Annu. Rev. Pharmacol. Toxicol. 2000. 40:177–91
Copyright © 2000 by Annual Reviews. All rights reserved

DRUG DISCOVERY IN THE NEXT MILLENNIUM

Eliot H. Ohlstein, Robert R. Ruffolo Jr., and John D. Elliott

Departments of Cardiovascular Pharmacology and Medicinal Chemistry, Research & Development, SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406–0939; e-mail: ohlstein@sbphrd.com, robert_r_ruffolo@sbphrd.com, john_d_elliott@sbphrd.com

Key Words target validation, functional genomics, proteomics, high-throughput screening, combinatorial chemistry, cheminformatics, gene therapy

■ **Abstract** Selection and validation of novel molecular targets have become of paramount importance in light of the plethora of new potential therapeutic drug targets that have emerged from human gene sequencing. In response to this revolution within the pharmaceutical industry, the development of high-throughput methods in both biology and chemistry has been necessitated. This review addresses these technological advances as well as several new areas that have been created by necessity to deal with this new paradigm, such as bioinformatics, cheminformatics, and functional genomics. With many of these key components of future drug discovery now in place, it is possible to map out a critical path for this process that will be used into the new millennium.

INTRODUCTION

The Human Genome Project was initiated on October 1, 1990, and according to the original plan, the complete DNA sequence of the human genome would be achieved by the year 2005 (1). However, improvements in technology and an increase in interest in human DNA sequence have placed the project 2 years ahead of schedule; it is now anticipated that the Human Genome Project will be completed in the year 2003. Furthermore, a partial blueprint of the human genome is scheduled to be available as early as March 2000. Gene identification provides a new paradigm for understanding human disease at its most fundamental level; a knowledge of the genetic control of cellular functions will constitute the conceptual underpinnings of future strategies for the prevention and treatment of disease. It is of interest that the identification in the 1980s of the gene believed to be responsible for cystic fibrosis took researchers approximately 9 years to discover, whereas the gene responsible for Parkinson's disease was recently identified within a period of several weeks (2). This quantum leap in the ability to associate a specific gene with a disease can be attributed primarily to the extraordinary

0362–1642/00/0415–0177$14.00

progress that has been made in the areas of gene sequencing and information technologies.

TARGET VALIDATION

Several thousand molecular targets have been cloned and are available as potential novel drug discovery targets. These targets include more than 750 G-protein–coupled receptors (GPCRs), over 100 ligand-gated ion channels, more than 60 nuclear receptors and 50 cytokines, and approximately 20 reuptake/transport proteins (3). A new potential therapeutic approach for the treatment of a known disease is published nearly every week, as a result of the exponential proliferation of novel molecular and biochemical targets. The sheer volume of genetic information being produced has shifted the emphasis from the generation of novel DNA sequences to the determination of which of these many new targets offer the greatest opportunity for drug discovery. Thus, with several thousand potential targets available, target selection and validation has become the most critical component of the drug discovery process and will continue to be so in the future.

An example of the new paradigm of target selection comes as a result of the pairing of the orphan GPCR, GPR-14, with its cognate neuropeptide ligand, urotensin II. Urotensin II is the most potent vasoconstrictor identified to date, being approximately one order of magnitude more potent than endothelin-1 (4). Thus, GPR-14/urotensin II represents an attractive therapeutic target for the treatment of disorders related to or associated with enhanced vasoconstriction, such as hypertension, congestive heart failure, and coronary artery disease, to name but a few.

The human genome contains approximately 100,000 genes, and any individual tissue expresses between 15,000 and 50,000 of these genes in differing amounts. In diseased tissue, gene expression levels often differ from those observed in normal tissues, with certain genes being over- or underexpressed, or new genes being expressed or completely absent. The localization of this differential gene expression is one of the first crucial steps in identifying an important potential molecular target for drug discovery. In addition to the traditional techniques of Northern blotting analysis, there are a number of newer methods used to localize gene expression. The techniques that typically yield the highest-quality data are in situ hybridization and immunocytochemistry, both of which are labor intensive. For example, in situ hybridization or immunohistochemical localization of a prospective molecular target to a particular tissue or subcellular region is likely to yield valuable information concerning gene function. Recent examples of the success of this approach include the case of the orexin peptides and receptors whose hypothalamic regional localization suggested an involvement in feeding (5). Furthermore, positional cloning of the gene encoding the leptin peptide from ob/ob mice and its subsequent localization to adipocytes has identified this peptide as an important lipostatic factor (6).

