THE β-SHEETS OF PROTEINS, THE BIOSYNTHETIC RELATIONSHIPS BETWEEN AMINO ACIDS, AND THE ORIGIN OF THE GENETIC CODE

MASSIMO DI GIULIO
International Institute of Genetics and Biophysics, CNR, Via G. Marconi 10, 80125 Naples, Napoli, Italy

Received 12 Feb. 1996

Abstract. Two forces are generally hypothesised as being responsible for conditioning the origin of the organization of the genetic code: the physicochemical properties of amino acids and their biosynthetic relationships (relationships between precursor and product amino acids). If we assume that the biosynthetic relationships between amino acids were fundamental in defining the genetic code, then it is reasonable to expect that the distribution of physicochemical properties among the amino acids in precursor-product relationships cannot be random but must, rather, be affected by some selective constraints imposed by the structure of primitive proteins. Analysis shows that measurements representing the 'size' of amino acids, e.g. bulkiness, are specifically associated to the pairs of amino acids in precursor-product relationships. However, the size of amino acids cannot have been selected per se but, rather, because it reflects the β-sheets of proteins which are, therefore, identified as the main adaptive theme promoting the origin of genetic code organization. Whereas there are no traces of the α-helix in the genetic code table.

The above considerations make it necessary to re-examine the relationship linking the hydrophilicity of the dinucleoside monophosphates of anticodons and the polarity and bulkiness of amino acids. It can be concluded that this relationship seems to be meaningful only between the hydrophilicity of anticodons and the polarity of amino acids. The latter relationship is supposed to have been operative on hairpin structures, ancestors of the tRNA molecule. Moreover, it is on these very structures that the biosynthetic links between precursor and product amino acids might have been achieved, and the interaction between the hydrophilicity of anticodons and the polarity of amino acids might have had a role in the concession of codons (anticodons) from precursors to products.

1. Introduction and Hypothesis

Numerous observations point to the relationship between the physicochemical properties of amino acids and the organization of the genetic code (Pelc, 1965; Woese *et al.*, 1966; Epstein, 1966; Goldberg and Wittes, 1966; Volkenstein, 1966; Alff-Steinberger, 1969; Nagyvary and Fendler, 1974; Nelsesteun, 1978; Weber and Lacey, 1978; Jungck, 1978; Wetzel, 1978; Wolfenden *et al.*, 1979; Jurka *et al.*, 1982; Lacey and Mullins, 1983; Swanson, 1984; Sjostrom and Wold, 1985; Taylor and Coates, 1989; Di Giulio, 1989a, b; Di Giulio, 1991; Haig and Hurst, 1991; Lacey *et al.*, 1992; Szathmary and Zintzaras, 1992; Di Giulio, 1992; Siemion and Stefanowicz, 1992; Szathmary, 1993; Goldman, 1993; Baumann and Oro, 1993; Lacey *et al.*, 1993; Di Giulio, 1994a; Di Giulio *et al.*, 1994; Talstrup *et al.*, 1994). These relationships provide evidence in favour of hypotheses on the origin of the genetic code, such as the physicochemical hypothesis (Sonneborn, 1965; Woese *et*

Origins of Life and Evolution of the Biosphere **26:** 589–609, 1996.

al., 1966) which suggests that the origin of genetic code organization was determined by selective pressure tending to reduce the deleterious effects of mutations (Sonneborn, 1965) or translation errors (Woese *et al.*, 1966). Another hypothesis supported by the relationship between the physicochemical properties of amino acids and genetic code organization is the ambiguity reduction hypothesis (Woese, 1965; Fitch, 1966; Fitch and Upper, 1987). The latter hypothesis suggests that groups of similar codons were initially assigned to groups composed of structurally similar amino acids, and that the genetic code reached its current organization as a result of a selective pressure tending to reduce the codifying ambiguity existing within and between groups of amino acids (Woese, 1965; Fitch, 1966; Fitch and Upper, 1987). Whereas Lacey and Mullins (1983) discuss the correlation between the physicochemical properties of amino acids and the properties of anticodonic nucleotides, thus favouring the anticodon hypothesis (Lacey *et al.*, 1992). Furthermore, several models have been proposed to establish a stereochemical relationship between an amino acid and its corresponding codons or anticodons (Gamow, 1954; Dunnill, 1966; Pelc and Welton, 1966; Melcher, 1974; Nelsestuen, 1978; Shimizu, 1982; Balasubramanian *et al.*, 1980; Hendry *et al.*, 1981; Yarus, 1991). The latter models are also favoured by the correlations between the physicochemical properties of amino acids and those of codons or anticodons (Pelc and Welton, 1966; Nagyvary and Fendler, 1974; Weber and Lacey, 1978; Jungck, 1978; Lacey and Mullins, 1983).

In short, all the above hypotheses suggest that the physicochemical properties of amino acids must be reflected in the organization of the genetic code which, as indicated in the bibliographic references cited at the beginning of this section, seems to be widely accepted.

On the other hand, there are several observations and suggestions (Pelc, 1965; Jukes, 1966; Dillon, 1973; Wong, 1975, 1976; Wong and Bronskill, 1979; Wong, 1980, 1981; McClendon, 1986, 1987; Jurka and Smith, 1987a, b; Wachtershauser, 1988; Miseta, 1989; Taylor and Coates, 1989; Danchin, 1989; Szathmary and Zintzaras, 1992; Morowitz, 1992; Di Giulio, 1992a; Szathmary, 1993; Di Giulio, 1993, 1994a, 1995a;) that favour the hypothesis of a coevolution of the biosynthetic relationships between amino acids and genetic code organization (Wong, 1975). In other words, the structuring of the genetic code can, according to this hypothesis (Wong, 1975, 1976), be primarily attributed to the imprint of the prebiotic pathways of amino acid formation on the organization of the genetic code.

