DRUG DISCOVERY

AND GENOMIC TECHNOLOGIES

This month, *BioTechniques* launches a new bimonthly section with an introductory column and articles that will focus on innovative developments in Drug Discovery. Our intent is to cover the full breadth of technologies that underlie the development of pharmaceutical products and therapeutic strategies. The section editor is Tomi K. Sawyer, Vice-President, Drug Discovery, ARIAD Pharmaceuticals, and the current President of the American Peptide Society.

Genes to Drugs

Tomi K. Sawyer

ARIAD Pharmaceuticals,
Cambridge, MA, USA

The year 2001 marks the dawn of the third millennium and a new epoch in the "genomic era". It is with great expectation that we integrate drug discovery and genomic technologies to advance our understanding of disease and our capability to create yet more effective and safe medicines. In retrospect, the past 50 years have witnessed tremendous achievements in chemistry, biology, and informatics that have accelerated the process of drug discovery. More recently, such examples include combinatorial chemistry, high-throughput biological screening, and structure-based drug design. Similarly, a plethora of genomics technologies is contributing to an emerging "genes to drugs" concept that will significantly impact the next generation of therapeutics, including small-molecule and peptide-based drugs, recombinant proteins, vaccines, monoclonal antibodies, antisense RNA, ribozymes, and DNA. Collaborative networks among pharmaceutical and biotechnology companies, academic and/or government research groups, are combining their strengths and resources toward an amalgamation of drug discovery and genomics technologies. In this first column for this new section of *BioTechniques* entitled "Drug Discovery and Genomic Technologies", I will touch upon a few points to highlight recent progress in unraveling the human genome, exploring the genetic basis of disease, and integrating genomic technologies and drug discovery. Future columns will provide more focused perspectives on the critical elements and scientific approaches that are key to bringing the "genes to drugs" concept into reality.

Unraveling the Human Genome

In 1990, the Human Genome Project (4) was launched in the United States as a multinational, 13-year effort to accomplish several goals, including the identification of an estimated 100 000 genes in human DNA and the determination of the sequence of the approximate 3 billion chemical base pairs that constitute human DNA. Additional objectives involve informatics (databases), technol-

Table 1. Some Microbial and Animal Genomic Sequence Determinations

Organism	Number of Genes	Genome Size (base pairs)
<i>Mycoplasma genitalium</i>	470	580 070
<i>Mycoplasma pneumoniae</i>	706	816 394
<i>Rickettsia prowazekii</i>	834	1 111 523
<i>Thermotoga maritima</i>	1014	1 860 725
<i>Helicobacter pylori</i>	1590	1 667 867
<i>Haemophilus influenza</i> Rd	1746	1 830 140
<i>Mycobacterium tuberculosis</i>	3974	4 411 529
<i>Bacillus subtilis</i>	4221	4 214 814
<i>Escherichia coli</i> K-12	4668	4 639 221
<i>Saccharomyces cerevisiae</i>	6526	12 147 823
<i>Caenorhabditis elegans</i>	19 099	97 000 000
<i>Drosophila melanogaster</i>	13 601	120 000 000
<i>Homo sapiens</i>	100 000 (estimated)	3 000 000 000 (estimated)

ogy transfer (e.g., to the private sector), and responsibility to address other issues (e.g., ethical, legal, and social) that may arise from this project. In 2000, it was acknowledged that the Human Genome Project and an independent effort at Celera Genomics, a private company, had accomplished to a significant measure the sequencing of the human genome. However, the complete annotation, sequence database editing for error corrections, and identification of genes and their functions remain to be determined. The latter point, indeed, provides the basis to fully understand the genetic basis of disease and a systematic approach to further identify the proteins involved (proteomics) and their overall manifestation in vivo (functional genomics). In other words, the best is yet to come as science, technology, and medicine forge ahead toward deciphering the human genome from the standpoint of chemical structure to that of biological activity. To date, a physical map of more than 30 000 human genes, including most genes that encode proteins of known function, has been determined (6) and provides a resource tool for analysis of complex genetic traits, positional cloning of disease genes, cross-referencing of mammalian genomes, and validating human-transcribed sequences for large-scale studies of gene expression. As related to informat-

