THEORETICAL FOUNDATIONS
OF PALEOGENETICS

RICHARD HOLMQUIST
SPACE SCIENCES LABORATORY
UNIVERSITY OF CALIFORNIA, BERKELEY

1. Introduction

This paper presents mathematical reflections that were completed during the Spring of 1966 (Holmquist [9]). In light of increasing knowledge and interest in the evolutionary significance of the primary amino acid or nucleotide base sequences of homologous (see Section 3.2 for definitions) proteins or DNA's, both within and among various phylogenetic species, these calculations are given here in the hope that they may prove useful to a wider audience. The mathematics, though straightforward, is complex. Therefore, a conscientious effort has been made to relate the mathematical equations to concrete physical phenomena so that the paper may be more readable to mathematicians for whom the historical jargon of molecular biology may be unfamiliar as well as to the biologists, biochemists, and anthropologists who may want to use the mathematics as a tool for interpreting their experimental data.

The problems in molecular evolution that are soluble by a study of protein sequences and homologies may for convenience be divided into three classes: (a) the construction of phylogenetic trees; (b) the deduction of the primary amino acid sequences of the common ancestral proteins at the branch points of the phylogenetic tree; and (c) the assignment of a time scale to each leg of the phylogenetic tree. Historically, three concepts have been extremely useful in solving these problems: amino acid differences between homologous proteins (Zuckerkandl and Pauling [28]; Needleman and Wunsch [18]), the minimum mutation distance between homologous proteins (Jukes [10]; Fitch and Margoliash [4], [5]), and the path of least information or maximum entropy (Reichert and Wong [21]) between two homologous proteins.

It has seemed plausible to attempt to correlate the amino acid or base differences between two homologous proteins or nucleic acids with a time of origin, measured from the present, of a "common ancestor" protein or nucleic acid which is homologous to both. In favorable cases a possible primary sequence for the common ancestral molecule can be deduced. Proceeding in this way, one can build up a biochemical tree of life which can be compared to those