated in biological screening efforts in order to confirm that the predictions provided by molecular modeling are representative of the real system.

Irrespective of the initial screening method employed (physical or virtual), a successful screening effort will produce a subset of potential “hit” compounds that will need to be examined in order to determine whether or not follow-up efforts are warranted. This determination, also referred to as “lead discovery,” (Figure 1.7) can be quite complex in itself depending on
the number and nature of the “hits” identified. If 500,000 compounds are screened and only 0.1% of the compounds in the library provide interesting biological results, this still leaves 500 compounds to be evaluated. Ideally, the initial “hits” will belong to a relatively small number of structural classes, and then each structural class can be independently analyzed to determine if further effort in the class is warranted. Small groups of related compounds demonstrating the desired biological activity can provide a significant advantage in further efforts, as structure–activity relationship data may become apparent at an early stage. (The concept of structure–activity relationships will be covered in more detail in Chapter 5.) Also, the preparation of additional analogs may be simplified, as synthetic methods may already be available. On the other hand, the presence of set of related compounds within a library suggests that they may have been prepared for a project with a different biological target. Intellectual property issues may also exist, as patent rights and ownership could become a serious question, especially if the compounds were part of a set that has been previously patented, previously published, or purchased from a commercial vendor. Intellectual property consideration will be explored in more detail in Chapter 12.

In some instances, “hit” compounds may be singletons. Isolated compounds can be more difficult to follow up on, as the original HTS data set will not provide any additional guidance on how to proceed. It is, however, still possible to generate more data on related compounds that may be available from outside of the original compound library through either commercial sources or additional synthetic efforts.

Once the initial “hit” compounds have been identified, confirmed, and a compound class (or perhaps more than one compound class, depending on the available resources) has been selected for further study, an iterative process of compound acquisition/synthesis, biological screening, and data evaluation begins with the goal of improving the potency of the compounds (Figure 1.16). In each cycle of the “lead optimization” process,

![Figure 1.16](https://example.com/fig1_16.png)

**FIGURE 1.16** The lead optimization cycle begins with the identification of a lead structure (“hit”) in a relevant biological assay. New analogs with structural modifications are prepared and screened in the biological assay. If the assay results improve, then the changes are kept and the cycle is repeated. If the changes are detrimental, then the changes are discarded and the cycle is repeated. This process continues until a candidate compound with the desired properties is identified.
new data are produced as changes in the molecular structure are made to the “lead” compounds, and these data are used to design the next generation of compounds. This cycle of generating structure–activity relationship data continues until a compound suitable for clinical evaluation has been identified. The nature of this process and the associated medicinal chemistry will be discussed in greater detail in Chapter 5.

IDENTIFY A CLINICAL CANDIDATE: JUGGLING THE PROPERTIES

Simply identifying a compound that is potent at the target of interest is a difficult task to begin with, but sheer potency is not enough to allow a compound to be considered for clinical development and ultimately commercialization. The process of discovering a suitable drug candidate is, in many ways, a juggling act performed by drug discovery scientists (Figure 1.17). As programs progress from hit and lead identification to lead optimization and an eventual clinical study, hundreds if not thousands of compounds will be examined. It is the drug discovery scientist’s responsibility to identify a compound that will not only modulate the target of interest but also possesses the correct balance of properties required to create a usable drug. Potency at a biological target is only the beginning of a long series of screening processes that must be performed in order to demonstrate that a compound will survive the rigors of a discovery program. The specific strategies employed are different for each program, but in general they can be mapped in a screening cascade (Figure 1.18) that sets gating guidelines for each level of the screening process, from initial activity screening through in vivo animal efficacy studies. The screening cascade is designed to decrease the number of

![FIGURE 1.17](Image) The identification of a clinical candidate requires consideration of a variety of properties beyond activity at the biological target of interest. Drug discovery and development programs seek to optimize as many of these properties as possible in order to identify the best opportunity for success.
compounds examined at each level in order to ensure that compounds with flaws are removed as early as possible. The cascade, also referred to as a screening tree, begins with \textit{in vitro} profiling and then transitions into \textit{in vivo} studies designed to determine a compound’s pharmacokinetic profile and demonstrate efficacy in an appropriate animal model. At the top of the cascade, compounds are screened for activity against the biological target and a threshold of interest is generally set to determine if compounds are active enough to warrant further investigation. Potency is, of course, an important issue, as dosing requirements are lower for compounds that are more potent. All other things being equal, compounds with higher potency can be dosed at lower levels, decreasing the likelihood of side effects. A compound with target potency of 5 nM in theory could be provided to a patient at a significantly lower dose than a compound serving the same function but with a potency of 5 μM.

