COMPARISON OF POLYPEPTIDE SEQUENCES

THOMAS H. JUKES
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

Evolution is able to take place because the inherited information in living organisms is subject to change. Most of this information is constant from generation to generation, but changes that take place in a very small part of it are necessary for evolution. Hereditary information is stored in the base sequences of DNA. Molecular evolution is, therefore, concerned with alterations in DNA. Changes in DNA are of three general types: *point mutations*, in which one base pair is substituted for another; *recombination*, in which two double strands of DNA cross over, so that there is a lengthening of one double strand and a shortening of the other; and *duplication*, proliferation, or multiplication, in which the amount of DNA per cell becomes increased. Any change of these three types is evolutionary if it is predominantly adopted into the genome of a species.

Part of the DNA, probably only a small fraction, consists of base sequences that provide information for the synthesis of proteins, and it is through proteins that most of the phenotypic expression of hereditary information takes place. This paper is concerned with evolutionary changes in proteins, as detected by analyzing proteins and by comparing the proteins of different species of living organisms with each other. Differences between two similar proteins, such as the hemoglobins obtained from man and monkey, will represent evolutionary changes in DNA that have entered into the makeup of the species. The last-mentioned two types of change, recombination and duplication, are infrequently adopted; for example, there seem to have been only six evolutionary events of recombination in the hemoglobins in about 500 million years. In contrast, base replacements take place incessantly. These produce the "evolutionary clock" that ticks slowly in proteins, independent of speciation, generation time or gene duplication, but *not* independently of the type of protein as defined by its essential function. We shall illustrate this by comparing the amino acid sequences of polypeptide chains, with the use of the amino acid code. A family of proteins, the globins, that occurs in all vertebrates, will be used as illustrative examples.

Supported by a grant from the National Aeronautics and Space Administration NgR 05-003-020 "The Chemistry of Living Systems" to the University of California.