Amino Acid Sequences in Proteins

One of the most important research activities today is the determination of amino acid sequences of proteins. The results are leading to improved biological classification and insight into the mechanisms of evolution. The results will provide new evidence concerning the compositions of the proteins of organisms that lived billions of years ago and will furnish valuable clues to the understanding of present-day metabolic diseases.

One of the most significant studies thus far made is a detailed examination of cytochrome c. Comparison of the amino acid sequences in variants of this protein obtained from such diverse forms as yeast and man showed that, in many positions in the molecules, the amino acids were identical. The extent of differences among the various specimens was related to differences in the forms studied. When specimens of cytochrome c from more animals have been examined and when other kinds of protein have likewise been studied, an important new dimension will be added to phylogeny.

Understanding of mechanisms of evolution has already been enhanced through sequence studies of the globins. In a lower vertebrate, the lamprey, hemoglobin consists of a single protein chain of molecular weight 17,000. Apparently in higher animals the gene for this protein was duplicated. The two genes evolved separately so that they could code for two different chains, α- and β-hemoglobins. Later the gene for the β-chain evolved further, so that, in addition to β-hemoglobin chains, γ- and δ-hemoglobin chains are produced.

Neurath, Walsh, and Winter*, in their studies of proteases, have found the same type of behavior, and they describe their view of this important phenomenon as follows: “The principal mechanism for increasing the size of the genome in the population is the process of gene duplication. . . . The utility of this process lies in the fact that the original function of the duplicated gene is preserved, thus freeing the new gene from selection pressure.” Comparison of amino acid sequences in proteins is a means for determining the extent to which this mechanism has been employed in evolution.

Another type of difference in proteins is introduced by single mutations. Already more than 22 variants of α-hemoglobin and 43 variants of β-hemoglobin have been found in humans;† One of these, a change from glutamic acid to valine in the 6-position of β-hemoglobin, results in sickle-cell anemia.

A host of additional proteins are being examined, and the number of sequences determined has been doubling each year. Techniques for the determination of sequence continue to improve. Especially notable is the Edman Sequenator.‡ In this device, phenylisothiocyanate reacts with the terminal α-amino group of a protein or peptide. The modified NH₂-terminal residue is then selectively removed, the chain being thus shortened by one unit. Repetitive cycles with identification of the removed amino acid derivative have established sequences of more than 40 amino acids.

Even with improved methods, the research will not be completed soon. There are countless organisms, and each contains unnumbered proteins. Further progress in this important field will depend on the quality of the intuitive judgment with which scientists select materials for study.

—PHILIP H. ABELSON

‡P. Edman and G. Begg, European J. Biochem. 1, 80 (1967).