Components of the Protein Sequence Identification Service: A Demonstration

W.C. Barker, B.C. Orcutt, L.T. Hunt, and D.G. George

National Biomedical Research Foundation
Georgetown Univ. Med. Center, Washington, D.C. 20007

The National Biomedical Research Foundation operates a Protein Sequence Identification Resource that includes many tools and services to aid researchers in identifying unknown proteins, detecting distant relationships between proteins, detecting duplications within protein sequences, and in predicting of corresponding nucleotide coding sequences, features of secondary structure, and possible antigenic sites. Some of these tools are available on-line to subscribers to the Protein Sequence Database. Others are available through the Sequence Identification Service of the Atlas of Protein Sequence and Structure. Components that will be demonstrated on-line by telephone access to our computer at Georgetown University include databases of protein and nucleic acid sequences and sophisticated retrieval programs for preliminary analysis of sequence data. The programs contain powerful routines for sequence and text searching as well as for retrieval on the basis of calculated numerical properties of sequences (e.g., molecular weight, percent cysteine), taxonomic categories, and protein relationships. Off-line components include programs for detecting distant relationships between proteins, for detecting duplications in protein sequences, and for predicting secondary structure and possible antigenic sites.