CONSERVATION OF PHYSICO-CHEMICAL AMINO ACID PROPERTIES DURING THE EVOLUTION OF PROTEINS

M. ANGELICA SOTO and JOSÉ TOHÁ C.

Laboratorio de Biofisica, Departamento de Fisica, Facultad de Ciencias Fisicas y Matemáticas, Universidad de Chile, Casilla 5487, Santiago, Chile.

(Received 27 May, 1982; in revised form 19 September, 1983)

Abstract. Based on a similarity ring constructed from a substitution probability matrix, we have analyzed the conservation of some amino acid properties in the evolution of proteins. Refractive index and bulkiness are highly conserved, hydrophobicity and polarity are fairly well conserved while optical rotation appears to be a less relevant property. On the other hand, the analysis of the correspondence between phenotype and genotype shows that the most frequent amino acid substitutions in proteins do not always correspond to the most feasible codon changes. The apparent disagreement between amino acid substitions in modern proteins and the primordial amino acid-codon assignment is discussed.

1. Introduction

Much effort has been spent in the analysis of the conditions that allow amino acid substitutions in homologous proteins of different species. A worthy contribution has been presented by Argyle (1980) who used an algorithm which compared all the observed amino acid substitutions in a group of contemporary proteins, following data published by Hasegawa and Yano (1975). The protein families used in those studies were: hemoglobin, cytochrome c, fibrinopeptide, immunoglobulin, lysozyme, toxin, insulin, proinsulin, growth hormone, virus coat protein, ferredoxin and calcitonin. The resulting arrangement was expressed as a symmetric matrix where those amino acids with higher substitution probabilities were neighbours in a ring. In this ring the amino acids became ordered following chemical similarities.

It is the aim of this paper to evaluate the role that has been played by certain groups of physico-chemical properties of free amino acids in the frequency of mutation during evolution. In addition, the contribution of the genetic code in these mutations is analyzed.

The results of this analysis emphasize the degree of conservation of some physico-chemical amino acid properties during the evolution of proteins and the relative discrepancy between the most frequent amino acid changes and the most feasible codon changes. This discrepancy can be explained by considering that the amino acid changes in proteins could be conditioned by parameters other than those involved in codon-amino-acid recognition.

2. Methods

The substitution probability matrix described by Argyle (1980) shows a distribution of

Origins of Life **13** (1983) 147-152.0302-1688/83/0132-0147$00.90.

the frequency of amino acid substitutions which diminishes progressively with the distance from the main diagonal. This matrix is constructed in such a way as to provide the minimum variance about the main diagonal.

Thus:

σ = standard deviation

$$\sigma^2 = \frac{\sum_{k=-9}^{10} k^2 D_k}{\sum_{k=-9}^{10} D_k}$$

$k \neq 0$

where

k = distance of each diagonal from the main diagonal
D_k = the sum of the elements of the k^{th} diagonal.

This σ-value represents the distribution of the values of probability of substitution of the amino acids in the matrix.

An ideal matrix (theoretical), with a limited minimum value for σ, can be calculated by ordering the elements of each column vertically so as to minimize the variance about the diagonal element in that column, independently of all the other columns.

Using the similarity ring, we constructed separately, matrices for each of the physico-chemical properties of the amino acids. In these matrices, the elements correspond to the similarity degree of the amino acid physico-chemical property considered. For every matrix, the σ-value obtained was compared with the respective ideal σ-value.

In addition, as the frequency of amino acid substitutions are codon-dependent, we constructed a matrix using the ring described by Argyle (1980), but in this case, the elements correspond to the respective probabilities of codon changes. The σ-value calculated for this matrix was also compared with the σ of an ideal matrix.

3. Results

The relevant physico-chemical properties of amino acids which we used in this work to characterize the degree of likeness of the molecules were: refractive index, bulkiness, hydrophobicity, polarity and optical rotation. The values for these amino acid properties are shown in Table I. Using these values, we constructed the matrices corresponding to the percentage of similarity of these properties, following the amino acid sequence given by Argyle (1980). For every property, the relevant ideal matrix was also constructed.

As an example, in Table II we show the matrix of similarity for the refractive index of the