Table 2. Transgenic Animal Models to Study the Genetic Relationship to Disease

Specific Gene Product	Gene Modification	Disease Phenotype
Src tyrosine kinase	deletion	osteopetrosis
cathepsin-K	deletion	osteopetrosis
Nf- κ B	deletion	osteopetrosis; impaired B cell function
osteoprotegerin	deletion	osteoporosis
α inhibin	deletion	gonadal stromal tumors
angiotensin-converting enzyme	deletion	hypotension
angiotensinogen	deletion	hypotension
PTP1B tyrosine phosphatase	deletion	diabetes; obesity
Zap70 tyrosine kinase	deletion	severe combined immunodeficiency
recombinant <i>bcr-abl</i>	addition	leukemia
transforming growth factor- α	addition	psoriasis
interleukin-5	addition	eosinophilia
melanocortin-4 receptor	mutation	obesity syndrome
leptin	mutation	obesity syndrome
leptin receptor	mutation	obesity syndrome

ics, accessibility to databases will be key to the success of exploiting the sequence of the human genome. Examples include genome mapping databases, nucleic acid and protein sequence databases, and nucleic acid and protein structural databases.

The complete sequences of several microbial and animal genomes, including the human genome, have been determined (Table 1). These efforts include major milestones such as the first complete genome of a free-living organism, *Haemophilus influenza* Rd, and that of the first eukaryote, the yeast *Saccharomyces cerevisiae*. With respect to drug discovery, it is thought that such microbial genomic information may lead to the identification of new targets, especially those that may not give rise to drug resistance, such as that which has emerged among many existing antimicrobial agents. The recent achievement in sequencing the complete genome of the nematode *Caenorhabditis elegans* has revealed more than 19 000 protein-coding genes, including a significant number of seven-transmembrane receptors, protein tyrosine and serine/threonine kinases, zinc fingers, RNA recognition motifs, protein tyrosine phosphatases, ion channels, and other intracellular signal transduction and gene-related proteins (7). Ultimately, the human genome will be mapped to determine the predicted 100 000 protein-coding genes and their correlation to disease states at the genetic level. Also, as relevant to drug discovery in terms of animal models of disease, the National Institutes of Health has recently launched a major research effort (Mouse Genome Sequencing Network) to decipher the mouse genome, which is anticipated to be completed by 2005.

Exploiting the Genetic Basis of Disease

There is no question that molecular biology has served to revolutionize the drug discovery process over the past two decades and continues to provide momentum for the development of a greater armamentarium of genomic technologies. Indeed, the application of molecular biology to identify drug targets, decipher complex disease mechanisms, and support biochemical screening and structural determination studies is particularly important. In determining the genetic basis of disease, some success has been achieved relative to target validation using transgenic animals involving deletion, addition, or mutation of specific genes to produce phenotypes that manifest a particular disease (Table 2). Gene deletions may be used to mimic the effects of a single gene defect or the physiological function of a specific gene product. In mice, for example, deletions of either c-Src, a tyrosine kinase, or cathepsin K, a cysteine protease, result in osteopetrosis (14,15), while the deletion of osteoprotegerin results in osteoporosis (2). Gene additions may be used to analyze the pathophysiological consequences of inappropriate protein expression. Such studies may utilize constructs carrying sequences of proto-oncogenes, growth factors, and cell-surface antigens that are under transcriptional control of heterologous promoters (10,12). For example, recombinant methods used to prepare a *bcr-abl* gene fusion in mice have been shown to mimic the Philadelphia chromosome translocation that results in leukemia (8). Finally, gene mutations may give rise to loss of function or, in some cases, constitutive activity. Several independent gene mutations in mice (each effecting loss of function) result in obesity syndrome. These include mutations in the melanocortin-4 receptor (9), leptin (22), and the leptin receptor (19).

