TWO DIMENSIONAL ANALYSIS OF POLARITY CHANGES IN GLOBIN AND CYTOCHROME *C*

HELMUT VOGEL
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, MONTPELLIER

1. Introduction

The classification of amino acid substitutions between protein chains has led to considerable success especially in the construction of phylogenetic trees that correctly, and in complete independence of the paleontological record or morphological facts, retrace many aspects of evolution (for example, M. O. Dayhoff [3]). The changes of physical properties that accompany those substitutions have not been as thoroughly investigated, at least not in close statistical conjunction with the substitutions themselves. The following attempt may open some new perspectives in this direction.

2. Procedure

By "class of proteins" we mean a set of homologous chains, generally functionally defined, that have been completely sequenced and that can include or exclude the reconstructed nodes in a phylogenetic tree (examples: all globins; all cytochromes *c*).

By "group of proteins" we mean a certain subset of a class of proteins which may or may not correspond to functional or taxonomic characteristics (examples: the α globins; all monohemic globins, the cytochromes *c* of all birds).

A "combination of chains" is a nonordered pair of chains.

The "combination of two groups" is the set of all combinations of chains, one member of the pair stemming from one of the groups, the other member from the other group. If the two groups are identical, we speak of an "in-group combination," otherwise of an "out-group combination."

In the present context, every combination is characterized by a point with the coordinates: N the number of sites where the two chains have different amino acids; P the difference of total polarities of the two chains. The polarities p_i of the amino acids are adapted from Woese [5] (Table I). Every combination is, according to the definitions above, listed only once, the sign of P being determined by the arbitrary order in which the chains are numbered. This has been done for 39 globin chains (see Table II) and for 25 cytochrome *c* chains (Table