Once a compound has satisfied the potency criteria, selectivity and physicochemical properties criteria are typically examined. Nature has developed exquisite systems to accomplish very specific tasks with highly selective systems, but many of these systems overlap structurally, and this can have a significant impact on the biological properties of a given test compound. Thus, the next biological screening step in a typical screening cascade is often an assessment of a compound’s potency at biological
systems that are closely related to the target of interest. The Kv1.5 channel, for example, is a voltage-gated potassium channel that has been the target of research programs atrial arrhythmia, and many compounds have been identified that can block this channel with a high level of potency.82 There are, however, over 70 other voltage-gated potassium channels with varying degrees of similarity to the Kv1.5 channel, and undesired activity at any of these related channels could create unwanted side effects in human or animal studies. For example, the Kv1.5 channel is closely related to the hERG channel. Blockade of the hERG channel has been linked to torsade de pointes and sudden cardiac death,83 so any compound moving forward in this area would need to be counterscreened for hERG activity in order to ensure that advancing compounds do not present a risk of sudden cardiac death in a clinical setting. This is a rather extreme example of the importance of proper selectivity, but it should be clear that failure to achieve proper target selectivity in this area represents a significant barrier to moving a program forward.

Similarly, there are over 500 known kinase enzymes,84 and any drug discovery program designed to target a single kinase, or even a family of kinases, has an associated risk of identifying compounds that are active at multiple members of this large family of related enzymes. In order to mitigate this risk, kinase programs routinely screen test compounds against panels of related kinases in order to understand the risks associated with off-target activity.

In general, compounds that are potent at the target of interest, but are also potent at a variety of other targets (“promiscuous compounds”), do not move forward in a drug discovery program, as the risk of undesired (or unpredicted) side effects is too high. The level of selectivity required, however, is dependent on the program, the nature of the potential side effect presented by off-target activity (off-target activity leading to excessive hair growth might be tolerable, whereas sudden cardiac death through poor hERG selectivity is not), the target patient population, duration of treatment (some side effects only appear upon extended exposure to a drug), and a variety of other factors. Overall, target selectivity is a major factor to consider.

An active and suitably selective compound, however, is not necessarily a good drug candidate. Physicochemical properties also play a major role in determining whether or not a compound is suitable for further investigation. In vitro screens designed to predict absorption, distribution, metabolism, and excretion (in vitro ADME) are generally performed early in a program in order to ensure that candidates reaching the drug development pipeline are “druglike” in nature (Figure 1.19). Compounds that have poor aqueous solubility, for instance, are often difficult to develop as drugs. In order for a drug to exert an influence on a biological target, it must be soluble in biological fluid at a level consistent with its potency. Thus, the level of
solubility required for a given compound is directly linked to its potency. As potency increases, the requisite solubility decreases, as less drug is required to provide the intended effect. Solubility also has a direct impact on absorption, as a compound must be soluble in biological fluids in order for it to successfully pass through a biological membrane and reach its intended target.

The ability of a compound to penetrate cellular membrane (its permeability) can also be a determining factor in the success or failure of a given candidate compound. If a compound is potent, selective, and soluble, but unable to pass through a biological membrane, it may not be able to reach the target of interest and fail to demonstrate the desired efficacy. Orally active drugs must be absorbed in the gastrointestinal tract, and drugs that target intracellular system must also pass through the cell membrane in order to reach their intended targets. Extracellular targets, of course, do not face this added issue, but CNS drug candidates face the added complexity of required permeability through the blood–brain barrier (BBB). Additionally, there are efflux pumps (e.g., P-glycoproteins (Pgps)) designed to remove xenobiotic material that can limit permeability, preventing efficacy. The inability of a compound to penetrate a cell membrane represents a significant issue that could prevent further investigation of the candidate compound.

Metabolic and chemical stability are also important considerations. If all of the previously mentioned criteria are met, and a compound is able to enter the body, but is immediately metabolized, efficacy studies will fail to show the desired results. However, the relative rate of metabolism that can be allowed for a successful compound depends on the goals of the project. If, for example, the goal is to develop a new antibacterial agent, then high-metabolic stability will likely be desired so that the potential drug candidate will be available in the circulation long enough to kill the
invading organism. If, however, the goal is to develop a new surgical anesthetic, metabolic stability may be less of an issue, as it may be desirable for drug efficacy to fade rapidly upon termination of dosing regimens.

In a related sense, compounds that have chemical stability issues may be problematic as drugs. Special packaging systems, some as simple as amber bottles for light-sensitive compounds or cold storage, may be required in order for the drug to be available commercially when a patient is in need. While these kinds of issues are not insurmountable, generally speaking, more chemically stable compounds are preferred.