The coevolution hypothesis (Wong, 1975) sees the introduction of new amino acids into proteins, and thus the improvement of their performance, as the predominant selective advantage favouring genetic code origin (Wong, 1976, 1981) and, moreover, attributes the physicochemical properties of amino acids with only a subsidiary role in determining the allocation of amino acids in the genetic code (Wong, 1980). Therefore, a conflict apparently seems to exist between the physicochemical postulates (Woese, 1965; Sonneborn, 1965; Woese *et al.*, 1966; Fitch, 1966; Lacey and Mullins, 1983; Fitch and Upper, 1987) and the coevolution hypothesis (Wong,

1975, 1988) on the role that the physicochemical properties of amino acids played in the organization of the genetic code: a role that must have been fundamental for the physicochemical postulates and subsidiary for the coevolution hypothesis (Wong, 1980).

However, it can be reasonably hypothesized that, if the biosynthetic relationships between amino acids (relationships between a precursor and a product amino acid) were so important in defining the genetic code (Wong, 1975), then it is to be expected that the distribution of the physicochemical properties among amino acids in precursor-product relationships cannot have been random but must have been affected by some selective constraints imposed by the structure of primitive proteins. According to the genetic code coevolution hypothesis, the origin of the code is equivalent to the origin of the first messenger RNAs which, by definition, must have been translated into proteins that must have necessarily possessed certain structural themes, such as α- helix, β-sheets, which, in turn, must have reflected some physicochemical properties of amino acids. Consequently, the pairs of amino acids in precursor-product relationships (Wong, 1975) must reflect some physicochemical properties if these pairs were selected because of their capability to give rise to structural elements characterizing primitive proteins, as would seem logical to surmise. More generally, if the biosynthetic relationships between amino acids had an important role to play in forging the genetic code (Pelc, 1965; Dillon, 1973; Wong, 1975; Taylor and Coates, 1989; Miseta, 1989), then it can be hypothesized that the primary structures of primitive proteins displayed a certain 'colinearity' with the biosynthetic transformations and flows between amino acids; furthermore, these primary structures, and therefore also the secondary ones, must have necessarily reflected some physicochemical properties of amino acids, sharing them with the biosynthetic relationships between amino acids.

The above considerations led me to analyze the way in which the physicochemical properties of amino acids are distributed among the pairs of amino acids in precursor-product relationships and among the ones that do not have such a relationship but are, nevertheless, defined in the genetic code table.

2. Materials and Methods

Table I shows the 32 physicochemical properties of amino acids or scales that should in any case reflect the properties of amino acids, along with 3 measurements of dinucleoside monophosphates. All these scales were subjected to the statistical analysis described below.

A certain property was taken (Table I) and its values were arranged in increasing order of magnitude. Each value was then attributed with its specific rank. If two or more amino acids were found to have the same value for the physicochemical property then they were attributed with the mean rank. Then all the possible (190) absolute values of differences was calculated for the twenty ranks. This matrix of

TABLE I

The physicochemical properties of the amino acids and dinucleoside monophosphates used in the analysis are shown and the determination procedure or a short description is specified.

Property	Determination procedure or description	Data sources
A. Amino acids (a.a.)		
Scales of size		
Molecular weight	Handbook value	Numerous sources
Molecular volume	Residue volume minus the constant peptide volume	Jungck, 1978
Refractivity	Refractivity scale based on Gly as the zero point	Jungck, 1978
Bulkiness	Ratio of the side chain volume to length	Zimmerman *et al.*, 1968
Specific volume	Residue partial molal volumes	Jungck, 1978
van der Waals volume of residue	The volumes computed according to a simple additivity rule using the additivity terms calculated by Richards	Liquori and Sadun, 1978
Scales of hydrophobicity-hydrophilicity		
Polarity Woese *et al.*, 1966	Slope of the straight line resulting when log R_F for that a.a. is plotted against mole fraction of H_2O in the pyridine-H_2O solvent employed in paper chromatography	Woese *et al.*, 1966
Polarity Zimmerman *et al.*, 1968	48 + dipole moment for ionized side chains, dipole moment otherwise	Jungck, 1978
Polarity Grantham, 1974	Woese *et al.*'s value averaged with R_F in another system	Jungck, 1978
Hydrophilicity Weber and Lacey, 1978	R_F of a.a. in a high salt solvent for chromatography	Weber and Lacey, 1978
Hydrophobicity Zimmerman *et al.*, 1968	Index based on solubility data of Tanford	Zimmerman *et al.*, 1968
Hydrophobicity Bull and Breese, 1974	Measures the effect of each a.a. on the surface tension of water, and considers the slope of the surface tension relative to the concentration of the a.a.	Cornette *et al.*, 1987
Aboderin, 1971	Measures the mobilities of the a.a. using the monophasic apolar solvent system ethyl acetate/pyridine/water (8:2:1, by vol.)	Cornette *et al.*, 1987
Egelman, Goldman and Steitz, 1986	Scale of transfer free energies of the a.a.	Tolstrup *et al.*, 1994
Kyte and Doolittle, 1982	Combination of three scales	Kyte and Doolittle, 1982
Olsen, 1980	The average internal preference of the a.a. in proteins	Cornette *et al.*, 1987

TABLE I
Continued.

