Genome revolution: impact on drug discovery and development

In the past few years, we have seen an exponential increase in the rate at which genomic sequence information has become available from a variety of organisms, including humans. Such information is now influencing drug discovery, with knowledge of the members of extended families of human isoenzymes and receptor subtypes facilitating screening for novel mechanistic prototypes. As human genetics begins to reveal specific molecular mechanisms underlying multifactorial diseases (e.g. hypertension, asthma, diabetes and cancer), and transgenic technologies permit the ranking of individual genes as targets within relevant biochemical pathways, refinement of therapeutic strategies to tackle such traditionally recalcitrant chronic diseases may soon be possible. The human genome also holds the key to understanding the molecular mechanisms involved in drug disposition, metabolism and toxicity; information that may soon give the drug developer the ability to predict *in vivo* performance from *in vitro* data. The combination of such sequence-based knowledge with newly-developed methods of high speed synthesis and combinatorial chemistry promises a quantum leap in the productivity of drug discovery.

It is useful to put this explosion of genetic information in a chronological context. The first genome to be sequenced, in 1977, was that of the double-stranded bacteriophage ΦX174, consisting of 5,386 bp, arranged as nine genes encoding 10 proteins [Sanger, F. *et al.* (1977) *Nature* 265, 687–695]. In 1988, the genomic sequence of the double-stranded DNA herpes simplex virus (HSV-1), comprising 152,260 bp and containing 72 genes encoding 70 proteins, became available [McGeoch, D.J. *et al.* (1988) *J. Gen. Virol.* 69, 1531–1574]. The sequence of another herpes virus, cytomegalovirus (CMV), was reported in 1991. Both HSV-1 and CMV are human pathogens. Their genomes completely define the biochemical potential of the viruses, and the essentiality of specific genes to the viral life cycle indicates potential drug targets. This target discovery strategy has also been used in the search for anti-HIV and anti-HCV agents [Ratner, R. *et al.* (1985) *Nature* 313, 277–284; Houghton, M. *et al.* (1991) *Hepatology* 14, 381–388].

The complete sequence of the genome of the first free-living organism, the bacterium *Haemophilus influenzae* Rd, has recently been reported [Fleischmann, R.D. *et al.* (1995) *Science* 269, 496–512; Casari, G. *et al.* (1995) *Nature* 376, 647–648]. With 1,830,137 bp, this genome contains some 1,750 predicted open reading frames. Bioinformatic analysis has identified more than 1,007 genes with specific biochemical roles, many of which will represent potential drug or vaccine targets if present in pathogenic strains.

Although progress in fully sequencing eucaryotic genomes has appeared slower than that for viral and bacterial genomes, much of the key genetic information required to support drug discovery is already at our disposal. A consortium of 35 European laboratories published the first full sequence of a eukaryotic chromosome, chromosome III of *Saccharomyces cerevisiae* [Oliver, S.G. *et al.* (1992) *Nature* 357, 38–46], and this has been followed by the publication of the full sequences of a further 9 chromosomes: the complete yeast genome sequence, estimated to contain some 6,000 genes arranged on 16 chromosomes, will become available in 1996 [Williams, N. (1995) *Science* 268, 1560–1561]. This information, coupled with the biochemical identification of genes essential to the growth of fungi, is already leading to the identification of novel therapeutic targets for antifungal drug discovery.

Meanwhile, public and commercial initiatives to sequence most human expressed genes as expressed sequence tags (ESTs), have placed an inventory of over 216,000 sequences in the public domain, while the recent publication of an analysis of over 300,000 ESTs from The Institute for Genomic Research (TIGR) has provided a 'directory' for the human genome [Adams, M.D. *et al.* (1995) *Nature* 377 Suppl., 3–174]. The completion of the sequencing of the human genome, and with it the unambiguous identification of the genes (approximately 100,000) that define human biochemical potential, is predicted by the year 2000.

The challenge for the pharmaceutical industry now lies in refining this information and turning it to therapeutic advantage. The relative scale of this enterprise dictates a growing dependence on computational assessment and rapid functional analysis to identify key target genes, an area in which substantial progress is urgently needed. It is interesting to note that the pharmaceutical industry has to date produced drugs which target some 150 specific gene products [Prous, J.R. (1994) *The Year's Drug News*, p. 522]. This leads to the obvious question of how many of the estimated 100,000 human gene products are likely to represent productive targets for therapeutic intervention. Perhaps the answer represents the final frontier for pharmaceutical drug discovery?

The impact of genomics is not confined to leveraging small molecule drug discovery and development. It opens up radically new opportunities for biological intervention, with important roles for novel biological and synthetic therapeutic agents identified through understanding how the human genome functions in health and disease. The focus of the genome revolution is now moving from the current phase of *defining* new targets, to that of *designing* new therapeutic strategies and products, a trend that can only increase as nature's own design principles are revealed.

David Bailey and Martin Mackay
Discovery Biology, Pfizer Central Research, Sandwich, Kent, UK

Emerging molecular targets

New target for male contraceptive

The B isoform of *N*-acetyl-β-D-hexosaminidase is found exclusively on the acrosome of sperm and plays a critical role in fertilization by allowing the sperm to recognize, bind and penetrate the membrane of the oocyte. The A isozyme is found in all somatic cells, and its deficiency is the cause of Tay-Sachs disease. A competitive inhibitor that acts only on the B isozyme has

  1359-6446/96/$15.00