PART I

DRUG DISCOVERY APPROACHES AND TECHNOLOGIES

1

THE DRUG DISCOVERY PROCESS

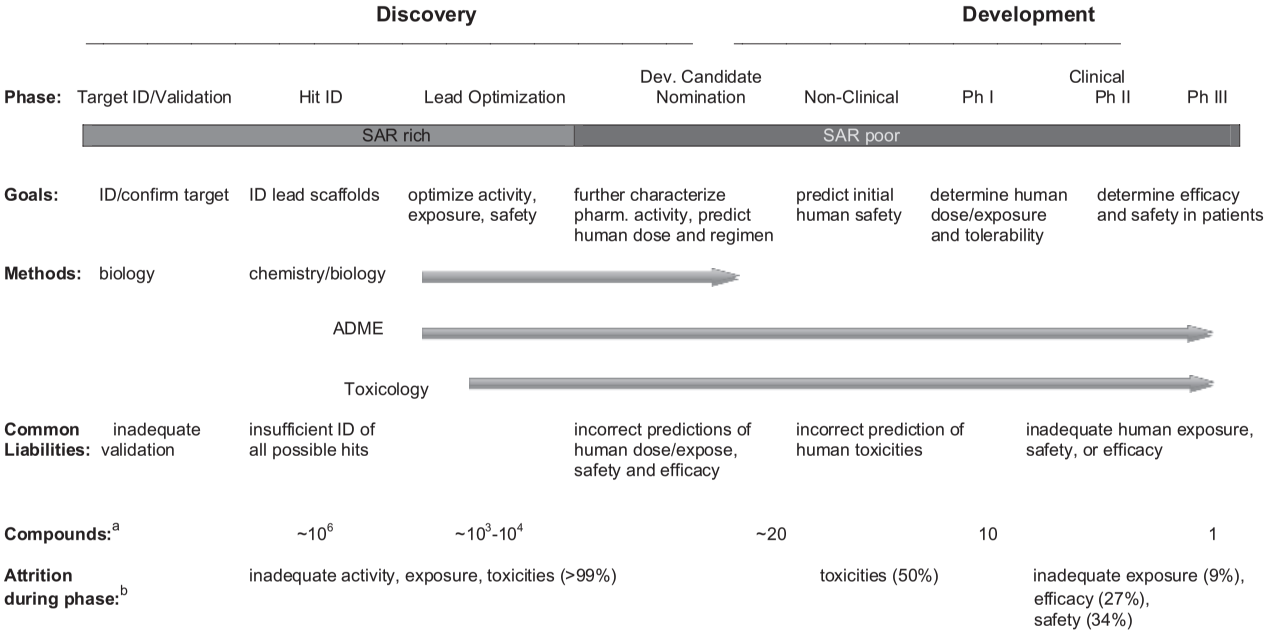
Gerald T. Miwa

1.1 INTRODUCTION

The discovery and development of drugs is an inefficient process. Only one of approximately 10,000 compounds synthesized reaches the marketplace, and this requires approximately 10–15 years and $800,000,000 in R&D expenditures (Khosla and Keasling, 2003). The objective of drug discovery should be to identify a compound that will prove to be safe and effective against the intended disease with minimal attrition due to toxicities, inadequate exposure, or unsuccessful translation of target modulation to clinical benefit against the disease. Drug discovery, the process used to select a compound for drug development, is common in most pharmaceutical and biotechnology companies. This process is composed of the following phases: target identification and validation, hit identification, lead optimization, and development candidate nomination (Fig. 1.1).