Property	Determination procedure or description	Data sources
Frommel, 1984	The apolar accessible surface area of a side-chain is the total accessible surface area of the side-chain minus a constant	Cornette *et al.*, 1987
Chothia, 1976	For each a.a. X, it computes the proportion of all X-residues in a certain set of six proteins that are 95% buried in the native structure of the protein	Cornette *et al.*, 1987
Levitt, 1976	Supplemented estimates to Nozaki and Tanford, data based on free energy of transfer of a.a. side chains from H_2O to 100% ethanol or dioxane	Cornette *et al.*, 1987
Transfer energies Chothia, 1976; Wertz and Scheraga, 1978; Janin, 1976	Transfer energies computed from the respective scales. For example, in Janin it is simply RTlnf, where f is the ratio of buried to accessible molar fractions	Cornette *et al.*, 1987
Other scales		
Isoelectric point (pI)	Tritation	Zimmerman *et al.*, 1968
Alpha pK_1	Tritation	Jungck, 1978
Sequence frequency	Frequency of occurence of a.a. in 68 heterologous evolutionarily diverse proteins	Jukes *et al.*, 1975
Composition Grantham, 1974	It is defined as the atomic weight ratio of hetero (noncarbon) elements in end groups or rings to carbons in the side chain	Grantham, 1974
Secondary structure scales of proteins		
β-turns	Overall a.a. composition of β-turns for 29 proteins as compiled by Chou and Fasman	Jurka and Smith, 1987a
α-helix		
O'Neil and DeGrado, 1990	α-helical versus random coil states have been obtained through the design of a peptide that forms a noncovalent α-helical dimer, which is in equilibrium with a randomly coiled monomeric state	O'Neil and DeGrado, 1990
Chou and Fasman, 1974	Helix conformational parameter P_α calculated from the frequency of helical residues in proteins	Chou and Fasman, 1974
Wojcik, Altmann and Scheraga, 1990	Helix stability constants, s, from host-guest experiments	O'Neil and DeGrado, 1990
β-sheets		
Chou and Fasman, 1974	Conformational parameter P_β	Chou and Fasman, 1974

TABLE I
Continued.

Property	Determination procedure or description	Data sources
Kim and Berg, 1993	Thermodynamic β-sheet propensities	Kim and Berg, 1993
Minor and Kim, 1994a	Scale of the relative propensity for β-sheet formation of the a.a. in a variant of the small, monomeric, β-sheet-rich, IgG-binding domain from protein G	Minor and Kim, 1994a
Random coil regions	Conformational parameter P_c	Chou and Fasman, 1974
B. *Dinucleoside monophosphates*		
Hydrophilicity Weber and Lacey, 1978	R_F in (10/90:v/v) 1.0 M ammonium acetate/saturated ammonium sulphate, pH 7.0 at 25 C	Weber and Lacey, 1978
Hydrophilicity Barzilay *et al.*, 1973	R_F from paper chromatography using saturated ammonium sulphate/ 1M sodium acetate/isopropanol (80/18/2:v/v/v) as a developing solvent	Barzilay *et al.* 1973
Hydrophobicity Garel *et al.*, 1973	The partition coefficient is $K=R_F/(1-R_F)$; R_X was calculated as the product of the R_X's of the constituents	Garel *et al.* 1973; Jungck, 1978

190 absolute values of the differences between ranks was then used to calculate the mean (μ) and standard deviation (σ) for each amino acid property. Obviously these statistics refer to those of the population and not those of the sample, as the 190 differences represent the entire population of distances.

Figure 1 in Wong (1975) reports 19 biosynthetic relationships between amino acids. For each of these 19 biosynthetic relationships, I calculated the absolute value of the difference between the ranks of the two amino acids participating in the definition of a precursor- product pair. I then separated these pairs according to whether they belonged to the rows or the columns of the code (see for instance Table II) as it is known that the biosynthetic relationships are distributed principally over the rows (Dillon, 1973; Taylor and Coates, 1989; Miseta, 1989), while the physicochemical properties of amino acids are mainly distributed over the columns of the code (Nelsestuen, 1978; Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a; Taylor and Coates, 1989). (The pairs of amino acids, such as Asp-Glu, which belong to a row and a column of the code at the same time, were attributed to the columns because, as mentioned above, the genetic code organizes amino acid properties into columns). I then calculated the mean (m) of the absolute values of the differences between the ranks for the set of considered pairs, such as, for instance, the 11 pairs of amino acids in precursor-product relationships

TABLE II

Significance spectrum of bulkiness (Zimmerman *et al.*, 1968) obtained by applying the statistical test (see Materials and Methods). m_s and m_w indicate the simple and weighted means, respectively, of the absolute value of the differences between ranks; n_s and n_w indicate the simple and weighted numerousness, respectively; and P is the probability that the mean value indicated will be obtained. The mean (μ) and the standard deviation (σ) of the population of the 190 absolute values of the differences between ranks are μ=6.99 and σ=4.58, respectively. See Materials and Methods for further information

	m_s	n_s	P	m_w	n_w	P
Precursor-product amino acids						
columns	2.44	8	6.2×10^{-3}	2.54	28	2.5×10^{-6}
rows	4.86	11	0.12	4.24	23	9.2×10^{-3}
columns plus rows	3.84	19	5.8×10^{-3}	3.30	51	3.3×10^{-7}
Non precursor-product amino acids						
columns	6.41	28	0.46	6.17	82	0.26
rows	7.73	33	0.94	8.29	71	1.00
columns plus rows	7.12	61	0.86	7.15	153	0.96

belonging to the rows of the genetic code (see, for instance, Table II). Finally, I have calculated the probability with which the mean value m of the absolute values of the differences between ranks can be observed in the sample of pairs considered using the normal distribution: $P(Z\leq(m-(\mu-0.5))/(\sigma/(n)^{1/2}))$ (Balaam, 1972), thus using continuity correction, and where Z represents the standardized normal variable.

For every subset of pairs of amino acids, the absolute values of the differences between ranks were associated with two types of mean, a simple one (arithmetic mean) and a weighted one. The simple mean (m_s) associates every distance with a numerousness (n_s) of 1, while the weighted mean (m_w) associates every distance with a numerousness (n_w) equal to the number of times that the codon of the i-th amino acid is transformed, on the basis of the genetic code structure and by means of a single base change, into the codon of the j-th amino acid (for instance, for the Val-Leu pair the weight is 6 which represents the number of times that the four codons of Val transform, on the basis of the genetic code, into the six codons of Leu when its codons undergo a single base change). The weighted mean thus shows the effect of the genetic code on the distances specified in it.

These calculations were performed by making a clear distinction between the 19 pairs of amino acids in precursor-product relationships or in close biosynthetic relationships (Wong, 1975; Figure 1, p. 1910) and the 61 pairs that are not in a precursor-product relationship but are nevertheless defined in the genetic code table, i.e. pairs codified by codons differing in a single base. A further classification was then made in pairs of rows or columns, as mentioned above, and the row calculations

were joined to those of the columns. A typical table of these calculations is reported in Table II.