DRUG DISCOVERY

AND GENOMIC TECHNOLOGIES

Comprehensive compilations of animal (e.g., mouse) gene knockout and mutation data exist, and such information provides insight toward transgenic animal models to study disease. Nevertheless, the correlation of genetic mutations to pathogenesis poses significant challenges to drug discovery, especially considering the approximately 5000 inherited diseases known in humans.

Integrating Drug Discovery and Genomic Technologies

The "genes to drugs" concept is fundamentally based on gene identification and target validation. Some major genomic technologies for gene identification include expressed sequence tag (EST) sequencing, secreted protein analysis, differential display, expression profiling, and positional cloning (20). These gene identification methods vary significantly in terms of throughput (genes/year) and relevance to disease. In general, the EST and secreted protein approaches rank among the most likely to identify therapeutic targets. More specifically, EST methods facilitate the ability to sift through relatively large numbers of novel gene sequences for subsequent selection of targets on the basis of homology to known classes of drug (e.g., G-protein coupled receptors, ion channels, secreted protein hormones, kinases, and proteases). With respect to target validation, key genomic tools include transgenic knockouts, ribozymes, antisense oligonucleotides, and DNA microarray screening. Transgenic knockouts include germ-line genomic deletion of a target gene (21), whereas conditional transgenic knockouts using inducible, gene-specific knockout provide an opportunity to eliminate a gene in a particular tissue rather than the entire organism (11). As previously mentioned, transgenic animals are useful tools for the identification of therapeutic targets, distinction of target isoforms, differentiation of signaling pathways, generation of disease animals, and toxicological testing (13). Ribozymes are RNAs that can catalytically cleave specific target sequences and may be used to functionally knock out genes both in vitro and in vivo (3). Antisense oligonucleotides provide the opportunity to reduce message levels for any gene (18). Antisense oligonucleotide technologies have also led to potential therapeutics (17). DNA microarray screening can be used to measure gene expression levels in a high-throughput manner for target identification and to monitor changes in gene expression affected by drug treatment (5). It is expected that genomic technologies will significantly impact drug discovery, and other methods to further enhance the development of novel therapeutics are most certainly forthcoming.

It is estimated that up to 10 000 targets for drug discovery may exist in humans. This will add significant scope to the current focus on approximately 500 targets that have been correlated with diverse diseases ranging from cancer to cardiovascular disease, osteoporosis, atherosclerosis, hematologic diseases, diabetes, obesity, and central nervous system diseases. Numerous strategic

alliances have emerged among pharmaceutical and biotechnology companies, academic and/or government research groups. The spectrum of such collaborations varies from large-scale gene sequencing and analysis to focused genomic research in one or more specific disease areas, including various applications of high-throughput biological screening, combinatorial chemistry, structure-based drug design, and other drug discovery technologies (1,16). Nevertheless, the ultimate task to identify the causative genes for complex polygenic diseases (e.g., diabetes, asthma, atherosclerosis, and Alzheimer's disease) poses an extraordinary challenge to genomic technologies and drug discovery.

In this "Drug Discovery and Genomic Technologies" section, we have included articles that highlight progress in cancer vaccines, recombinant vascular targeting agents, and cDNA microarrays. In future issues, the concepts and impact of specific aspects of drug discovery and genomic technologies, including combinatorial chemistry, informatics, genomics and proteomics, high-throughput screening, structural biology, molecular modeling, and the development of new therapeutics, will be addressed.