Target identification and validation is the process used to identify and confirm that the modulation of a biological target will produce a desired therapeutic effect. The methods employed for this phase are mainly biological. The confirmation or validation of the utility of the target in modifying a human disease is critical and remains the greatest potential issue of this phase because of its implication in clinical failures due to inadequate efficacy. Hit identification is the process employed to initially identify molecules that interact with the target. Both biological and chemical methods are used to identify hits. The methods employed, although more comprehensive than in the past, may still not identify all possible hits. Usually, more stringent selection criteria are

Enzyme Inhibition in Drug Discovery and Development, edited by Lu and Li



[a] adapted from Khosla 2003

[b] major causes of attrition during a phase and (percentage of total attrition due to these causes), adapted from Kola and Landis 2004

Figure 1.1. Goals and challenges in the processes of drug discovery and development.

implemented to screen hits in order to identify lead compounds that have the potential to improve further with structure modifications. Lead optimization is the chemical structure–activity optimization process that identifies the best possible drug-like molecule. Usually absorption, distribution, metabolism and excretion (ADME) and toxicology assays are added to the biological and chemical methods during lead optimization. Development candidate nomination is the process used to further characterize the potential exposure, efficacy, and safety of the nomination candidate and to judge if the molecule is suitable for drug development. The most common liabilities remaining from this phase are inaccurate predictions of human exposure, safety and efficacy. Although the objectives are common among companies engaged in this process, the methods, issues, and acceptance standards vary. Nevertheless, the best measure for the success of drug discovery is the demonstration of a compound's effectiveness and safety in patients. This chapter describes the elements of the drug discovery process and comments on its successfulness and areas currently under evaluation to improve success.

1.2 TARGET IDENTIFICATION AND VALIDATION

Target identification is the process used to identify potential therapeutic targets amenable to modulation by drug molecules, antibodies, aptamers, or gene modulators such as siRNAs and antisense oligonucleotides. The identification of potential targets for therapeutic intervention by drugs has greatly improved in the past two decades. Prior to the mid-1980s, drug targets were identified by the serendipitous discovery of active agents such as the penicillins and the benzodiazepines, through the symptomatic changes in disease models in animals such as cardiovascular drugs or through activity in suitable *in vitro* systems such as for anti-infective drugs. Biochemical targets were identified through their postulated relevance in pathways thought to be involved in disease processes such as HMG-CoA reductase in cholesterol biosynthesis and coronary heart disease (Tobert, 2003), and the H2-receptor in gastric acid secretion and gastric ulceration. In rare cases, targets could be identified and validated in humans through existing genetic mutations in the human population such as the deficiencies in 5-alpha reductase (Johnson et al., 1986) which led to the 5-alpha reductase inhibitor class of drugs for benign prostatic hypertrophy and propecia.

During the past two decades, greater knowledge and newer biological methods have permitted the mining of patient samples and animal models of diseases to elucidate the probable genes implicated in the disease etiology. Techniques such as gene expression profiling and comparative genomics have been valuable in identifying potential targets. For example, the capability to identify the overexpression of HER2 in the diseased tissues of some breast cancer patients was used to identify HER2 as a new target for metastatic breast cancer and led to the discovery of Herceptin (Chang, 2007).

Single-nucleotide polymorphisms (SNP) genotype–phenotype correlations from patient samples have also yielded many more, albeit clinically less validated, potential targets for drug design (Wunber et al., 2006). In addition, greater knowledge of the role of specific receptors, cell cycle regulators, enzymes, and other proteins in biochemical pathways has led to the identification of potential drug targets at the organism and biochemical levels.

The complete sequencing of the human genome ushered in the potential to vastly increase the number of drug targets. It has been estimated that there are as many as 10,000 potential drug target genes in the human genome, which is comprised of approximately 30,000 to 50,000 genes (Venter et al., 2001; IHGSC, 2001). In addition, this may underestimate the total number of possible protein targets because of splice variants and post-translational modifications (Pillutla et al., 2002). More recently, nonreceptor or noncatalytic protein targets such as protein–protein interactions and nonprotein targets such as DNA and mRNA have also been recognized. This is an enormous increase over the approximately 43 protein targets for the 100 best selling drugs (Zambrowicz and Sands, 2003) and is the basis for optimism for the future of the pharmaceutical industry. Of course, many, perhaps most, of the proteins will not be suitable for pharmaceutical development. Since there are thousands of potential protein targets, a growing effort has been made to develop high-thoughput screens to identify potential drug targets through ligand–protein (Pellecchia et al., 2004) and protein–protein interactions with protein microarrays (McBeath and Schreiber, 2000) or cell microarrays (Bailey et al., 2002). This and other technology for screening large protein libraries is comprehensively reviewed by Pillutla et al. (2002).