3. Results

3.1. Scales of amino acid 'size'

I have analyzed six scales measuring the size of amino acids (Table I) and the frequency of amino acids in proteins, which is correlated to them (Jungck, 1978; Di Giulio, 1989a), and I have essentially observed the pattern shown in Table II. All these properties, with the exception of specific volume, display, with non-significant variations, the behaviour typified by bulkiness (Table II). In other words, the statistical test (see Materials and Methods) is significant or highly significant only for the pairs of amino acids in precursor-product relationships, whereas it is never significant for pairs that are not in precursor-product relationships (Table II). In particular, for the pairs in precursor-product relationships the simple mean (m_s) is significant or highly significant for the column pairs (Table II), whereas it is not significant for row pairs but is, vice versa, significant for the 19 precursor-product pairs (columns plus rows) (Table II). The conclusion to be drawn from the above is that the amino acids in precursor-product or close biosynthetic relationships (Wong, 1975) have, on average, a similar size value. In other words, the biosynthetic pathways between amino acids preserve their size. Furthermore, if we consider the effect of the genetic code on these distances, i.e. if we look at the weighted mean (m_w) (see Materials and Methods), we can see that the significance increases by at least two orders of magnitude (Table II). Hence, the structure of the genetic code has strengthened the distances specified in the biosynthetic relationships between amino acids.

For the pairs of amino acids that are not in precursor-product relationships, no significance is ever observed in the statistical test (Table II) in any of the seven properties analyzed (Table I), suggesting that the size of the amino acids in these pairs was not an important variable.

In the literature there are several indications linking the size of amino acids to the genetic code structure (Alff-Steinberger, 1969; Jungck, 1978; Swanson, 1984; Di Giulio, 1989a; Taylor and Coates, 1989).

3.2. Scales of hydrophilicity and hydrophobicity

I have used the statistical test on fourteen scales measuring the hydrophilicity or hydrophobicity of the amino acids, whether free, linked to a protein context or in scale combinations (Table I). The behaviour of these scales can be exemplified using the polarity of Woese *et al.* (1966) (Table III) as a reference. As can be seen in Table III, these scales display a significance or only a marginal significance in the 19 pairs of amino acids in precursor-product relationships: these pairs are significant

TABLE III

Significance spectrum of amino acid polarity (Woese *et al.*, 1966). For the symbols used and further information, see the legend for Table II and Materials and Methods. The mean (μ) and the standard deviation (σ) of the population of the 190 absolute values of the differences between ranks are μ=6.99 and σ=4.59, respectively

	m_s	n_s	P	m_w	n_w	P
Precursor-product amino acids						
columns	2.31	8	5.0×10^{-3}	2.46	28	1.7×10^{-6}
rows	6.68	11	0.55	6.41	23	0.47
columns plus rows	4.84	19	0.058	4.24	51	2.3×10^{-4}
Non precursor-product amino acids						
columns	4.66	28	0.017	3.80	82	5.6×10^{-8}
rows	7.54	33	0.90	7.25	71	0.92
columns plus rows	6.22	61	0.32	5.40	153	1.6×10^{-3}

or highly significant for the columns while they are never significant for the rows (Table III). The conclusion that seems to emerge is that the hydrophilicity and hydrophobicity of amino acids was, likewise, selected on the 19 pairs in precursor-product relationships and, in particular, that a strong influence was exerted on the 8 pairs belonging to the code's columns. Here too, we can see that the effect of the genetic code on these distances was to strengthen them by at least two orders of magnitude, when the results of the weighted mean are analyzed (Table III).

As regards the pairs of amino acids not in precursor-product relationships, most of the scales display statistical significance only for the column pairs of the simple mean, while both the row pairs and those of the rows plus columns are never significant (Table III). This leads to the conclusion that there was a pressure that allocated the amino acids with similar hydrophilicity-hydrophobicity values only on the genetic code columns, which is in agreement with the literature (Nelsestuen, 1978; Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a; Taylor and Coates, 1989). The interesting feature here lies in the fact that the set of these pairs (columns plus rows) does not display significance in any of the 14 scales used (Table I). This indicates that, as far as the simple mean is concerned, the set of pairs not in precursor-product relationships underwent a marginal influence by these scales in certain stages of the origin of the genetic code. As far as the weighted mean is concerned, the genetic code brought about a considerable strengthening of the distances on the pairs not in precursor-product relationships belonging to the columns; moreover, this strengthening was such that in 50% of the scales used, the weighted means of the columns plus rows pairs were made significant (Table III).

In the literature there are several indications linking the scales of hydrophilicity and hydrophobicity to the genetic code (Woese *et al.*, 1966; Alff-Steinberger, 1969; Nelsestuen, 1978; Swanson, 1984; Wolfenden *et al.*, 1979; Di Giulio, 1989a, b;

TABLE IV

Significance spectrum of measurement of amino acid propensity to enter the α-helix (O'Neil and DeGrado, 1990). For the symbols used and further information, see the legend for Table II and Materials and Methods. The mean (μ) and the standard deviation (σ) of the population of the 190 absolute values of the differences between ranks are μ=6.99 and σ=4.58, respectively

	m_s	n_s	P	m_w	n_w	P
Precursor-product amino acids						
columns	7.12	8	0.65	7.86	28	0.94
rows	7.91	11	0.85	8.76	23	0.99
columns plus rows	7.58	19	0.85	8.26	51	1.00
Non precursor-product amino acids						
columns	7.38	28	0.85	8.08	82	1.00
rows	6.82	33	0.66	7.51	71	0.97
columns plus rows	7.07	61	0.84	7.82	153	1.00

TABLE V

Significance spectrum of measurement of amino acid propensity to enter β-sheets (Kim and Berg, 1993). For the symbols used and further information, see the legend for Table II and Materials and Methods. The mean (μ) and the standard deviation (σ) of the population of the 190 absolute values of the differences between ranks are μ=6.95 and σ=4.60, respectively

	m_s	n_s	P	m_w	n_w	P
Precursor-product amino acids						
columns	2.56	8	8.4×10^{-3}	2.43	28	1.9×10^{-6}
rows	5.04	11	0.15	4.63	23	0.029
columns plus rows	4.00	19	0.010	3.42	51	1.3×10^{-6}
Non precursor-product amino acids						
columns	5.89	28	0.26	5.95	82	0.16
rows	7.98	33	0.97	8.68	71	1.00
columns plus rows	7.02	61	0.83	7.22	153	0.98

Taylor and Coates, 1989; Haig and Hurst, 1991; Di Giulio *et al.*, 1994; Talstrup *et al.*, 1994).

