CODE DEPENDENT CONSERVATION OF THE PHYSICO-CHEMICAL PROPERTIES IN AMINO ACID SUBSTITUTIONS

MARÍA PIEBER and JOSÉ TOHÁ C. **

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

(Received 5 January; in revised form 24 May, 1983)

Abstract. The frequency of amino acid replacements in families of typical proteins has been elegantly analyzed by Argyle (1980) showing that the most frequent replacements involve a conservation of the amino acid chemical properties. The cyclic arrangement of the twenty amino acids resulting from the most frequent replacements has been described as an amino acid chemical ring.

In this work, a novel amino acid replacement frequency ring is proposed, for which a conservation of over 90% of the most general physico-chemical properties can be deduced.

The amino acid chemical similarity ring is also analyzed in terms of the genetic code base probability changes, showing that the discrepancy that exists between the standard deviation value of the amino acid replacement frequency matrix and its respective ideal value is almost equal to that deduced from the corresponding base codon replacement probability matrices. These differences are finally evaluated and discussed in terms of the restrictions imposed by the structure of the genetic code and the physico-chemical dissimilarities between some codons of amino acids which are chemically similar.

1. Introduction

The amino acid sequences of typical proteins have changed slowly during evolution, as can be observed from the analysis of sequence data. Based on the amino acid replacement frequency in families of modern proteins, Argyle (1980) has constructed an amino acid transition probability matrix. The amino acid arrangement of this matrix which contains the replacement frequencies as matrix elements, was deduced by rearranging these elements in such a way as to provide the minimum variance about the main diagonal. In this amino acid arrangement the last amino acid is next to the first, generating the so-called amino acid ring. In this ring, amino acids which are chemically similar appeared grouped together. The amino acid sequence proposed by Argyle is Asp-Asn-Gly-Ser-Ala-Thr-Pro-Cys-Val-Ile-Leu-Met-Phen-Trp-Tyr-Arg-His_Gln_Lys-Glu, where Glu, is next to Asp generating the amino acid ring. As pointed out by Argyle, nearly 56% of all replacements for an amino acid involved its nearest or second nearest neighbor in the ring.

Recently, it has been evaluated the amino acid chemical similarities for that ring in terms of some physico-chemical properties (Soto and Tohá, 1983). In this amino acid arrangement it was found that the standard deviation, $\sigma_0$, for the refractivity, bulkiness and hydrophobicity agree with the $\sigma$ value of the corresponding ideal distributions in a 93%, 92% and 83% respectively, $\sigma_i/\sigma_0 \times 100$. However, polarity and optical rotation appeared less conserved, as the $\sigma_i/\sigma_0$ values were found to be 77% and 59% respectively.

** This work was partially supported by OEA and Departamento de Desarrollo de la Investigación.

Origins of Life 13 (1983) 139-146.0302-1688/83/0132-0139$\$01.35.
In this work, a new amino acid ring is proposed, for which a higher amino acid chemical similarity can be deduced.

Furthermore, the amino acid replacement frequency matrix is analyzed in terms of the corresponding codon replacement matrix. This would allow the assessment of the possible interrelationship between the observed most frequent amino acid replacements and the expected most probable base codon exchanges that can be inferred from base transition transversion probability considerations. The expected most probable codon exchanges where considered as those which involve the fewest number of base exchanges and the fewest number of base transversions. The similarity of the standard deviation value of this codon matrix with respect to that of the amino acid replacement frequency matrix should give a measure of the amino acid codon dependent exchanges.

Finally, the most frequent amino acid codon dependent replacements are analyzed in terms of the restrictions imposed by the genetic code structure and the larger physicochemical discrepancy that exist between some codons corresponding to amino acids which are chemically similar, i.e. codons related by one transversion or more than one base change, rather than only one base transition.

2. Methods

A. In all cases, except that of Table I, the variance of the matrices were minimized about the main diagonal and was calculated according to the described equation (Argyle, 1980)

\[
\sigma^2 = \frac{\sum_{k=-9}^{10} k^2 D_k}{\sum_{k=-9}^{10} D_k} \quad \text{for} \quad k \neq 0,
\]

where

\[\sigma^2 = \text{variance and } \sigma = \text{standard deviation.}\]
\[k = \text{the distance of the lesser diagonal from the main diagonal (main diagonal } k = 0).\]
\[D_k = \text{the sum of the matrix elements } F_{ij} \text{ in the } k\text{-th diagonal.}\]
\[D_k = \sum_{j=1}^{20} F_{n+1,j}\]

where

\[n = (j + k - 1) \mod 20.\]

B. In the case of \(\sigma\) values for matrices of Table I (where the matrix elements correspond to differences in chemical indexes), the variance appeared minimized about \(k = 10\) diagonal and not about the main diagonal along which the smallest values are grouped.