Target validation is the assessment and evidence supporting the linkage of a target to the etiology of the disease or pathology and the amelioration of the disease by modulation of the target. In other words, it is the convincing demonstration that modulation of the target, usually through inhibition, results in amelioration of the disease. Target validation is critical and, arguably, the most important step in translating a new potential target into a viable drug target because of its role in achieving efficacy in patients, currently one of the major reasons for drug attrition in the clinic (Kola and Landis, 2004).

The most accepted criteria for target validation during drug discovery are based on three categories: (1) demonstration of the target protein expression or mRNA in relevant cell types or in the target tissues from animal models or patients, (2) demonstration that modulation of the target in cell systems results in the desired functional affect, and (3) demonstration that the target has a causal role in producing the disease phenotype in animal models and/or patients (Windler et al., 2003).

The demonstration of protein expression is usually accomplished in diseased or target tissues by *in situ* hybridization or immunocytochemistry. For example, the localization of orexin peptides and receptors in the hypothala-

mus provided evidence for its role in the regulation of feeding (Sakurai et al., 1998). *In situ* hybridization permits gene detection only at the transcriptional or mRNA level, whereas immunocytochemistry identifies and locates the protein expression product but requires appropriate antibodies.

Today, the association of the target protein with diseased or target tissue is rarely considered sufficient for target validation. The functional association of the target with disease modification is also required. The demonstration that modulation of the target results in the desired functional affect can often be explored or demonstrated through inhibition of the target by transgenic (Tornell and Snaith, 2002) and gene knockout mice, small molecule inhibitors, antisense oligonucleotides, and small interfering RNA (siRNA). Antisense oligonucleotides inhibit gene expression through complementary binding to single-stranded DNA or RNA. They have been used to characterize the function of genes as well as potential drugs. In contrast, siRNAs inhibit gene expression through complementary binding to mRNA followed by catalytic destruction of the mRNA (Natt, 2007; Behike, 2006; Hammond et al., 2000). Consequently, siRNAs have rapidly eclipsed antisense oligonucleotides as a tool for inhibiting gene expression and target validation (Szymkowski, 2003). Scholarly reviews of techniques and examples of target identification and validation can be found in Natt (2007), Behike (2006), Pillutla et al. (2002), and Ohlstein et al. (2000). Proteomic methods have been less successful in identifying or validating potential drug targets because analytical techniques for exploring the entire proteome have not been adequate (Kopec et al., 2005). However, the recent development of cell microarrays (Bailey et al., 2002) may provide more feasible methods for quickly assessing disease-based changes in the proteome.

Despite the enormous advances in the technologies for examining the genome and proteome in cell-based and animal models of diseases and for validating potential drug targets with these methods, there remains considerable uncertainty about the prospective translation of these findings to human diseases (Williams, 2003). Today, target validation still ultimately depends on retrospective verification from clinical studies while the lack of efficacy in patients continues to be a major source of drug attrition. Consequently, the current performance of target validation has been criticized (Sams-Dodd, 2005). This had led to more holistic approaches for target validation such as the confirmation of activity in cell systems (Kenakin, 2003) and target deconvolution (Terstappen et al., 2007). Target deconvolution is the identification of both target and off-target proteins and their functional roles affected by a compound in mammalian cells or model organisms such as zebrafish or nematodes. Both target and off-target proteins are identified by a growing battery of techniques that include the use of the compound as an immobilized ligand for binding and isolating these proteins through affinity chromatography and through the gene expression changes caused by the compound. A more integrated preclinical and clinical paradigm has also been advocated for drug discovery (Schadt et al., 2003).

1.3 HIT IDENTIFICATION

Up to this point in the discovery process, the focus has been on the target and not the drug. Following target validation sufficient to create confidence that the target is involved in the etiology of a disease and that modulation of the target will result in the reduction of the phenotypic expression of the disease, the search for a drug molecule is initiated. Hit identification is the first stage of this process. For small molecules, the goal is to identify the core molecular structure that demonstrates promise for further structure optimization.