Each of these localization techniques has its advantages and disadvantages. In situ hybridization can be initiated immediately following gene sequencing and cloning; however, gene detection is only at the transcriptional mRNA level. Immunocytochemistry, on the other hand, offers the ability to measure protein expression but requires the availability of antibodies having the requisite affinity and selectivity, which may often take several months to generate. With either of these techniques, target localization within the cell is possible at the microscopic level, but it is dependent on the availability of high-quality normal and diseased human tissues, which often represents yet another problem.

The localization of a gene in a particular tissue does not necessarily shed light on all the functions of that gene. As an example, the previously mentioned discovery of orexin as a putative regulator of energy balance and feeding was initially concluded as a result of localization in the dorsal and lateral hypothalamic regions of the brain (5). However, more recently, this gene product was discovered through a positional cloning approach to be a major sleep-modulating neurotransmitter that may represent the gene responsible for narcolepsy (7).

In recent years, new technologies such as microarray gridding (gene-chip) and TaqMan polymerase chain reaction have emerged that would appear destined to play a more prominent role in the high-throughput localization of genes, and the identification of their regulation in disease (8). Gene-chip technology has evolved into two platforms: oligonucleotide (9, 10) and cDNA fragment arrays (11). The inherent problem with oligonucleotide arrays stems from the use of a short hybridization sequence, which can lead to artifacts. cDNA arrays, as a result of their extended sequence, provide higher-quality information due to the specificity and stringency of hybridization. The latter approach does, however, require the generation of long (300–750 bp) cDNAs.

Microarray gridding is already evolving into a procedure that will allow for the comprehensive evaluation of differences in gene expression patterns in normal, diseased, or pharmacologically manipulated systems (8). For genes expressed in low abundance, more sensitive techniques may be required, and reverse transcriptase polymerase chain reaction–based TaqMan technology offers the ability to detect changes in gene expression with as little as two copies per cell. TaqMan technology also has the potential to be developed into a robust methodology for high-throughput tissue localization.

PROTEOMICS

Proteomics offers an alternative, and complimentary, approach to genomic-based technologies for the identification and validation of protein targets, and for the description of changes in protein expression under the influence of disease or drug treatment. Much interest has been expressed by the pharmaceutical industry in proteomics in anticipation of the value of this technology to both discovery and development of new drugs (12).

Proteomics involves the identification and quantitation of gene expression at the protein level. Additionally, proteomics may help to identify protein interaction partners and members of multiprotein complexes. Furthermore, this technique may assist in following time-dependent changes in protein expression levels resulting from selective excitation of a biological pathway, and thereby delineating a cellular protein network, a methodology that has been referred to as functional proteomics.

A recent example of the successful application of a proteomics approach was demonstrated by the identification of a protein (HMG-1) that is a potential late mediator of endotoxin lethality in mice (13). This particular study is noteworthy because it detected changes in HMG-1 release from macrophages induced by endotoxin treatment, whereas the mRNA expression of HMG-1 was not affected by endotoxin treatment, indicating that protein release is not necessarily regulated by gene transcription. These authors further demonstrated that HMG-1 itself is toxic and that anti–HMG-1 antibodies prevent lethality, implicating HMG-1 as a potential target for therapeutic intervention (13). This example clearly demonstrates the complementary nature of proteomic- and genomics-based methods.

In contrast to genomic-based approaches, proteomics is not nearly as well developed as a high-throughput methodology, for a number of reasons. First, proteins have more variable physicochemical properties than does DNA, affecting their behavior, separation, and identification. Second, the abundance of proteins often varies widely. For example, transcription factors are present only at the level of a few copies per cell, whereas very abundant proteins, such as actin, may be present at 10^8 copies/cell. Recently, however, considerable progress has been made in improving detection of low copy proteins through enhancement of gel-loading techniques and enrichment strategies such as affinity-based purification two-dimensional gel separation. Finally, enhanced protein staining/detection methods are now becoming available, and mass spectrometry is pushing the bounds of detection to even more sensitive limits (14). Notwithstanding the technical difficulties that remain, sufficient evidence exists, even at this early stage of the technology, to warrant that proteomics will provide crucial information for the discovery and development of novel therapeutic targets.

HIGH-THROUGHPUT SCREENING

Throughout the 1990s, the pharmaceutical industry has sought to expand its collections of compounds for the purpose of high-throughput screening (HTS) against novel molecular targets (15). A notable early success using this paradigm was the discovery of CP-96,345 (16) (Figure 1), a potent, nonpeptide, neurokinin-1 receptor antagonist. During the past decade, many lead structures have subsequently been unearthed through HTS, particularly for GPCR targets (17, 18). Although the example cited above was a spectacular success (i.e. CP-96,345 has a 50% inhibitory concentration of 3 nM for the NK1 receptor), it has been more

