DNA Sequence Analysis

Barry Yedvobnick and Victoria Finnerty

Department of Biology, Emory University
Atlanta, Georgia 30322
(404) 727-4203/4219

During the past 20 years rapid progress has been made in DNA cloning and sequencing technologies. The goal of much of this technology is to provide tools to help us understand gene structure and regulation, and also to characterize the protein sequence encoded by genes. It is certain that more knowledge in these areas will provide insight into basic cellular processes and ultimately assist in the diagnosis and treatment of disease. Consequently, it is important to introduce these ideas early to undergraduates with interests in medicine as well as basic and biomedical research. At Emory we have developed an advanced laboratory course in Genetics to address this issue. One section of this course was designed to introduce students to computer-assisted analyses of DNA.

The careful analysis of large amounts of DNA sequence is now routinely performed with the assistance of computers. Many analyses cannot be performed in a reasonable time frame without such programs. Even a slow computer will perform analyses in a fraction of a second.

GENEPRO (Riverside Scientific, 206/842-9498) is a simple-to-use, yet sophisticated program that runs on a microcomputer. In our course students work with a *Drosophila* sequence. The sequence derives from a gene named "mastermind"; this gene is required for normal *Drosophila* nervous system development. In the lab each student is given an "unknown" sequence. Initially students learn to work with the DNA sequence and find restriction enzyme sites as well as areas of the sequence likely to encode proteins. The deduced protein sequence is then examined for particular structural features that may provide initial clues to function (i.e., hydrophobic regions that may span a membrane). Finally, students use the deduced amino acid sequence to search sequence databases for related proteins. All students score hits in the database and then use associated literature references to write a short paper. The following is an outline of the types of analyses that can be performed on this sequence:

DNA analysis: (1) restriction sites, (2) composition, (3) inverse, (4) translate, (5) open reading frame search, (6) codon usage, (7) highlight a sequence.

Protein analysis: (1) structural tests, (2) composition, (3) reverse translate, (4) highlight a sequence.

Homology analysis: (1) dot matrix, (2) homology search.