There are a number of strategies for identifying hits. With the advent of large compound libraries through combinatorial synthesis, proprietary compound collections, and natural products and the increased feasibility of screening provided by automated analysis and data informatic methods, high-throughput screening (HTS) of large compound libraries against potential targets emerged as a major source of compound hits. This approach was eagerly pursued, especially by large pharmaceutical companies, during the 1990s. Although the random screening of large combinatorial libraries is less common today, many companies continue to use HTS of focused compound libraries as a source of hits. Focused libraries are subsets of chemical libraries that are chosen based on a company's proprietary knowledge of the target structure or by other chemoinformatic criteria (Goodnow, 2006). In addition to the use of smaller, more focused compound libraries to screen for chemical hits, a growing effort to identify higher-quality hits through further characterization has been adopted by some pharmaceutical companies. The most common approach is the characterization of desirable drug-like properties using assays that can be easily implemented in a high-capacity and rapid-turnaround format such as those for drug absorption, metabolism, drug interactions, and cardiac ion channel modulation. For example, *in vitro* assays for oral absorption such as the Caco-2 cell system, metabolic stability such as the hepatic liver microsomal system, the potential for drug interactions with CYP inhibition assays, and the potential for cardiac QT interval prolongation with hERG binding have been moved to the end of the hit identification stage to provide higher-quality characterization to the leads they generate (Bleicher et al., 2003).

Computational or modeling techniques have also been employed as a source of hits (Balakin et al., 2006). For example, ligand-based design employs known binding ligands to map out the structural binding features of the target. This computational method does not require knowledge of the structure of the binding domain of the target but does depend on a comprehensive structure–activity relationship with known ligands. A more refined computational method, structure-based design, can be employed when the three-dimensional structure of the ligand-binding domain of the target protein is known. Modeling of ligands docking to the target provides a means for identifying potential hits (Kitchen et al., 2004). Actual HTS screening with focused compound libraries can then be used to confirm these potential hits or identify new hits.

An alternative or complementary approach to HTS with compound libraries is fragment-based lead design (FBLD). This method has been gaining popularity in the last decade (Hajduk and Greer, 2007; Everts, 2008) and is based on measuring the binding of smaller chemical fragments or functional groups to target proteins rather than the functional assay of the interaction of the larger molecules containing more chemical fragments employed with HTS. The strategy of FBLD is to identify individual chemical functional groups (such as amino, carboxylate, carbonyl, aryl, etc.) that bind to different sites on the target and then combine them in a single molecule in order to create compounds with higher binding affinities. One strategic advantage of FBLD over HTS is that a smaller library of compounds is required to define a greater fraction of the total number of possible chemical scaffolds (Hajduk and Greer, 2007). For example, it has been estimated that for scaffolds of MW 500, there are 10^{62} possible scaffolds while the usual HTS screen is limited to about 10^6 compounds leaving a vast amount of chemical space unscreened. In contrast, the smaller size of FBLD fragments, MW < 200, results in approximately 10^7 possible combinations requiring a much smaller fragment library of 10^4–10^5 to give much better screening coverage. The low coverage of all possible scaffolds possible with HTS may be a significant contributor to the disappointing success of this approach in identifying clinical candidates (Drews, 2000).

Although the functional assays of HTS are often not sensitive enough to detect the lower affinity binding of fragments, a number of general methods have been applied to successfully screen FBLD libraries such as NMR detection of the chemical shifts produced from fragment binding to the target protein, X-ray crystallographic detection of ligand binding, and surface plasmon resonance, which measures the change in protein surface refractive index upon ligand binding. To date, FBLD has yielded several clinical candidates from a number of companies such as Abbott, Astex Therapeutics, Novartis, Plexxikon, and Sunesis (Everts, 2008).