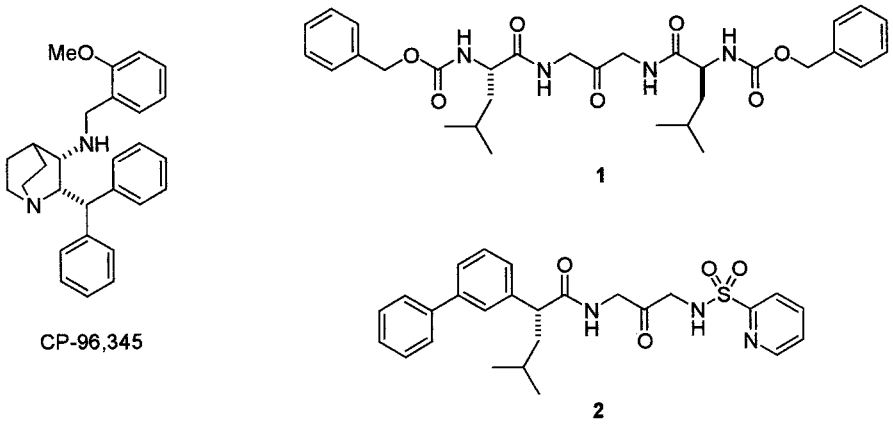

Figure 1 The structure of CP-96,345, a potent neurokinin-1 receptor antagonist obtained by high-throughput screening and compounds 1 and 2, rationally designed cathepsin K inhibitors.

commonly the case that ''hits'' emerging from HTS require more substantial chemical optimization to provide therapeutic agents with the desired level of potency, selectivity, and suitable pharmacokinetic properties (19). Furthermore, the data available from HTS efforts have been of limited utility from the point of view of generating structure-activity relationships capable of directing medicinal chemistry efforts. The combinatorial chemical revolution has led to a situation in which chemotypes previously represented by a few examples in compound collections are now available as arrays of analogs. Thus, at the outset of research programs in the future, more valuable data will be available to guide a medicinal chemistry effort. To fulfill the potential that is promised by this technology, several challenges are currently being addressed. First, industrial compound collections have grown from tens of thousands of compounds in the mid-1990s to hundreds of thousands of compounds today. Fueled by combinatorial chemistry, most major pharmaceutical companies will be screening a million or so discrete compounds against each novel molecular target early into the next millennium.

In rising to the challenge of providing this quantity of data in a timely fashion, the scientist involved in high-throughput screening has sought increasing use of automation, as well as miniaturization, to reduce the demands on precious protein reagents and chemical supplies. Traditional radioligand binding assays are giving way to more rapid and easily miniaturizable homogeneous fluorescence-based methods, a trend almost certain to continue in the future. The increased efficiency of ultra-HTS offers the potential, which will be realized in the not-too-distant future, to screen discrete collections of a million or more single compounds, at multiple concentrations. This vast body of useful structure-activity relationship information will be made available at the nascence of a medicinal chemical effort.

Traditional medicinal chemical endeavors have involved the analysis of detailed biological data from hundreds or perhaps thousands of compounds. It is not surprising that the prospect of such an explosive growth of information both from screening and from program-directed combinatorial chemistry has driven the evolution of cheminformatics (20), in much the same way that genomic sequencing gave rise to the science of bioinformatics.

So what should the medicinal chemist expect from HTS when beginning a new research program early into the third millennium? In many cases, the molecule(s) of requisite potency needed for biological proof-of-concept studies will come directly from the HTS. As the HTS of compound collections becomes more dependent on combinatorial chemistry, the properties of lead structures may be enhanced rapidly using previously established high-throughput synthetic methods. The latter effort may be significantly aided by the incorporation of desirable developability characteristics (i.e. bioavailability, drug metabolism, and pharmacokinetics) into the design of libraries for lead generation, an area of further impact for the cheminformatics practitioners of the future (21).

MEDICINAL CHEMISTRY

During the past 50 years, the total synthesis of many exceedingly complex natural products has been achieved, representing a major accomplishment for organic chemistry (22). Although these efforts have been greatly facilitated by advances in the use of spectroscopic techniques, the underlying scientific principles are the same as those that guided synthetic organic chemical pioneers at the turn of this century. Organic synthesis is, and is likely to be for some time, the cornerstone upon which medicinal chemistry is built. However, as part of an exciting development still taking place as the millennium closes, these disciplines have witnessed a paradigm shift as great as any in their history, namely the introduction of parallel or combinatorial synthesis. For some time now, the synthesis of peptides and oligonucleotides has been conducted by automated methods, a paradigm rendered more accessible by the modular nature of these macromolecules. The revolution that has gripped medicinal chemistry over the past 5 years has brought a similar philosophy to the construction of more ‘‘drug-like’’ small-molecules, using both solid- and solution-phase methods. Although early attempts to automate solid-phase small-molecule synthesis used modified peptide synthesizers, such machines were less than optimal. However, custom designed devices are now available that handle both solution and solid-phase protocols. Further refinement of these devices will continue into the new millennium, and the use of parallel synthesis will likely become part of the armamentarium of every medicinal chemist. In addition to the introduction of combinatorial methods, the 1990s have witnessed further refinement of rational drug design based upon molecular modeling and the use of protein X-ray crystallography. In the future, the avail-

ability of structural information early in a research program will no doubt enhance its impact, especially with the further development of de novo–design molecular modeling protocols.

