

While genetic information is linear and static, cellular systems are highly nonlinear, dynamic systems that respond to their environment and regulate gene expression. Over the next decades a focus of biochemical engineering, in conjunction with other disciplines, will be relating this linear sequence information to those nonlinear dynamical systems.

The role of the bioengineer is twofold. One role is as an enabler, by making better tools for rapid analysis of DNA sequences, of expression of *m*-RNAs, and of a cell's total proteins (*proteomics*). A second role is as an interpreter and organizer of genetic information; this role usually involves mathematical modeling.

8.5.1. Experimental Techniques

The primary tools of genomics are used for DNA sequencing, detecting which *m*-RNAs are expressed, and determining which proteins are present in a cell or tissue. The nucleotide sequence of DNA fragments can be determined on a sequencing gel. The key to this method is the use of dideoxynucleoside triphosphates, which are derivatives of the natural deoxyribonucleoside triphosphates and lack the hydroxyl (OH) group at the 3' position. For example, ddATP is the derivative dATP. If a strand of DNA is being replicated and if ddATP is inserted into the position normally occupied by dATP, replication is stopped. The OH group at the 3' position is essential for continued replication. The basic process for sequencing is shown in Fig. 8.9. It is easiest to imagine reactions in four separate tubes labelled A, T, C, and G. To each tube is added the DNA fragment to be sequenced, DNA polymerase, a stoichiometric excess of dATP, dTTP, dCTP, and dGTP, and an oligonucleotide primer for the DNA polymerase to use. The reaction is analogous to the first step in the PCR reaction. Only one strand of the DNA fragment to be sequenced is read. In addition to the above reactants, in each tube a small amount of ddNTP is added. For example, in tube A ddATP is added, in tube T, ddTP, and so on. Since the amount of ddNTP added is small in tube N, several reaction products are formed. In tube N, the first time dNTP or ddNTP must be added to the growing copy of the original DNA fragment, there is a high probability that dNTP will be added and the chain can be extended. However, there is also a finite probability (determined by the ratio of ddNTP to dNTP) that ddNTP will be added and the chain extension will be terminated. At the second position where N is required, either ddNTP or dNTP will be added. As above, some chains will be terminated, and others will continue to extend. This reaction continues, and tube N will generate fragments of different sizes, all of which end in the letter N. These fragments are separated by gel electrophoresis, and the sequence of the DNA fragment can be read directly from the gel (as shown in Fig. 8.9). This technique is limited to relatively short DNA fragments (a few hundred nucleotides).

For sequencing genomes this technique can be modified and automated. The basic approach to sequence a large genome is to cut it into millions of overlapping fragments of 2,000 to 10,000 base pairs in length. Each fragment is ligated into a plasmid, which is transformed into *E. coli*. This approach is termed a shotgun approach, and the *E. coli* form a living genomic library. These colonies are robotically picked, identified with a barcode, and placed in a 384-well plate. The amount of the cloned DNA fragment is amplified using PCR. From each end of the fragment 500 letters are replicated using the ddNTP