0273-2289/01/9601-3/0009/$11.00

Keynote

Functional Genomics and DNA Chips

Samir K. Brahmachari

Functional Genomic Unit, Center for Biochemical Technology, Mall Road, Delhi 110007, India

Genomics: A New Frontier

The impact of the information technology revolution in biologic research and recent developments in genomic technologies have opened up new horizons for biotechnology, which is entering a new era of "genomics" that promises to transform both biology and medicine. This new biology is expected to take us from curative and preventive medicine to an era of predictive medicine.

At last we seem to be succeeding in bringing together all that we have learned in chemistry, physics, and biology toward an understanding of what humans are. The human genome initiative is a worldwide research effort with the goal of understanding the hereditary instructions that make each of us unique. The goal of this effort is to find the location of the 100,000 or so human genes and to read the entire genetic script, all 3 billion bits of information, by the year 2005. This will likely provide an in-depth understanding of and possible treatment for more than 6000 genetic diseases that affect humans as well as genetic alterations that increase the risk of developing some common diseases. In addition, the variability in response to pathogenic organisms and the basis of neoplastic proliferation and human behavior are also likely to be unraveled, thus paving the path of an era of predictive medicine.

Research on recombinant DNA over the last two decades focused on understanding single gene function at a time and expression of human foreign genes in microorganisms. Genomics is a new way of looking at gene function in its natural context, and complete genome sequence provides a holistic view of the organism.

Genomics is likely to deliver a new generation of diagnostics and therapeutics for health care. The widespread use of DNA sequencing technology and the polymerase chain reaction have helped increase our understanding of the molecular basis of many genetic disorders as well as multifactorial diseases such as cancer, hypertension, asthma, and neurologic and psychiatric disorders. The Human Genome Project is moving 3 to 4 yr ahead of

time, and it is apparent that most of the 88,000 human genes will be completely sequenced along with large regions of noncoding sequence by next year. It is expected that by 2003, the complete human genome (3 billion nucleotides) will be sequenced. With the arrival of the new millennium, the first draft of the human genome sequence will become available. Simultaneously, one of the biggest challenges in the biologic sciences will emerge: unraveling the functional role of sequenced data. This vast amount of information will open up opportunities for discovering a relationship between gene and phenotypic variability in humans. This will also help us to unravel rules of the biologic world. It is expected that biotechnology of tomorrow will be dominated by the knowledge derived from genomics research. However, this will require determining all the variations in sequences in the genome for all populations (i.e., 15×10^{15} nucleotides).

Single Nucleotide Polymorphisms and DNA Chips

The new focus of human genomics will be single nucleotide polymorphism (SNP) studies. The initial results of the sequencing project have already indicated that 1 in 1500 nucleotides in humans is substantially polymorphic. SNPs can be found in both coding and noncoding regions of the genome. They may be upstream or downstream from a gene or in a gene itself. They might be associated with a phenotype (as in mutation), and some SNPs can be thought of as mutations because they may cause some biologic effect. However, not all polymorphisms are mutations. The identification of SNPs at roughly 100 kb or closer in the human genome and subsequent establishment of associations of specific SNPs with specific diseases will enable us to discover the molecular basis of many polygenic diseases such as cancer, hypertension, asthma, epilepsy, schizophrenia, and bipolar disorder. This will enable us to identify new molecular targets for many diseases—for which no drugs are available. Understanding the molecular basis of 6000 known genetic disorders will allow the development of future tools for gene therapy.

This is where the role of electronics and physics emerges. The detection of nucleotide variation by conventional sequencing will be replaced by the DNA chip, in which an array of oligonucleotides is fixed on a silicon/glass surface. An array is an orderly arrangement of samples that provides a medium for matching known and unknown DNA samples based on base-pairing rules and automating the process of identifying the unknowns. An array experiment can make use of common assay systems such as microplates or standard blotting membranes and can be created by hand or make use of robotics to deposit the sample. In general, arrays are described as macroarrays or microarrays, the difference being the size of the sample spots. Macroarrays contain sample spot sizes of about 300 μm or larger, and they can be easily imaged by existing gel and blot scanners. The sample spot sizes in microarrays are typically <200 μm in diameter, and these arrays usually contain thousands of spots. Microarrays require

specialized robotics and imaging equipment that generally are not commercially available as a complete system.

DNA microarrays, or DNA (gene) chips, are fabricated by high-speed robotics on glass or nylon substrates, for which probes with known identity are used to determine complementary binding allowing massive parallel gene expression and gene discovery studies. An experiment with a single gene chip can provide researchers information on thousands of genes simultaneously—a dramatic increase in throughput. The density of the microarray that will be required will be on the order of 4000×4000 in a $2\ \text{cm} \times 2\ \text{cm}$ surface. Ultimately, a multicolor fluorescent detection system and advanced scanners will be required to analyze such an array. The DNA chip technology is already under way in this direction.

Impact of Genomic Pharmaceutical Industry

The drugs of tomorrow are likely to be genotype specific, and, therefore, the focus should be on pharmacogenomics or pharmacogenetics. Understanding how drugs are metabolized and which gene products responsible for drug response are sensitive to genetic variation within a population will open up new frontiers in medicine. New drug trials will probably be undertaken based on population genetic data. The detailed information on individual genetic variation relevant to drug treatment could eliminate the use of ineffective or even dangerous treatments. Drugs that show adverse responses in one population may be useful to another population based on their genetic makeup. The implication of the knowledge of "genomics" to pharmaceutical industries is likely to be far reaching as more and more genome sequences of pathogenic organisms become available. Highly specific drugs without side effects are likely to be developed for infectious diseases caused by organisms such as *M. tuberculosis*, *H. pylori*, and *P. falseferum*, leading to the alleviation of suffering of millions from tuberculosis, stomach ulcers, and malaria.

Functional Genomics

DNA sequence data provided by genome projects have spawned the new field of functional genomics. This approach will yield exciting insights into the pathways to which specific genes belong and will provide clues to their role in health and diseases. It is estimated that there are between 70,000 and 100,000 genes in the mammalian genome. To turn this genetic "blueprint" into a functioning organism, each of these genes must be expressed in specific temporal and spatial contexts. Large-scale, high-throughput experimental methods require information processing and an analysis system to match. Software and database systems to design arrays, track materials, and collect, analyze, and interpret data from gene expression studies are still in their infancy. Among other things, such systems have to catalog the expression behavior of thousands of genes in a single

experiment using a DNA chip and, subsequently, make comparisons across tissues, developmental and pathologic states, or cellular perturbations. It is exciting to anticipate a time when data from thousands of gene expression experiments will be available for meta-analysis, which has the potential to balance out artifacts from many individual studies, thus leading to more subtle findings and robust results. Investment in this area of research by national laboratories and industrial research and development is essential, because this DNA chip technology will probably change the way molecular biologic research is done at present. In addition, computational genomics will provide several new molecular targets for new drug discovery. It would be important to develop computational methods to replace or reduce drug trials for toxicity testing. This is likely to be the Holy Grail of the drug discovery process in the post-genome-sequencing era.