The positive impact that structural information has on a medicinal chemistry program is evident from many efforts in recent years, notably those directed toward inhibitors of HIV protease (23) and cathepsin K (24). Thus, an initial design hypothesis for cathepsin K inhibitors, based upon the X-ray cocrystallography of aldehyde-type structures bound to papain, led to the synthesis of potent and selective 1,3-bis(acylamino-2-propanone, compound 1 ($K_{i,\mathrm{app}} = 22$ nM) (Figure 1) (24). Subsequent elaboration of compound 1 based upon further cocrystallization experiments with cathepsin K itself, and molecular modeling, led to the synthesis of even more potent analogs, such as compound 2 ($K_{i,\mathrm{app}} = 1.4$ nM) (Figure 1) (25).

Given the importance of GPCRs as targets in the pharmaceutical industry, the recently disclosed X-ray crystal structure of the archetypal protein of this class, bacteriorhodopsin (26), assumes special significance. In the first decade of the next century, it is likely that the three-dimensional structure of therapeutically relevant GPCRs will become available, enhancing our understanding of these targets at a molecular level and opening the way for rational drug design.

The impact of nuclear magnetic resonance (NMR) spectroscopy on rational drug design has recently come to the fore with the description of the so-called structure-activity relationships by NMR technique (27). This and other NMR techniques (28) that probe molecular interactions will no doubt receive attention in the coming decade, providing a new technique for HTS. Furthermore, NMR structural studies may become applicable to larger proteins than heretofore examined through the development of recently described segmental labeling protocols (29).

The successful medicinal chemical drug discovery effort for the new millennium will rely on a hybrid approach of parallel and iterative (single-molecule) synthesis. As HTS collections are built up through parallel synthesis, lead structures will be amenable to high-throughput follow-up. Iterative analog preparation directed at specific questions, which will influence the design of the parallel analog syntheses, will, however, continue to play a key role in the medicinal chemical effort of the future. As a research program begins to define more clearly the chemical structure of a compound that is appropriate for drug development, a greater level of iterative synthesis will likely become necessary to fine-tune the molecule to enhance specific properties of the drug in order to make the compound more suitable for drug development (e.g. potency, selectivity, pharmacokinetic profile). Structure-based design will continue to positively impact medicinal chemical efforts. In this area, the ability of parallel synthesis to explore hypotheses with a greater number of analogs should offer a distinct advantage in overcoming inevitable uncertainties with even the best X-ray or NMR-derived models.

CHEMINFORMATICS

The ability of combinatorial synthetic methods to provide large numbers of compounds rapidly does not ensure that screening collections built up by this means make the most effective use of HTS resources. To maximize the opportunity for lead generation from a given compound collection, the constituents of the collection should be as diverse as possible. However, such comparisons are complicated by the fact that any measure of chemical diversity is dependent on the parameter being considered. Certain molecular features have been considered desirable in drug candidate molecules, such as molecular weights of <500, a limited number of H-bond acceptors (<10) and donors (<5), and a cLogP of <5, and these characteristics have been recognized in what have become known as the Lipinski ''rule of five'' (21). Notwithstanding a desire to populate screening databases with molecules embodying these favorable characteristics, attempts are being made to maximize diversity within these limits. The development of such diversity tools by cheminformatic scientists is still in its infancy, and their value will only really become apparent when chemical libraries thus designed have been screened against a battery of structurally diverse targets. Database mining of these screening data may then provide an understanding of which molecular characteristics are most important for a given target. In this manner, commonalties of molecular properties of compounds known to interact with a given protein family can be used to enrich databases with small molecules bearing these features, so-called biased libraries. Thus, the likelihood of finding a lead structure for a novel member of such a class of proteins may be enhanced. In essence, this approach, although more rigorous, is a development of the empirical observation that certain structural similarities exist among small-molecule ligands of GPCRs, the so-called permissive or promiscuous structures (8). Thus, at least two classes of endothelin receptor antagonists arose from the discovery of a lead structure that had been previously prepared as an antagonist of another GPCR, the angiotensin AT-1 receptor (30) (Figure 2).