It is also important to consider how candidate compounds may impact the normal metabolic processes, potentially altering the metabolism of drug products used in tandem with the candidate compound. Inhibition of key metabolic enzymes in the liver, such as Cyp3A4, Cyp2D6, and Cyp2C9, members of the cytochrome P450 (Cyp450) family of metabolic enzymes, are often studied using in vitro screening methods (liver microsomes) in order to determine the risk of drug–drug interactions. A compound that meets all other in vitro criteria and demonstrates efficacy, may still fail as a drug candidate if it is determined that there is significant risk of drug–drug interactions. The withdrawal of Seldane from market is a classic example of the risks associated with unintended inhibition of the normal metabolic processes, and is discussed in greater detail in Chapter 13.

Positive results through the in vitro screening portion of a discovery program represent a significant accomplishment, but are still not necessarily indicative of success. The pharmacokinetic properties (PK) of a candidate compound must be determined in order to answer key questions about the in vivo fate of the potential drug candidate. For example, if the candidate compound is dosed orally, what percentage of the oral dose actually reaches the systemic circulation? How rapidly is the candidate compound excreted or metabolized? Does the compound reach systemic concentrations high enough to suggest that in vivo efficacy should be expected in an animal model? Is the compound freely distributed through the body, or does it concentrate in a particular organ or tissue type? The answer to these and a number of similar questions will have a significant impact on the ability of any given compound to provide the desired in vivo efficacy in a given animal model. Irrespective of the positive results of in vitro screening, compounds with poor PK profiles are not likely to be successful drugs.

Compounds found to possess suitable PK profiles must, of course, demonstrate activity in key animal efficacy trials before they can be considered for clinical study. The type of efficacy studies required is based on the desired biological endpoint (disease state), and a full discussion of in vivo efficacy models is well beyond the scope of this text. Some examples are given in chapter 7, but it should be clear that the ultimate goal of a discovery program is to identify compounds that meet all of the aforementioned
in vitro criteria, demonstrate efficacy in the appropriate animal model, and have PK properties consistent with the desired dosing regimen.

Safety and side-effect profiles are also major concerns, and there are a number of in vitro and in vivo screens that can be used to assess the risks associated with a compound (e.g., in vitro hERG screening,88 Ames mutagenicity screening,89 dog cardiovascular safety assessment90). The nature and scope of safety studies is well beyond the scope of this text, but should always be a major concern in the minds of drug discovery scientist as any project moves forward. It is also important to realize that the side-effect profiles for a potential clinical candidate are somewhat dependent on the intended use. For example, compounds used to prevent life-threatening illness, such as cancer, AIDS, and ALS, may be given more latitude with their side-effect profiles, given the severity of the illness. On the other hand, treatments designed for chronic use or non-life-threatening conditions, such as osteoarthritis or neuropathic pain, must be scrutinized for possible safety issues or side effects. The concern for safety bridges through all aspects of discovery and development of novel therapeutics. It is impossible to guarantee that compounds entering clinical development will be safe, but compounds with “red flags” in safety screens are generally not pursued as drugs.

Finally, the ability to identify patentable compounds will also gate the progress of any drug discovery program. As mentioned earlier, drug discovery and development is an exceptionally expensive endeavor. Market exclusivity through patent protection provides the necessary financial incentive required for companies to invest in new drug development. Compounds and compound classes that cannot be protected through the issuance of patents are unlikely to be pursued by private organizations, as recouping the significant investments required to move a compound into clinical use becomes challenging. Further information regarding the importance of patent protection in the pharmaceutical industry is provided in Chapter 12.

Drug discovery scientists walk on the edge of several precarious slopes in attempting to identify potential new therapeutic entities. Balancing the needs of potency, selectivity, solubility, stability, pharmacokinetics, safety, and novelty is critical to the success of any project, and failure to deliver in any one of these areas can terminate the forward progression of a test compound.

QUESTIONS

1. What are the three major phases of drug discovery?
2. What are the four major phases of drug development?
3. Describe the lead optimization cycle.
4. What is a screening cascade (also referred to as a screening tree)?
5. Why is compound selectivity an important aspect of drug discovery?
6. What does the term *in vitro* ADME refer to?
7. Name five properties that are a part of a compound’s *in vitro* ADME profile.

References


31. Add Notation for Medicinal Chemistry Text.


64. Chemical Abstracts Services. A Division of the American Chemical Society. website: [http://www.cas.org/content/chemical-substances](http://www.cas.org/content/chemical-substances).


c. In High-Throughput Screening in Drug Discovery; Hüsér, J., Ed., 2006. (Mannhold, R. (Series Editor); Kubinyi, H. (Series Editor); Follers, G. (Series Editor). Wiley-VCH.

REFERENCES