3.3. SCALES MEASURING THE PROPENSITY OF AMINO ACIDS TO ENTER THE SECONDARY STRUCTURE OF PROTEINS

I have analyzed three scales reflecting the propensity of amino acids to enter the α-helix (Table I), some of which are not correlated. None of these three scales shows traces of statistical significance. Table IV shows the significance spectrum

of the measurement taken by O'Neil and DeGrado (1990) as an example of the three scales (Table I). The conclusion to be drawn from this data is that the α-helix did not originate with the genetic code but was a structural element of proteins that evolved only more recently.

I have also performed the statistical test on three β-sheet scales (Table I). Two of these scales (Kim and Berg, 1993; Minor and Kim, 1994a) show a similar behaviour. Table V gives an example of the significance spectrum. As can be observed in Table V, the significance can be seen only in the pairs in precursor-product relationships. In particular, there is a highly significant probability for the column pairs in one scale (Table V) and a non-significant one in the other (m_s=5.00, n_s=8, P=0.18; m_w=5.07, n_w=28, P=0.049). While for the row pairs there is a marginal significance in one scale (Table V) while a significant value is obtained in the other (m_s=4.00, n_s=11, P=0.035; m_w=3.96, n_w=23, P=3.9×10^{-3}). In both scales there is a good strengthening of the distances when the weighted mean is analyzed (Table V). Therefore, as the 19 pairs in precursor-product relationships display significance towards the β-sheets (Table V), this seems to suggest that this structural element of proteins might have been the driving force behind the origin of genetic code organization and, obviously, that primitive proteins contained this theme. Vice versa, the measurement of the β-sheets by Chou and Fasman (1974) (Table I) is significant (m_s=5.07, n_s=28, P=0.049; m_w=4.72, n_w=82, P=2.2×10^{-4}) only for the column pairs not in precursor-product relationships, thus casting a shadow of doubt on the conclusion referred above; otherwise this discrepancy can be attributed to the fact that β-sheet propensity seems to be context-dependent (Minor and Kim, 1994b).

I have also analyzed the distribution of β-turns (Table I) in the genetic code, and observed a significance only for the precursor-product pairs belonging to the columns (m_s=2.69, n_s=8, P=9.6×10^{-3}; m_w=2.96, n_w=28, P=2.4×10^{-5}), even if in the 19 precursor-product pairs there is a behaviour that might suggest a sort of statistical significance (m_s=5.68, n_s=19, P=0.22; m_w=4.82, n_w=51, P=4.7×10^{-3}).

Finally, the coil regions (Table I) show significance or quasi-significance only for the pairs of amino acids belonging to the columns of the genetic code, both for the precursor-product pairs (m_s=4.31, n_s=8, P=0.092; m_w=4.28, n_w=28, P=$5.8\text{x}10^{-3}$) and for the non precursor-product pairs (m_s=4.89, n_s=28, P=0.034; m_w=4.89, n_w=82, P=$9.1\text{x}10^{-4}$).

In the literature, attempts to link secondary protein structures to the genetic code have proved to be unsuccessful (Salemme *et al.*, 1977; Goodman and Moore, 1977), with the exception of β-turns (Jurka and Smith, 1987a, b).

3.4. OTHER AMINO ACID SCALES

The isoelectric point (Table I) shows no significant probability. Whereas, α-pK_1 (Table I) shows significance only for the precursor-product pairs belonging to the columns (m_s=3.56, n_s=8, P=0.035; m_w=3.86, n_w=28, P=1.2×10^{-3}) and a marginal

TABLE VI

Significance spectrum of the dinucleoside monophosphates of anticodons (Weber and Lacey, 1978). For the symbols used and further information, see the legend for Table II and Materials and Methods. The mean (μ) and the standard deviation (σ) of the population of the 190 absolute values of the differences between ranks are μ=6.97 and σ=4.59, respectively

	m_s	n_s	P	m_w	n_w	P
Precursor-product amino acids						
columns	2.00	8	2.9×10^{-3}	1.71	28	2.0×10^{-8}
rows	6.32	11	0.46	6.41	23	0.48
columns plus rows	4.50	19	0.031	3.83	51	2.0×10^{-5}
Non precursor-product amino acids						
columns	4.43	28	9.3×10^{-3}	3.90	82	2.0×10^{-7}
rows	8.00	33	0.97	7.92	71	1.00
columns plus rows	6.36	61	0.43	5.76	153	0.028

significance in the 19 pairs (m_s=5.53, n_s=19, P=0.18; m_w=5.07, n_w=51, P=0.014). Finally, the composition index of Grantham (1974) (Table I) displays a significant behaviour only on the columns for the non precursor-product pairs (m_s=4.93, n_s=28, P=0.057; m_w=4.74, n_w=82, P=1.1×10^{-3}).

3.5. THE PROPERTIES OF THE DINUCLEOSIDE MONOPHOSPHATES OF ANTICODONS AND CODONS

I have used the statistical test to analyze the properties of the dinucleoside monophosphates (Table I) as if they were amino acid properties.

For the dinucleoside monophosphates of anticodons, it is possible to observe a behaviour that is almost identical in all three scales. Table VI reports the significance spectrum for the measurement of Weber and Lacey (1978). As can be seen in Table VI, for the anticodons of amino acids in precursor-product relationships, we can observe a significance both for the 19 pairs and, more strongly, on the columns, while no significance is detected on the rows (Table VI). A highly significant probability was observed on the columns for the pairs not in precursor- product relationships, while no significance is detected on the rows (Table VI). Overall, these properties (Table VI) behave like scales of hydrophilicity-hydrophobicity (Table III).