Although many classes of receptors and enzymes have had small-molecule agonists/substrates and antagonists/inhibitors discovered for them, protein-protein interactions have proven to be more difficult to influence, an exception being certain integrins. This has been found to be the case despite a large number of such targets having been subjected to HTS (15). It is also true, however, in the case of the integrins, that structures synthesized for one target, such as GP II_b/III_a, have provided lead molecules for another member of the integrin family, $\alpha_v\beta_3$ (31). In the mid-1990s, a study of the complex between human growth hormone and the extracellular domain of the human growth hormone receptor highlighted, through site-directed mutagenesis, a relatively small area of interaction, or a ''hot spot,'' responsible for the more than three quarters of the binding free energy (32). Although this observation bodes well for the possible ability to identify an inhibitor of the protein-protein interaction with a small molecule, the

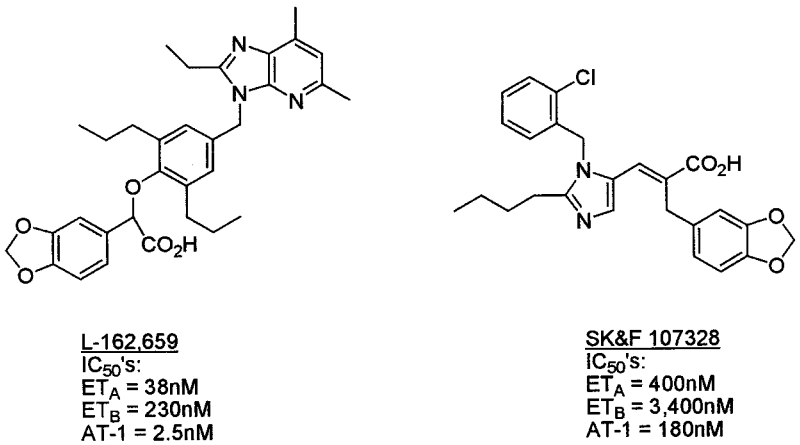

Figure 2 Two classes of endothelin receptor antagonists obtained by screening collections of compounds made for another G-protein–coupled seven-transmembrane receptor, angiotensin II AT-1. IC_{50}, 50% inhibitory concentration; ET_A, endothelin-A receptor subtype; ET_B, endothelin-B receptor subtype; SK&F, SmithKline and French.

generality of this paradigm remains uncertain. As structural information for protein-protein interactive targets becomes more plentiful, combined with site-directed mutagenesis data, the medicinal chemist of the future may be able to predict which of these targets are more likely to be susceptible to small-molecule therapeutic intervention. A further breakthrough in this area has been achieved recently with the discovery of a small-molecule agonist of the GCSF receptor, which acts by oligomerization of receptor proteins (33). As examples of small molecules influencing protein-protein interactions accumulate, cheminformatic analysis of the data, as described above, could expand the influence of medicinal chemistry within this important area.

FUNCTIONAL GENOMICS

The term functional genomics is now being used to describe the post–genome project era, which will begin early in the new millennium and will encompass the many efforts needed to elucidate gene function. Indeed, the phenotyping of genetically manipulated animals will be critical in the determination of biological function of a particular gene. But, in reality, the discipline of functional genomics has its foundation in the physiological and pharmacological sciences. This is gratifying to the ''traditional'' pharmacologist, whose expertise will be drawn on even more in the future to unravel the mysteries of genetics. Although the evaluation of genetically manipulated animals will require a thorough understanding of physiology and pharmacology, the experimental approach will involve many

new technologies. These methods will include in vivo imaging (i.e. magnetic resonance imaging, micro–positron emissions tomography, ultrafast computed tomography, infrared spectroscopy), mass spectrometry, and microarray hybridization, all of which should enhance the speed and accuracy at which functional genomics is achieved.

There are two general systematic approaches to the generation of mutations in the mammalian genome, one genotype driven and the other phenotype driven. Genotype-driven mutagenesis involves classic transgenic approaches whereby constructs introduced into the genome by pronuclear injection lead to insertional mutagenesis. Alternatively, homologous recombination in embryonic stem cells can be used to introduce new mutations into known genes. This approach is not easily scalable to the recovery of a large number of mutations on a genome-wide basis.

Because the phenotyping of mutant animals is usually driven by preconceived notions concerning the biological function of the manipulated gene, critical data can be overlooked. Thus, in the case of the endothelin-B receptor knockout mouse, anticipated to be hypertensive, more rigorous phenotyping implicated this receptor as an important sensory pain mediator (34).

Phenotype-driven approaches have focused on random mutagenesis systems to identify novel genes and pathways. One type of phenotype-driven mutagenesis employs the chemical mutagen, N-ethyl-N-nitrosourea. This potent mutagen can deliver mutation frequencies of approximately 1 in 1000 gametes and is therefore effective in accomplishing saturation mutagenesis of the entire mouse genome. Obviously, this approach draws heavily on the appropriate biological screens used to identify phenotypes of interest.

