problem is that in much of the human DNA there are long repeated sequences which complicate alignment. Only about 5% of human DNA encodes genes. The coding fraction is much higher in bacteria. Because of the smaller genome size and fewer complications, such as repeated sequences, bacterial genomes are easier to sequence.

DNA sequencing is essentially a problem in information technology. Many feel that the current sequencing technology is comparable to computer technology in 1970. There are tremendous opportunities for engineers to design and construct devices to read and analyze genomic information much more rapidly and cheaply. The intrinsic scale of genomic information is the size of a nucleotide (about 0.34 nm). Recent advances in nanotechnology may allow us to generate such instruments. Indeed, engineers have contributed significantly to a "lab-on-a-chip" device to do DNA sequencing. This device begins to marry microtechnology and biotechnology.

Another technology that is having a large impact on genomics is *microarrays* for measuring which genes are being expressed by measuring the corresponding *m*-RNA levels. These microarrays are high-density oligonucleotide arrays. These oligonucleotides hybridize with the corresponding *m*-RNA. For a known gene or protein an oligonucleotide can be synthesized that binds the corresponding *m*-RNA. Using photolithography, the manufacturing of arrays containing 280,000 individual oligonucleotides on glass substrates of 1.64 cm^2 is now routine. Such arrays can simultaneously analyze expression from 6,000 to 10,000 genes. For bioprocesses the array can be used to determine which genes are up regulated or down regulated in response to a process change (e.g., temperature).

An increase in *m*-RNA levels does not immediately correspond to a change in protein level. Because different *m*-RNAs have different rates of degradation, efficiency of translation, and so on, changes in *m*-RNA expression may not translate directly into changes in the protein content of a cell or tissue. Since the proteins are the primary components responsible for biological activity, a knowledge of the protein content of a cell is highly valuable. Because proteins cannot be easily amplified like nucleic acids, measurements of all of the proteins in a cell (i.e., proteomics) is very difficult. Also proteins vary greatly in properties such as hydrophobicity. Currently there is no method that can measure all proteins in a cell. The best available technique is two-dimensional gel electrophoresis. This technique is derived from the SDS-PAGE technique described earlier in this chapter. For 2D gel electrophoresis two methods are combined. The first is to separate proteins in one direction (say the *x*-axis) according to their *isoelectric point*—the pH at which the protein has no net charge. Using a special set of buffers, a pH gradient can be established in a polyacrylamide gel. When the proteins are subjected to an electric field, they will move to the pH equal to their isoelectric point and remain there. After equilibration with an anionic surfactant, the proteins are then subjected to an electric field in the other direction (say the *y*-axis). The proteins will remain at the same pH but move down the *y*-axis according to their molecular weight. By separating protein mixtures using these two parameters (isoelectric point and molecular weight), a complex protein mixture can be resolved into separate spots. Often those "spots" can be removed and analyzed by tandem mass spectrometry to identify the protein's sequence.

The technique of 2-D gel electrophoresis is time consuming and expensive. Significant improvements will be necessary for routine analysis of cell and tissue protein content. Better gels or methods of protein separation are needed. Promising approaches