Chapter 2 – OVERVIEW OF BIOTECHNOLOGIES

by

Paul L. Knechtges, Ph.D.
U.S. Army Center for Environmental Health Research, Fort Detrick, MD, U.S.A.

2.1 ABSTRACT

Relevant "–omics", informatics, and biosensor technologies are defined and briefly characterized in a summary of key tools of importance to laboratory investigations of toxic hazards.

2.2 SCOPE AND DEFINITION OF BIOTECHNOLOGY

The term biotechnology is used to describe practical applications of the life sciences, ranging from medicine and agriculture to bio-inspired materials. Although biotechnology has been around for centuries (e.g. wine and medicines), the biotechnology industry grew very rapidly during the 1990s. Part of the reason for this rapid growth was the integration of the life sciences with other enabling technologies such as computers and analytical chemistry by the pharmaceutical manufacturing industry. With the announcement in 1999 that the human genome had been sequenced, a new era of biotechnology known as "genomics" was ushered into research and development.

Given the broad scope of biotechnology, it is very difficult to define it in simple and inclusive terms. Nevertheless, under the auspices of the U.S. National Research Council, the Army Board on Science and Technology published a report that defined biotechnology as a technology with one or both of the following characteristics:

- It uses organisms, or tissues, cell, or molecular components derived from living things, to act on living things.
- It acts by intervening in the workings of cells or the molecular components of cells, including their genetic material [1].

A great number of applications are encompassed by this definition. For the purposes of this Task Group's report, the following section focuses on those biotechnologies judged to be most relevant in protecting NATO troops from toxic hazards during deployments.

2.3 KEY BIOTECHNOLOGIES

2.3.1 Genomics

The term genomics was coined to describe the scientific discipline interested in information about the entire complement of a cell's DNA, referred to as the genome. Prior to the era of genomics, DNA was sequenced on a relatively small scale, usually gene by gene. However, with the advent of modern sequencers and computer technology, institutions have undertaken the sequencing of the entire genomes of many types of organisms. As of January 2005, an estimated 1,251 genome sequencing projects were identified, with 250 genomes completely sequenced and published [2].

The widespread availability of DNA sequences in genomic databases led to additional terms such as ***functional genomics***, which refers to the study of gene functions. Other terms like transcriptomics, proteomics, and metabonomics/metabolomics have also emerged. These "-omics" technologies are described below.

The sequencing of human and animal genomes has permitted comparative study of DNA sequences using super computers and advanced computational algorithms. The results have revealed much about the similarity (i.e. homologies) of coding regions in the DNA (exons) between the different genomes and the total number of possible genes. For example, the genome of the simple nematode *Caenorhabditis elegans* has over 83% of its protein-encoded DNA conserved in the human genome [3]. Higher levels of homology to the human genome are observed with vertebrates, particularly mammals.

By using DNA sequence data in human genome databases, researchers are identifying sequence variations between groups of individuals, which may eventually help explain differences in resistance or susceptibility to genetic and environmentally-induced diseases. A particular emphasis has been placed on mapping variations, known as single nucleotide polymorphisms (SNPs) for use as genetic markers. Over 10 million SNPs are thought to be present in the approximately 3 billion sequences of the human genome. Many SNPs are inherited together in "blocks" that are called haplotypes; the number of haplotypes in the human genome is estimated at 300,000 or more. Compared with SNPs, haplotypes represent a more tractable number of genetic markers for identifying genes related to disease susceptibility or resistance. The human haplotype map was released in 2003 and is now publicly available [4, 5].

One of the greatest leaps forward in biotechnology is the development of gene microarrays, which permit highly multiplexed analysis of genes and gene transcripts (see Annex A and the next section). Several microarrays are already commercially available that can simultaneously determine over 100,000 SNPs in an individual. More densely packed SNP microarrays, and other technologies for measuring genetic variation among individuals, are likely to become available in the near future. These technologies represent unprecedented power to study human genetic variation, which could help identify individual resistance and susceptibility to diseases and responsiveness to medical treatment options.

2.3.2 Transcriptomics

The term ***transcriptomics*** refers to the global measurement and study of messenger RNA (mRNA) that has been transcribed from a cell's DNA. These pieces of mRNA can be translated into peptides or proteins, or used as guides for the synthesis of proteins by ribosomes. Other types of RNA perform other functions such as gene regulation and enzymatic activity. The amount of various mRNAs in the cell is indicative of its cell type and physiological state. The two most important biotechnologies which have enabled the discipline of transrciptomics are microarrays and real-time polymerase chain reaction (RT-PCR).