GENE THERAPY

Gene therapy as a therapeutic technique offers the possibility of introducing a functioning gene into somatic cells of a patient to correct a defective gene and thereby restore biological function. The major interest of the pharmaceutical industry in gene therapy will undoubtedly be centered around in vivo treatment protocols, although more invasive ex vivo methods (whereby cells are removed from the patient, transfected with the gene of interest, and then placed back into the patient) may be acceptable for certain serious diseases (e.g. cancer). Currently, genetic information can be transferred into cells by a number of protocols, including the use of DNA plasmids, DNA liposomes, or a variety of viruses. The most effective transforming agents are viral vectors, such as adenovirus, adeno-associated viruses, and retroviruses. Although retroviruses require cell division to incorporate the new information into the genome, adenovirus and adeno-associated viruses will transfer their information into nonreplicating cells. Despite considerable efforts that have gone into vector design during the past decade, delivery of genes still represents one of the major problems associated with the

gene therapeutic approach. Encouragingly, recent work has demonstrated the ability of certain viral vectors to stably incorporate genetic material in vivo, resulting in expression of the resultant proteins, which can be maintained for periods of several months (35).

The Food and Drug Administration has approved approximately 90 clinical studies involving gene therapy, which represents one of the fastest growing areas in biomedical research. Safety is, and will continue for some time to be, an issue with this approach, and many of the clinical trials involving gene therapy are directed toward patients with life-threatening diseases (e.g. cancer patients already receiving conventional therapy). Most reports to date have been phase I clinical studies confirming gene transfection and demonstrating safety, but evidence of efficacy has been anecdotal. Properly controlled phase II and phase III clinical trials involving gene therapy are yet to be done and will tax the current development capacity within the pharmaceutical industry.

Some published examples of clinical studies of gene therapy in the cardiovascular arena have appeared. For example, the intra-arterial gene transfer of a plasmid that encodes for vascular endothelial growth factor has been tested for its ability to increase coronary and peripheral angiogenesis (36, 37). The initial results from these early clinical trials suggest that vascular endothelial growth factor gene transfer produces angiogenesis and reduces ischemic symptoms. However, as mentioned above, these studies are inadequately controlled and the results can only be considered anecdotal.

Gene therapy has received much attention for the treatment of inherited metabolic diseases. One such disease in which preclinical data suggest promise targets familial hypercholesterolemia. Replacement of the low-density-lipoprotein receptor gene by adenovirus transfection into mice in which the gene for this receptor had been knocked out resulted in correction of the dyslipidemia (38).

PROTEIN THERAPEUTICS

The development of proteins as drugs has been the principal focus of the biotechnology industry as well as a component of the drug pipeline of several larger pharmaceutical companies for some time. With the anticipated disclosure of the blueprint of the human genome scheduled for the spring of 2000, further interest is likely to emerge in the area of novel proteins as drugs. Although the focus of the pharmaceutical industry is likely to remain principally on small-molecule agents, there have been some notable successes with protein agents (e.g. erythropoietin). Human genome sciences have already converted genomic information obtained from high-throughput sequencing into potential therapeutic proteins in clinical trials. Thus, myeloid progenitor inhibitory factor-1 is currently in phase II clinical trials as a stem cell protector in cancer therapy. In addition, keratinocyte growth factor-1 is in phase II clinical trials for wound healing.