As regards the dinucleoside monophosphates of codons, the three scales (Table I) show a behaviour similar to that in Table VII. The significance is limited to the codons of the precursor-product pairs and, in particular, to the column pairs (Table VII). This behaviour is thus similar to that of the size of amino acids (Table II). However, one exception can be observed. In the measurement of Garel *et al.*

TABLE VII

Significance spectrum of the dinucleoside monophosphates of codons (Weber and Lacey, 1978). For the symbols used and further information, see the legend for Table II and Materials and Methods. The mean (μ) and the standard deviation (σ) of the population of the 190 absolute values of the differences between ranks are μ=6.97 and σ=4.59, respectively

	m_s	n_s	P	m_w	n_w	P
Precursor-product amino acids						
columns	3.62	8	0.040	3.07	28	4.4×10^{-5}
rows	6.14	11	0.41	6.50	23	0.51
columns plus rows	5.08	19	0.093	4.62	51	2.0×10^{-3}
Non precursor-product amino acids						
columns	6.54	28	0.53	6.02	82	0.19
rows	5.64	33	0.15	5.75	71	0.093
columns plus rows	6.05	61	0.24	5.90	153	0.062

(1973), the pairs of amino acids not in precursor-product relationships show signs of significance (m_s=5.16, n_s=28, P=0.072; m_w=5.07, n_w=82, P=3.8×10^{-3}).

This data (Tables VI and VII) shows that the anticodons display a wider significance spectrum than the codons and could, therefore, have played a greater role in structuring the genetic code (Nagyvary and Fendler, 1974; Weber and Lacey, 1978; Jungck, 1978; Lacey and Mullins, 1983). It is also extremely clear that the properties of the dinucleoside monophosphates of anticodons are arranged in columns (Table VI).

4. Discussion

4.1. The biosynthetic pathways between amino acids might have been selected in order to build the β-sheets of proteins and might have represented a more ancient code that led to the characterization of the earliest messenger RNAs

The main result of the present analysis is the demonstration that the physicochemical properties of amino acids are not randomly distributed over amino acids in precursor-product relationships (Tables II, III, and V; see Results). However, some properties, e.g. bulkiness of amino acids, are specifically associated to the precursor-product relationships (Table II). [This is contrary to the literature (Szathmary, 1991; Szathmary and Zintzaras, 1992; Szathmary, 1993)]. If we consider that the propensity of amino acids to enter β-sheets is related to these amino acid pairs (Table V), then these fundamental themes of protein structure might have been the driving force behind the origin of the genetic code. In particular, the observation

that the pairs of amino acids in precursor-product relationships preserve their size (Table II; see Results) seems to be linked to the analogous observation that β-sheets are related to these pairs (Table V; see Results), as the correlation coefficient (r) between the bulkiness values (Zimmerman *et al.*, 1968) and those of the β-sheets (Kim and Berg, 1993) is highly significant (r=–0.749, F=23.07, df=(18, 1), $P<10^{-4}$). [This moreover gives weight to the suggestion that bulkiness is the simplest and one of the most fundamental amino acid properties related through steric hindrance to secondary and tertiary protein structure (Zimmerman *et al.*, 1968)]. Therefore, the selection in favour of amino acid size (Table II) might not have been an important element per se but, rather, because it made possible the construction of proteins formed mainly of β-sheets (Table V) (see also the following section).

The idea that β-sheets might have characterized primitive proteins is certainly not new (Orgel, 1972; Orgel, 1975; Brack and Orgel, 1975; von Heijne *et al.*, 1978; Marlborough, 1980; Hartman, 1995). Furthermore, Jurka and Smith (1987a, b) suggest that β-turns became objects of selection in the prebiotic environment and affected the evolution of the genetic code and the biosynthetic pathways of amino acids. In particular, they find (Jurka and Smith, 1987a, b) that the most abundant amino acids in β-turns are also precursors of other amino acids, and they suggest that these were implanted in the code early on. These observations are consistent with the contents of the present paper and also lend themselves to a more general interpretation. This can be seen in Orgel (1977) which discusses the importance of β-turns stabilized by β-sheets as plausible sites of the early enzymatic activity. Therefore, there might have been an initial phase in the development of the genetic code in which a selection was made for β-turns on precursor amino acids (Jurka and Smith, 1987a, b) followed by a phase in which the biosynthetic connections with product amino acids evolved through the selection of β-sheets, thus giving rise to the very origin of the genetic code, as the data shown here seems to support (see also Tables II and V).

All this seems to provide evidence in favour of the idea that the biosynthetic relationships between amino acids might have been a code that was even more ancient than the genetic code, which led to the characterization of the earliest messenger RNAs, and that the flows through these pathways might have specified a crude primary structure of primitive proteins which was perfected only after the genetic code was established, i.e. with the assertion of the messenger RNA. Evidence in favour of this idea seems to lie in the observation that, as far as the properties considered to be important are concerned, the statistical test is always significant on the simple mean for the 19 pairs in precursor- product relationships (Tables II, III and V; see Results). Furthermore, still on the simple mean, the set of pairs not in precursor-product relationships is never significant (Tables II, III and V; see Results), once again favouring the idea of an ancient code based on the biosynthetic pathways of amino acids. I feel that this is, in part, different from 'The code within the codons' discussed by Taylor and Coates (1989) because, as I suggest in the present paper, the fundamental core of the genetic code was

organized around the biosynthetic relationships between amino acids (Pelc, 1965; Dillon, 1973; Wong, 1975; Miseta, 1989). While the above authors (Taylor and Coates, 1989) propose that the information content of the three codon positions does indeed reflect an intracodonic code that served as the most ancient language of cells, they also attribute a fundamental role to the amino acid-codon (anticodon) interactions, which I do not think is immediately visible. However, I will return to this point later on.