REFERENCES

1. Beeley, L.J., D.M. Duckworth, and C. Southan. 2000. The impact of genomics on drug discovery. *Prog. Med. Chem.* 37:1-43.
2. Bucay, N., I. Sarosi, C.R., Dunstan, S. Morony, J. Tarpley, C. Capparelli, S. Schully, H.L. Tan et al. 1998. Osteoprotegerin-deficient mice develop early onset osteoporosis and arterial calcification. *Genes Dev.* 12:1260-1268.
3. Cech, T. 1992. Ribozyme engineering. *Curr. Opin. Struct. Biol.* 2:605-609.
4. Collins, F.S., A. Patrinos, E. Jordan, C. Elke, G. Aravinda, R. Gesteland, L. Walters, and the members of the DOE and NIH planning groups. 1998. New goals for the U.S. Human Genome Project: 1998-2003. *Science* 282:682-689.
5. Debouck, C. and P.N. Goodfellow. 1999. DNA microarrays in drug discovery and development. *Nat. Genet.* 21:48-50.
6. Deloukas, P., G.D. Schuler, G. Gyapay, E.M. Beasley, C. Soderlund, H.L. Rodriguez-Tome, T.C. Matise et al. 1998. A physical map of 30,000 human genes. *Science* 282:744-746.
7. *C. elegans* Sequencing Consortium (for a complete authorship listings, see: http://www.genome.wustl.edu/gsc/C_elegans/ and http://www.sanger.ac.uk/Projects/C_elegans/). 1999. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012-2018.
8. Heisterkamp, N., G. Jenster, J. ten Hoeve, D. Zovich, P.K. Pantengale, and J. Groffen. 1990. Acute leukaemia in bcr/abl transgenic mice. *Nature* 344:251-253.
9. Huszar, D., C.A. Lynch, V. Fairchild-Huntress, J.H. Dunmore, Q. Fang, L.R. Berkemeier, W. Gu., R.A. Kesterson et al. 1997. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* 88:131-141.
10. Jaenisch, R. 1998. Transgenic animals. *Science* 240:1468-1474.
11. Kuhn, R., F. Schwenk, M. Aguet, and K. Rajewsky. 1995. Inducible gene targeting in mice. *Science* 269:1427.
12. Merlino, G.T. 1991. Transgenic animals in biomedical research. *FASEB J.* 5:2996-3001.
13. Rudolph, U. and H. Mohler. 1999. Transgenic and targeted mutant mice in drug discovery. *Curr. Opin. Drug Disc. Dev.* 2:134-141.
14. Saftig, P., E. Hunziker, O. Wehmer, S. Jones, A. Boyde, W. Rommerskirch, J.D. Moritz, P. Schu, and K. Von Figura. 1998. Impaired osteoclast bone resorption leads to osteoporosis in cathepsin-

- K deficient mice. *Proc. Natl. Acad. Sci. USA* 95:15453-15458.
15. **Soriano, P., C. Montgomery, R. Geske, and A. Bradley.** 1991. Targeted disruption of the c-src proto-oncogene leads to osteopetrosis in mice. *Cell* 64:693-702.
 16. **Veber, D.F., F.H. Drake, and M. Gowen.** 1997. The new partnership of genomics and chemistry for accelerated drug discovery. *Curr. Opin. Chem. Biol.* 1:151-156.
 17. **Wahlestedt, C. and L. Good.** 1999. Antisense oligonucleotides—the way forward. *Curr. Opin. Drug Disc. Dev.* 2:142-146.
 18. **Wagner, R.W., M.D. Matteucci, J.G. Lewis, A.J. Gutierrez, C. Moulds, and B.C. Froehler.** 1993. Antisense gene inhibition by oligonucleotides containing C-5 propyne pyrimidines. *Science* 260:1510-1513.
 19. **Wang, M.-Y., K. Koyama, M. Shimabukura, C.B. Newgard, and R.H. Unger.** 1998. OB-Rb gene transfer to leptin-resistant islets reverses diabetogenic phenotype. *Proc. Natl. Acad. Sci. USA* 95:714-718.
 20. **Wiley, S.R.** 1998. Genomics in the real world. *Curr. Pharm. Design* 4:417-422.
 21. **Zambrowicz, B.P., G.A. Friedrich, E.C. Buxton, S.L. Lilleberg, C. Person, and A.T. Sands.** 1998. Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* 392:608-611.
 22. **Zhang, Z., R. Proenca, M. Maffei, M. Barone, L. Leopold, and J.M. Friedman.** 1994. Positional cloning of the mouse obese gene and its human homolog. *Nature* 372:425-432.

Suggestions for contributions to the “Drug Discovery and Genomic Technologies” section are welcomed by its editor, Dr. Tomi K. Sawyer (tomi.sawyer@ariad.com)