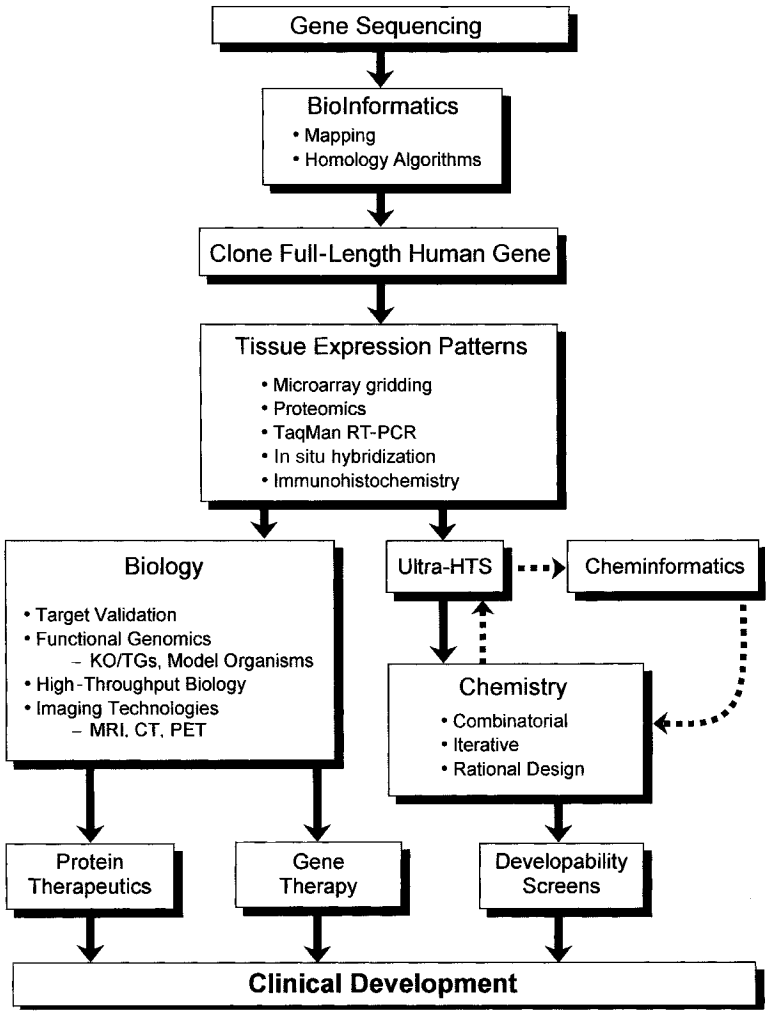

Figure 3 Progression of molecular targets to novel therapeutics under a new paradigm for drug discovery. HTS, High-throughput screening; RT-PCR, reverse transcriptase polymerase chain reaction; MRI, magnetic resonance imaging; CT, computed tomography; PET, positron emissions tomography.

CONCLUSIONS

The tremendous impact of genomic sequencing is currently being felt across all areas of drug discovery, and major challenges for the pharmaceutical industry into the next millennium will be in the areas of drug target selection and vali-

dation. The progression of new molecular targets into novel drugs under this new paradigm for drug discovery is shown in Figure 3. One can already anticipate the future availability of genetic structure and susceptibility to disease at the individual level. With such information available early in a research program, the drug discovery scientist is faced with the unprecedented opportunity to address the individual variability to drug therapy and safety prior to advancing a compound into clinical trials. The exponential growth of attractive novel molecular targets for potential drug therapy has heavily taxed the core disciples of drug discovery, and automated methods of compound synthesis and biological evaluation will play an even more dominant role in the pharmaceutical industry of the twenty-first century.

Visit the Annual Reviews home page at www.AnnualReviews.org.

LITERATURE CITED

1. Natl. Inst. Health. *The National Human Genome Research Institute.* http://www.nhgri.nih.gov.
2. Venkatesh TV, Bowen B, Lim HA. 1999. Bioinformatics, pharma and farmers. *Trends Biotechnol.* 17:85–88
3. Stadel JM, Wilson S, Bergsma D. 1997. Orphan G protein-coupled receptors: a neglected opportunity for pioneer drug discovery. *Trends Pharmacol. Sci.* 18: 430–37
4. Ames RS, Sarau HM, Chambers J, Willette RN, Aiyar N, et al. 1999. Human urotensin-II, the most potent vasoconstrictor identified, is a ligand at the novel receptor GPR-14. *Nature* 401:282–86
5. Sakurai T, Amemiya A, Ishii M, Matsuzaki I, Chemelli RM, et al. 1998. Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* 92:573–85
6. Zhang Y, Proenca R, Maffei M, Barone M, Leopold L, Friedman JM. 1994. Positional cloning of the mouse obese gene and its human homologue. *Nature* 372: 425–32
7. Lin L, Faraco J, Li R, Kadotani H, Rogers W, et al. 1999. The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* 98:365–76
8. Debouck C, Metcalf B. 2000. The impact of genomics on drug discovery. *Annu. Rev. Pharmacol. Toxicol.* 40:193–208
9. Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SA. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. USA* 91:5022–26
10. Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhard DJ. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet. Suppl.* 21:20–24
11. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. 1996 Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93: 10614–19
12. Blackstock WP, Weir MP. 1999. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol.* 17(3): 121–27
13. Wang H, Bloom O, Zhang M, Vishnubhakat JM, Ombrellino M, et al. 1999. HMG-1 as a late mediator of endotoxin lethality in mice. *Science* 9:248–51
14. Carr SA, Huddleston MJ, Annan RS. 1996. Selective detection and sequencing