In conclusion to this section, I feel that, regardless of the mechanism that linked the biosynthetic pathways of amino acids to the genetic code structure, whether transformations from a precursor-tRNA-like molecule to a product-tRNA-like molecule (Wong, 1975; see below), mechanisms predating the tRNA/ribosome system (Taylor and Coates, 1989), or even the existence of a more ancient code based on the biosynthetic flows and relationships between amino acids, as outlined here, the analysis (Tables II, III and V; see Results) shows that the biosynthetic pathways connecting the amino acids are intimately linked to the genetic code and played a fundamental role in defining its organization.

4.2. On the problem of the compatibility between the physicochemical postulates and the coevolution hypothesis

If the biosynthetic relationships between amino acids did effectively had a profound influence on genetic code organization (Pelc, 1965; Dillon, 1973; Wong, 1975; Taylor and Coates, 1989; Miseta, 1989) and this took place through the origin and selection of primitive mRNAs codifying for proteins having β-sheets as their main structural theme, then it is to be expected that some properties of amino acids are specifically reflected in the precursor-product relationships (Tables II and V) because these structural themes should be expressed through a few properties of amino acids (e.g. bulkiness), as suggested in the previous section. On the other hand, we do not expect a strong correlation between the bulkiness of amino acids and the properties of codons or anticodons because, if the size of amino acids was selected not per se but as a reflection of the β-sheets, then we can see no simple mechanism that should address the correlation between the values of bulkiness and those of the codon or anticodon properties.

However, Jungck (1978) found that the properties of the dinucleoside monophosphates of anticodons (ant) (Weber and Lacey, 1978) correlate with the bulkiness values (bul) (Zimmerman *et al.*, 1968) and, indeed, the correlation coefficient is highly significant ($r=-0.665$, $F=14.30$, $df=(18, 1)$, $P=1.4\times10^{-3}$). Furthermore, there are other significant coefficients (Jungck, 1978): between polarity (pol) (Woese *et al.*, 1966) and the dinucleoside monophosphates of anticodons (Weber and Lacey, 1978) ($r=+0.890$, $F=68.40$, $df=(18, 1)$, $P<10^{-4}$) and between bulkiness (Zimmerman *et al.*, 1968) and polarity (Woese *et al.*, 1966) ($r=-0.521$, $F=6.70$, $df=(18, 1)$, $P=0.018$). Hence at least part of the correlation between bul and ant might depend on the correlation between bul and pol. I have therefore calculated the partial lin-

ear correlation coefficient (R) (Kenny, 1979) between bul and ant, while keeping constant, and thus eliminating, the influence of pol from the correlation, obtaining a reduction in significance (R=–0.517, t=2.56, df=18, P=0.019). Therefore, part of the correlation between bul and ant is determined by pol. An analogous calculation shows that bul exerts little influence on the correlation between pol and ant (R=+0.853, t=6.93, df=18, $P<10^{-4}$). It would consequently appear that bulkiness effectively possesses a limited correlation with the values of the dinucleoside monophosphates of anticodons when polarity is taken into account, and this is to be expected if the bulkiness of amino acids was selected as an expression of the β-sheets and not per se.

In order to further investigate the above point, I have performed the following multiple linear regression analysis, in this case using the absolute values of the differences between pairs of amino acids for the three different properties: bulkiness (Δbul) (Zimmerman *et al.*, 1968), polarity (Δpol) (Woese *et al.*, 1966) and dinucleoside monophosphates of anticodons (Δant) (Weber and Lacey, 1978)*.

Overall, the above analysis (see also Footnote) suggests the existence of a robust correlation between the properties of the dinucleoside monophosphates of anticodons (Weber and Lacey, 1978) and the polarity of amino acids (Woese *et al.*, 1966), but a weak correlation between the former and bulkiness (Zimmerman *et al.*, 1968). This led me to re-examine the equation 5 of Jungck (1978): ant=0.032+0.037pol–0.0074bul (r=+0.921, F=47.30, df=(17, 2), $P<10^{-4}$). As already mentioned, analysis of the variance between ant and pol gives a strong significance (r=+0.890, F=68.40, df=(18, 1), $P<10^{-4}$), a good significance between ant and bul (r=–0.665, F=14.30, df=(18, 1), $P=1.4\times10^{-3}$) and there is also a significant correlation between bul and pol (r=–0.521, F=6.70, df=(18, 1), P=0.018). However, in the multiple equation (Jungck 1978; Eq. 5) only the regression coefficient of pol is highly significant (t=6.72, df=18, $P<10^{-4}$) while that of bul possesses a limited significance (t=2.50, df=18, P=0.023). It is thus clear that, consistently

* For the 19 pairs in precursor-product relationships, we get the equation Δant=0.017+0.027Δpol+0.0069Δbul with a significance in the regression coefficient of Δpol (t=4.20, df=17, $P=7.0\times10^{-4}$) and not in that of Δbul (t=1.56, df=17, P=0.14) (the general significance of the equation is: r=+0.738, F=9.60, df=(16, 2), $P=1.8\times10^{-3}$). Analogous results can be obtained if we analyze the pairs not in precursor-product relationships of the columns, of the rows and the total of the pairs. For the columns we get the equation Δant=0.022+0.032Δpol+0.0043Δbul with a significance in the regression coefficient of Δpol (t=5.81, df=26, $P<10^{-4}$) and not in that of Δbul (t=1.66, df=26, P=0.11) (the general significance of the equation is: r=+0.790, F=20.79, df=(25, 2), $P<10^{-4}$). For the rows we get the equation Δant=0.071+0.032Δpol–0.00092Δbul with a significance in the regression coefficient of Δpol (t=4.29, df=31, $P=2.0\times10^{-4}$) and not in that of Δbul (t=0.29, df=31, P=0.77) (the general significance of the equation is: r=+0.626, F=9.64, df=(30, 2), $P=6.0\times10^{-4}$). Finally, for the total of the pairs we get the equation Δant=0.037+0.030Δpol+0.0032Δbul with a significance in the regression coefficient of Δpol (t=8.16, df=77, $P<10^{-4}$) and not in that of Δbul (t=1.73, df=77, P=0.088) (the general significance of the equation is: r=+0.711, F=38.81, df=(76, 2), $P<10^{-4}$). Although for the last regression coefficient, and only in this case, the correlation between Δant and Δbul is significant (r=+0.267, F=5.91, df=(77, 1), P=0.017).