1.4 LEAD OPTIMIZATION

Hit identification, through the screening of compound libraries, FBLD, or other sources, provides the entry point for the drug discovery phase known as lead optimization (LO). Lead optimization is a systematic effort to maximize the pharmacological and drug exposure properties of a lead candidate while minimizing its toxicity and drug–drug interaction potential through structure modifications (refer to Chapter 2 in this book). Often, some assessment of certain acute toxicity markers and factors that can contribute to inadequate exposure or overexposure is either conducted systematically or conducted occasionally during LO or as part of the compound characterization stage prior to the nomination of a compound to development. The most common toxicity and exposure assays employed during this stage are: hERG binding for potential adverse cardiac ion channel binding, cytotoxicity assays

as predictors of organ toxicities, limited AMES assays to identify direct acting genotoxins, the evaluation of chemically reactive metabolites (Baillie, 2006), the identification of off target receptors with the NOVA screen (Kramer et al., 2007; Sasseville et al., 2004), the identification of major metabolic pathways (refer to Chapter 5 in this book), and the evaluation of drug–drug interaction potential through the identification of enzymes and transporters responsible for the clearance of the compound, the evaluation of polymorphic CYPs such as 2C9, 2C19, and 2D6 as primary clearance enzymes, the evaluation of the induction potential for CYPs involved in drug clearance (refer to Chapter 8 in this book) and the inhibition of these enzymes and drug transporters (refer to Chapters 7, 15, and 17 in this book; Balani et al., 2005). In some cases, acute single or repeat dose *in vivo* toxicology studies are conducted, usually based on the observation of organ toxicities from previous compounds in the class or LO program.

The first step in LO is the identification of the pharmacophore or component groups on the lead molecule responsible for the desired interaction with the target. Often, the pharmacophore is only a small portion of the entire molecule, and other portions are either (a) providing the framework to support the pharmacophore for either favorable or less favorable binding to the target or (b) extraneous and have no effect on binding of the pharmacophore. Identification of the pharmacophore, framework, and extraneous portions of the lead molecule requires systematic structural modifications, most commonly through selective deletion, of portions of the lead molecule to define these components. Lead optimization can then be pursued through selective additions, deletions, and functional group modifications of the lead molecule to maximize activity. Of course, it is also essential to optimize the biopharmaceutical and exposure or pharmacokinetic properties while minimizing the toxicity potential during LO. If LO is conducted *in vivo*, the structure can also be directly optimized for pharmacokinetic properties such as bioavailability, systemic exposure, distribution to target organs or tissues, and duration of action while minimizing the potential for drug–drug interactions via induction or inhibition of clearance enzymes (refer to Chapter 6 in this book). The systematic structure modifications lead to an understanding of the relationship between chemical structure and the activity, exposure, and safety properties of the lead. This structure–activity relationship (SAR) is often the most time-consuming phase of LO, requiring hundreds to thousands of compounds to optimize these properties.

1.5 CRITERIA FOR THE DEVELOPMENT CANDIDATE

During the last phase of LO, when one or a few compounds are judged to possess the optimal properties of exposure, activity, and safety, they are usually more fully characterized to verify their potential as drug development candidates. The most common discovery criterion for drug development is

that there is compelling enough evidence to merit confidence that the compound will have the desired efficacy, exposure, and safety properties in the clinic. The specific criteria in each of these areas are compound-specific and depend on factors such as the seriousness of the disease, the properties of previous development candidates, the strength of the patent position, and the needed ease of administration of the dose and dose regimen to ensure compliance. In addition, there are a number of nondiscovery criteria that also influence the decision to advance a compound into development such as the competition from other drugs in development and existing marketed drugs, the anticipated manufacturing requirements and cost of producing drug product, the projected development times, the anticipated development and regulatory issues, and the market potential and other commercial issues.

Although the majority of development candidates arise from the drug discovery process or candidates that are licensed in from other companies, there are other, less common, sources for drug candidates. Drug metabolites have been a source for development compounds (Fura, 2006), and some notable examples have gone on to become drugs such as oxazepam, pravastatin, and fexofenadine. Side effects or off-target effects observed in the clinic have led to new or increased indications for some drugs such as Viagra, marketed for erectile dysfunction although originally developed for angina and hypertension, and Zyban, marketed for smoking cessation although originally developed as an antidepressant.