of phosphopeptides at the femtomole level by mass spectrometry. *Anal. Biochem.* 1:239(2):180–92
15. Spencer RW. 1998. High-throughput screening of historic collections: observations on file size, biological targets, and file diversity. *Biotechnol. Bioeng.* 61(1):61–67
16. Snider RM, Constantine JW, Lowe JA III, Longo KP, Lebel WS, et al. 1991. A potent nonpeptide antagonist of the substance P (NK_1) receptor. *Science* 251: 435–37
17. Doherty AM, Patt WC, Edmunds JJ, Berryman KA, Reisdorph BR, et al. 1995. Discovery of a novel series of orally active non-peptide endothelin-A (ETA) receptor-selective antagonists. *J. Med. Chem.* 38:1259–63
18. White JR, Lee JM, Young PR, Hertzberg RP, Jurewicz AJ, et al. 1999. Identification of a potent, selective non-peptide CXCR2 antagonist that inhibits interleukin-8-induced neutrophil migration. *J. Biol. Chem.* 273:10095–98
19. Ohlstein EH, Nambi P, Douglas SA, Edwards RM, Gellai M, et al. 1994. SB 209670, a rationally designed potent nonpeptide endothelin receptor antagonist. *Proc. Natl. Acad. Sci. USA* 91:8052–56
20. Polinsky A. 1999. Combichem and cheminformatics. *Curr. Opin. Drug Discov. Dev.* 2(3):197–203
21. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 23:3–25
22. Service RF. 1999. Race for molecular summits. *Science* 285:184–85
23. Eyermann CJ, Jadhav PK, Hodge CN, Chang C-H, Rodgers JD, Lam PYS. 1997. The role of computer-aided and structure-based design techniques in the discovery and optimization of cyclic urea inhibitors of HIV protease. *Adv. Amino Acid Mimetics Pept.* 1:1–40
24. Yamashita DS, Smith WW, Zhao B, Janson CA, Tomaszek TA, et al. 1997. Structure and design of potent and selective cathepsin K inhibitors. *J. Am. Chem. Soc.* 119:11351–52
25. DesJarlais RL, Yamashita DS, Oh H-J, Uzinskas IN, Erhard KF, et al. 1998. Use of X-ray co-crystal structures and molecular modeling to design potent and selective non-peptide inhibitors of cathepsin K. *J. Am. Chem Soc.* 120:9114–18
26. Leucke H, Schobert B, Richter H-T, Cartailler J-P, Lanyi JK. 1999. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* 291(4):899–911
27. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. 1996. Discovering high-affinity ligands for proteins—SAR by NMR. *Science* 274:1531–34
28. Lin M, Shapiro MJ, Wareing JR. 1997. Diffusion-edited NMR—affinity NMR for direct observation of molecular interactions. *J. Am. Chem. Soc.* 119:5249–50
29. Yamazaki T, Otomo T, Oda N, Kyogoku Y, Uegaki K, et al. 1998. Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. Chem. Soc.* 120: 5591–91
30. Peishoff CE, Lago MA, Ohlstein EH, Elliott JD. 1995. Endothelin receptor antagonists. *Curr. Pharm. Design* 1:425–40
31. Keenan RM, Miller WH, Kwon C, Ali FE, Callahan JF, et al. 1997. Discovery of potent nonpeptide vitronectin receptor ($\alpha_v\beta_3$) antagonists. *J. Med. Chem.* 40: 2289–92
32. Clackson R, Wells JA. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science* 267:383–86
33. Tian S-S, Lamb P, King AC, Miller SG, Kessler L, et al. 1998. A small, nonpeptidyl mimic of granulocyte-colony-stimulating factor. *Science* 281:257–59
34. Griswold DE, Douglas SA, Martin LD, Davis L, Schultz LB, Ohlstein EH. 1999. Modulatory role of the endothelin B receptor in inflammatory pain and

cutaneous inflammation in gene target-disrupted knockout mice. *Mol. Pharmacol.* In press

35. Ye R, Rivera VM, Zoltick P, Cerasoli F Jr, Schnell MA, et al. 1999. Regulated delivery of therapeutic proteins after in vivo somatic cell gene transfer. *Science* 283:88–91
36. Losordo DW, Vale PR, Symes JF, Dunnington CH, Esakof DD, et al. 1998. Gene therapy for myocardial angiogenesis: initial clinical results with direct myocardial injunction of phVEGF165 as sole therapy for myocardial ischemia. *Circulation* 98:2800–4
37. Isner JM, Pieczek A, Schainfeld R, Clair R, Haley L, et al. 1996. Clinical evidence of angiogenesis after arterial gene transfer of phVEGF165 in patient with ischaemic limb. *Lancet* 348:370–74
38. Kozarsky KF, Jooss K, Donahee M, Strauff JF III, Wilson JM. 1996. Effective treatment of familial hypercholesterolaemia in the mouse model using adenovirus-mediated transfer of the VLDL receptor gene. *Nat. Genet.* 13:54–62