CONSERVATION OF THE SECONDARY STRUCTURE OF PROTEIN DURING EVOLUTION AND THE ROLE OF THE GENETIC CODE

M. ÁNGELICA SOTO, AQUILES SEPÚLVEDA, and JOSÉ TOHÁ C.*

Biofísica, Departamento de Física, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Casilla 5487, Santiago, Chile

(Received 8 July, 1985)

Abstract. In this communication we demonstrate, in a group of modern proteins, following an algorithm described by Argyle (1980), that the ordination of the amino acids in terms of the most frequent substitutions agrees with the conservation of the \( \alpha \)-helix, \( \beta \)-sheet, and \( \beta \)-turn formation tendencies during evolution. The same correspondence has been demonstrated for the conservation of the physico-chemical properties in the amino acid substitutions. Both parameters are similar in showing higher correlation with the most frequent amino acid substitutions than with the feasibility of changes at the level of the respective codons.

Some kind of restrictions for the expression of the genomic changes, due to the conservation of the secondary structure of proteins and/or the physicochemical properties of the substituted amino acids, could account for the differences found between the distribution of the amino acid substitutions and the most probable codon changes.

1. Introduction

In previous works (Soto and Tohá, 1983; Pieber and Tohá, 1983), we analyzed in a group of contemporary families of proteins (Hemoglobin, Cytochrome C, Fibrinopeptide, Immunoglobulin, Lysozyme, Toxins, Insulin, Growth hormone, Virus Coat Protein, Ferredoxin, and Calcitonin; Hasegawa and Yano, 1975), the conservation of some relevant physico-chemical properties during evolution. We used in that study an algorithm described by Argyle (1980) for the comparison of the frequencies of amino acid substitutions. The values of the frequencies of these substitutions were expressed in a symmetrical matrix, in such a way that those amino acids with higher substitution probabilities become neighbours in a similarity ring, in which chemically similar amino acids appeared grouped together.

We found that this amino acid ring corresponds in a high proportion (over 90\%) to the conversation of physico-chemical properties such as: Refractivity, Bulkiness, Optical Rotation, Hydrophobicity and Polarity.

On the other hand, there is only partial agreement (68\%) in the distribution of values in the matrix of most frequent amino acid substitutions, when compared with the ideal distribution of these changes. This disagreement could be explained, at least in part, because do not always the most frequent amino acid substitutions correspond to the

* This work has been partially supported by Departamento de Investigación y Bibliotecas, Universidad de Chile y Fondo Nacional de Investigación Científica y Tecnológica.
most feasible codon changes. When we compared the distribution of the most feasible
codon changes in a matrix that follows the order of the amino acid ring of the most
frequent substitutions against an ideal codon matrix, we found an agreement of only
63%, a figure smaller than the 68% agreement registered for the amino acid
substitution matrix when compared with its own ideal distribution.

The absence of correspondence between the values of the two matrices could be
explained, at least partially, by the presence of a restrictive control for the expression of
those amino acids that being dissimilar are, however, specified by similar codons. This
limitation could be imposed by restrictions due to the secondary or tertiary conserva-
tive structure of the proteins.

It is the aim of this communication to analyze the role of the conservation of the α-
helix, β-sheet, β-turn amino acid tendencies (Chou and Fasman, 1978; Ptitsyn and
Finkelstein, 1983) in amino acid substitutions and, moreover, to evaluate the degree of
correspondence of the respective codon probability of change and the structural
protein conservation maintained through the amino acid substitutions.

2. Methods

The original communication (Argyle, 1980) describes an amino acid substitution
matrix of a group of modern proteins in which the value of the frequency of amino acid
substitutions decreases progressively with the distance to the main matrix diagonal.
This matrix was constructed in such a way as to provide a minimum variance (σ^2) about
the main diagonal, which determines a singular amino acid ring.

Thus:

\[ \sigma^2 = \frac{\sum_{k=-9}^{10} k^2 D_k}{\sum_{k=-9}^{10} D_k} : k \times 0 \]

where \( k \) is the distance of each diagonal from the main diagonal and \( D_k \) is the sum of the
elements of the \( k \)th diagonal.

The \( \sigma \) value represents the distribution of the values of substitution probabilities of
the amino acids in the matrix.

The ideal matrix (theoretical) used as a reference in the comparisons gives a
minimum value for \( \sigma \), because it is constructed by ordering the elements of each matrix
column vertically so as to minimize the variance about the diagonal element in that
column, independently of all the other columns.

Conserving the order of the amino acids in the similarity amino acid ring, found as
the solution in the minimization of the variance of the substitution amino acid matrix,
we constructed new matrices in which the elements were, respectively, the degree of
similarity for the amino acid tendency to be found in an α-helix, β-sheet, or β-turn
formation.

This degree of similarity has been calculated using the data on conformational