

Computer Analysis of Amino Acid Sequences

The Case of Plant Virus Capsid Proteins

Eugene V. Koonin, Arcady R. Mushegian, and Valerian V. Dolja

1. Introduction

Genome sequences are of minimal use without an adequate interpretation of the sequences of putative protein products, which is only possible on the basis of detailed computer analysis. The approaches to amino acid sequence analysis can be roughly divided into those that explore intrinsic properties of proteins, such as hydrophathy, secondary structure, distribution of different types of amino acid sequences, and so on, and those that search for sequence similarity. Both approaches include numerous algorithms and computer programs. In this short chapter, we cannot describe all or even the most widely used and valuable of these methods. Instead, we present a minimal set of procedures that, in our experience, is useful in order to extract a substantial amount of information from an amino acid sequence in a relatively short time. For detailed descriptions of various computer methods for sequence analysis, the reader is referred to the recently published reviews and *Methods in Enzymology* collections (1–3).

Plant virus capsid proteins are the major focus of this book. Accordingly, we will use sequences of these proteins in all our examples. Because the amino acid sequences of capsid proteins are among the most variable in the virus genomes, they are a good test case to evaluate sensitivity of currently available computer methods. In the case of some other virus proteins, i.e., the enzymes that are generally better conserved, computer approaches can be even more rewarding (e.g., refs. 4–6).