Microarrays for transcriptomics research are designed with either complementary DNA (cDNA) or oligonucleotide probes that hybridize with mRNA from cells. A variety of methods are used to attach these probes to the microarray, and the number of probes incorporated in a single microarray can number in the tens of thousands. The commercially-available human genome microarray, for example, can detect over 30,000 known transcripts [6]. The transcription products are fluorescently labelled, and a scanner is used to convert light intensities for each probe spot into numeric values for data analysis.

The great number of probes on a microarray requires specialized data analysis techniques, because even a fraction of one percent experimental error can result in thousands of false positives or negatives. Researchers

have published various methods for reducing the likelihood of a false positives or negatives, and new computational approaches are continually being developed. In addition, computer software is commercially available that can perform many of the calculations and data manipulation functions needed for microarray data analysis.

2.3.3 Proteomics

The term ***proteomics*** refers to the global or comprehensive study and analysis of an organism's proteins (its proteome). The proteins of an organism catalyze enzymatic reactions, provide cell structures, serve as cellular signals, and provide many other cellular and inter-cellular functions. The complexity of the proteome greatly exceeds that of the genome. More than 130 different covalent modifications of proteins are known and variable RNA processing events such as alternative splicing and termination increase the number of possible proteins expressed from a single gene.

Advances in proteomics have been driven by increasingly sensitive analytical instrumentation, particularly mass spectrometry, and by powerful computational methods. Microarrays are also emerging for proteomics applications, but they have not yet matched the practicality of genomic and transcriptomic microarray technology. Nevertheless, many biomedical industry analysts predict that proteomics will yield more practical applications such as drug targets and biomarkers. Consequently, ongoing research and development is aimed at increasing the throughput of proteomics methods.

2.3.4 Metabolomics/Metabonomics

Technologies such as nuclear magnetic resonance (NMR) and mass spectrometry (MS) permit the profiling of an entire set of cellular metabolites (the metabolome); this area of investigation is known as ***metabolomics***. The term ***metabonomics*** refers to the study of the complement of small molecule metabolites that change in response to some challenge. Since metabolites are the products and by-products of many biosynthetic and catabolic pathways, the applications of metabolomics include disease diagnosis and the identification of drugs or chemical exposures.

2.3.5 Bioinformatics and Systems Biology

The burgeoning growth of sequence data from genomics, transcriptomics, and proteomics has necessitated development of a new discipline called ***bioinformatics***. This discipline manages and analyzes biological data, including microarray data analysis, protein structure determination, hypothesis generation, and other data-intensive biological research. Bioinformaticists must be knowledgeable of biology as well as computer and software technologies. New and future generations of biologists are being trained to routinely use bioinformatic tools along with advanced laboratory equipment to take full advantage of advances in biotechnology.

Systems biology encompasses analyzing and modelling biological systems ranging from genes, proteins, or metabolites to an overall model of a disease state. The approach in systems biology is iterative and involves interaction between computer simulations (*in silico*) with actual *in vivo* and/or *in vitro* experiments. The results are increasingly accurate models of biological processes over time that can be used to make predictions of responses to therapeutic agents, environmental exposures, or other perturbations to a biological system. Both bioinformatics and analytical technologies such as microarrays are essential to systems biology and to the modelling of complex biomolecular networks.

2.3.6 Biologically-Based Sensors (Biosensors)

Biosensors consist of a biological component (biomolecule, cell, or tissue) and a signal transducer. Over the past decade, technologies have emerged that permit the arrangement of specific biomolecules on a microchip and the culture of cells and tissues within microelectrical mechanical systems (MEMS). These biological components can act as sensors when coupled with transducers that convert biological responses into electrical signals or other types of signals that can be measured (e.g. optical, vibration, acoustic). The culture of cells and tissues in a very small package has many advantages over relatively cumbersome *in vitro* methods found in laboratories. Uses for these biosensors include high throughput screening of pharmaceuticals and the possibility of field-portable sensors for chemical hazards or pathogens.

2.4 REFERENCES

[1] Committee on Opportunities in Biotechnology for Future Army Applications, Board on Army Science and Technology, Division on Engineering and Physical Sciences, National Research Council. *Opportunities in Biotechnology for Future Army Applications*. Washington, D.C.: National Academy Press, 2001.

[2] www.genomesonline.org – (Last accessed 31 January 2007).

[3] Lai, C. et al. Identification of novel human genes evolutionarily conserved in caenorhabditis elegans by comparative proteomics. Genome Res 2000; 10(5): 703-713.

[4] International HapMap Consortium. The international HapMap project. Nature 2003; 426(18/25):789-796.

[5] www.hapmap.org – (Last accessed 31 January 2007).

[6] Addonizio, M. The great chip chase. Scientist 2003; 17(17):49.