1.6 PERSPECTIVE

Pharmaceutical companies pursue similar processes for discovering new drugs. Although each company has its own specific criteria and metrics for measuring success during drug discovery, the most appropriate metric should be the success in meeting safety and efficacy criteria in patients with minimal compound attrition. By this measure the pharmaceutical industry has not become more successful and actually may have become less successful during the past decade in its efforts to discovery new drugs (Mathieu, 2007). As an industry, pharmaceutical companies have greatly advanced in their knowledge and accessibility to new drug targets and the technology for screening for hits. During the last two decades, the addition of screening assays relevant to drug disposition and drug–drug interactions in humans and the optimization of ADME properties during hit identification and lead optimization has had remarkable success in reducing the attrition from inadequate drug exposure during clinical development (Kola and Landis, 2004). The process and methods employed for lead optimization of pharmacological activity have not seen as dramatic an improvement in speed or scope, but their success in identifying development candidates has been maintained.

The greatest challenges facing successful drug discovery are those closest to meeting the primary objectives for a new drug: safety and efficacy. Only 1

of 10 compounds entering clinical development is approved as a new drug. The main contributors to this 90% attrition during clinical development are inadequate efficacy (27%) and inadequate safety (34%) (Fig. 1.1). The absence of better models and methods to predict drug efficacy for many new disease targets and to predict toxicities in humans portend that these will continue to be the major sources for drug attrition during development. This is recognized by the industry, and there is a growing emphasis on better target validation and predictive toxicity screens that can be implemented earlier in drug discovery. Solutions to these needs will be crucial to improving the success of the discovery process.

REFERENCES

Bailey SN, Wu RZ, Sabatini DM. Applications of transfected cell microarrays in high-throughput drug discovery. *Drug Discovery Today* 2002;7:16.

Baillie T. Future of toxicology-metabolic activation and drug design: challenges and opportunities in chemical toxicology. *Chem Res Toxicol* 2006;19:889–893.

Balakin KV, Kozintsev AV, Kiselyov AS, Savchuk NP. Rational design approaches to chemical libraries for hit identification. *Curr Drug Discovery Technol* 2006;3: 49–65.

Balani SK, Miwa GT, Gan L-S, Wu J-T, Lee FW. Strategy of utilizing *in vitro* and *in vivo* ADME tools for lead optimization and drug candidate selection. *Curr Topics Med Chem* 2005;5:1033–1038.

Behike MA. Progress towards *in viv*o use of siRNAs. *Mol Ther* 2006;13:644–670.

Bleicher KH, Bohm H-J, Muller K, Alanine AI. Hit and lead generation: beyond high-throughput screening. *Nat Rev: Drug Discovery* 2003;2:369–378.

Chang JC. HER2 inhibition: From discovery to clinical practice. *Clin Cancer Res* 2007;13:1–3.

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

Everts S. Piece by piece. *Chem Eng News* July 21, 2008:15–23.

Fura A. Role of pharmacologically active metabolites in drug discovery and development. *Drug Discovery Today* 2006;11:133–142.

Goodnow RA Jr. Hit and lead identification: integrated technology-based approaches. *Drug Discovery Today: Technologies* 2006;3:367–375.

Hajduk PJ, Greer J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev: Drug Discovery* 2007;6:211–219.

Hammond S, Bernstein E, Beach D, Hannon G. An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 2000;404:293–296.

IHGSC International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001;409:860–921.

Johnson L, George FW, Neaves WB, Rosenthal IM, Christensen RA, Decristoforo A, Schweikert HU, Sauer MV, Leshin M, Griffin JE. Characterization of the testicular abnormality in 5 alpha-reductase deficiency. *Clin Endocrinol Metab* 1986;63: 1091–1099.

Kenakin T. Predicting therapeutic value in the lead optimization phase of drug discovery. *Nat Rev: Drug Discovery* 2003;2:429–438.

Khosla C, Keasling JD. Metabolic engineering for drug discovery and development. *Nat Rev: Drug Discovery* 2003;2:1019–1024.

Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev: Drug Discovery* 2004;3: 935–949.

Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nature Rev: Drug Discovery* 2004;3:711–715.

Kopec KK, Bozyczko-Coyne D, Williams M. Target identification and validation in drug discovery: the role of proteomics. *Biochem Pharmacol* 2005;69:1133–1139.

Kramer JA, Sagartz JE, Morris DL. The application of discovery toxicology and pathology towards the design of safer pharmaceutical lead candidates. *Nat Rev: Drug Discovery* 2007;6:636–649.

Mathieu MP, editor. Statistics on drug development: costs/complexity, development time, success rates. In: *Parexel's Bio/Pharmaceutical R&D Statistical Sourcebook 2007/2008*. Waltham, MA: PAREXEL International Corp.; 2007, p. 229.

McBeath G, Schreiber SL. Printing proteins as microarrays for high throughput function determination. *Science* 2000;289:1760–1763.

Natt F. siRNAs in drug discovery: target validation and beyond. *Curr Opin Mol Ther* 2007;9:242–247.

Ohlstein EH, Ruffolo RR, Jr, Elliott JD. Drug discovery in the Next Millennium. *Annu Rev Pharmacol Toxicol* 2000;40:177–191.

Pellecchia M, Becattini B, Crowell KJ, Fattorusso R, Forino M, Fragai M, Jung D, Mustelin T, Tautz L. NMR-based techniques in the hit identification and optimization processes. *Expert Opin Ther Targets* 2004;8:597–611.

Pillutla RC, Fisher PB, Blume AJ, Godstein NI. Target validation and drug discovery using genomic and protein–protein interaction technologies. *Expert Opin Ther Targets* 2002;6:517–531.

Sakurai T, Amemiya A, Ishii M, Matsuzaki I, Chemelli RM, Tanaka H, Williams SC, Richardson JA, Kozlowski GP, Wilson S, Arch JR, Buckingham RE, Haynes AC, Carr SA, Annan RS, McNulty DE, Liu WS, Terrett JA, Elshourbagy NA, Bergsma DJ, Yanagisawa M. Orexins and oresin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* 1998;92:573–585.

Sams-Dodd F. Target-based drug discovery: Is something wrong? *Drug Discovery Today* 2005;10:139–147.

Sasseville VG, Lane JH, Kadambi VJ, Bouchard P, Lee FW, Balani, SK, Miwa GT, Smith PF, Alden CL. Testing paradigm for prediction of development-limiting barriers and human drug toxicity. *Chem-Biol Int* 2004;150:9–25.

Schadt EE, Monks SA, Friend SH. A new paradigm for drug discovery: integrating clinical, genetic, genomic and molecular phenotype data to identify drug targets. *Biochem Soc Trans* 2003;31:437–443.

Szymkowski DE. Target validation joins the pharma fold. *Targets* 2003;2:8–9.

Terstappen GC, Schlupen C, Raggiaschi R, Gaviraghi G. Target deconvolution strategies in drug discovery. *Nat Rev: Drug Discovery* 2007;6:891–903.

Tobert JA. Lovastatin and beyond: the history of the HMG-CoA reductase inhibitors. *Nat Rev Drug Discovery* 2003;2:517–526.

Tornell J, Snaith M. Transgenic systems in drug discovery: from target identification to humanized mice. *Drug Discovery Today* 2002;7:461–470.

Venter JC, Adams MD, Myers EW. The sequence of the human genome. *Science* 2001;291:1304–1349.

Williams M. Target validation. *Curr Opin Pharmacol* 2003;3:571–577.

Windler H. Target validation requirements in the pharmaceutical industry. *Targets* 2003;2:69–71.

Wunber T, Hendrix M, Hillisch A, Lobell M, Meir H, Schmeck C, Wild H, Hinzen B. Improving the hit-to-lead process: data-driven assessment of drug-like and lead-like screening hits. *Drug Discovery Today* 2006;11:175–180.

Zambrowicz BP, Sands AT. Knockouts model the 100 best-selling drugs—will they model the next 100? *Nat Rev Drug Discovery* 2003;2:38–51.