As a further control, I have also performed the same analysis but using the absolute values of the differences between ranks of the pairs of amino acids for the three properties investigated. Here too, I never observed significance in the regression coefficients of Δbul ($P\geq0.16$).

with the analysis referred above, the part of data variance explained by bulkiness is extremely limited, whereas polarity explains most of the variation. All this seems to me to agree with the idea that the bulkiness of amino acids was not selected per se but because it reflected β-sheets and thus did not show, as expected, a strong correlation with the dinucleoside monophosphates of anticodons.

The data referred in the literature (Weber and Lacey, 1978; Jungck, 1978; Lacey and Mullins, 1983), the correlation analysis illustrated above, the non-specificity of the significance of polarity on the precursor-product pairs (Table III; see Results) and the perfect overlapping of the significance spectrum of amino acid polarity (Table III) and that of the dinucleoside monophosphates of anticodons (Table VI) suggest that there must have been interactions between amino acids and anticodons. However, as far as the simple mean is concerned, the non-significance on the 61 pairs of amino acids not in precursor-product relationships (Table III; see Results) contrasts with the significance on the 19 pairs that are in such a relationship (Table III; see Results), thus giving weight to the idea that the latter pairs might have represented a particularly important feature in the origin of the genetic code. However, we still have to try to understand at which stage in the origin of the genetic code the amino acid-anticodon interactions occurred, and whether or not this is compatible with the coevolution hypothesis.

From a general viewpoint, the coevolution hypothesis identifies two phases in the origin of the genetic code (Wong, 1988): phase 1 with the precursor amino acids occupying the code in an early stage, and phase 2 with the late entrance of product amino acids. The concession of codons (anticodons) from the precursor amino acid to the product seems to have been mediated by tRNA-like molecules on which the precursor-product transformations are supposed to have taken place (Wong, 1975; Wachtershauser, 1988; Danchin, 1989; Di Giulio, 1993; Di Giulio, 1994b). Obviously, if this was the codon (anticodon) concession mechanism, then the origin of the tRNA molecule must include some characteristics that make it possible to justify this amino acid-anticodon interaction. However, as the domain of the codons of the precursor had already been assigned in the first phase of code development, the product amino acid was able to choose only from these codons (anticodons) the one with the properties most similar to its own. Furthermore, as the genetic code developed, the product amino acids which were the last to enter were attributed with codons (anticodons), with very limited scope for choice. If this is true, the coevolution hypothesis is obviously in contrast with highly deterministic stereochemical models of genetic code origin, at least for the majority of product amino acids. Szathmary (1993) supports a different viewpoint.

Consequently, what, if any, are the characteristics of the ancestral tRNA molecules that can justify an amino acid-anticodon interaction such as the one analyzed here? It has been repeatedly suggested that a hairpin structure was the ancestor of the tRNA molecule (Hopfield, 1978; Eigen and Winkler-Oswatitsch, 1981; Bloch *et al.*, 1985; Moller and Janssen, 1990, 1992; Di Giulio, 1992; Maizels and Weiner, 1993; Schimmel *et al.*, 1993; Maizels and Weiner, 1994; Dick and Schamel, 1995;

Di Giulio, 1995b). Furthermore, a hypothesis has been formulated on the existence of an ancient code sited in the stem of the hairpin structure and made evident by the nucleotides that today form the determinants of tRNA identity (Hopfield, 1978; de Duve, 1988; Moller and Janssen, 1990, 1992; Mursier-Forsyth and Schimmel, 1993; Schimmel and Henderson, 1994) and, from this hairpin structure, by means of direct duplication, the tRNA molecule might have originated (Di Giulio, 1992, 1995b). The latter duplication might also have transferred the code sited in the stem on the hairpin structure into the anticodon nucleotides, thus giving rise to the anticodon loop (Moller and Janssen, 1990, 1992). It is thus clear that it is at this stage that the amino acid-anticodon interaction, suggested by the correlation between amino acid polarity and anticodon hydrophilicity, might have been expressed. We need merely postulate that the determinants of tRNA identity represent the vestiges of the ancient anticodon forms (Hopfield, 1978; de Duve, 1988) sited in the stem of the hairpin structures (Moller and Janssen, 1990, 1992). It also seems reasonable to hypothesize that the early synthesis of oligopeptides might have taken place on the hairpin structures transporting amino acids (precursors) (Orgel, 1989; Di Giulio, 1994b) and that, in this dynamic evolutionary scenario, these structures charged with precursors might have started to form products choosing the anticodons most similar to themselves that were still available to their precursor. This is therefore in agreement with the observed correlation between amino acid polarity and anticodon hydrophilicity. This point of view seems to find evidence in experiments conducted on hairpin structures accommodating anticodons (Shimizu, 1995) even if this research aimed to assay the reliability of a stereochemcial model of the genetic code (Shimizu, 1982).

Finally, in my opinion the perfect overlapping of the polarity significance spectrum (Table III) and that of the dinucleoside monophosphates of anticodons (Table VI), along with the analysis of correlation between the polarity and the anticodons mentioned above, does not provide evidence in favour of the ambiguity reduction hypothesis (Woese, 1965; Fitch, 1966; Fitch and Upper, 1987). Indeed, this parallelism (Tables III and VI) and correlation seem to imply more an amino acid-anticodon interaction than a selective pressure organizing the genetic code in columns (Nelsestuen, 1978; Wolfenden *et al.*, 1979; Sjostrom and Wold, 1985; Di Giulio, 1989a; Taylor and Coates, 1989) by means of ambiguity reduction, as the latter mechanism might have been favoured at least by the absence of this correlation, which is not the case.

In conclusion, there is clearly some conflict between the historical nature of the biosynthetic relationships of amino acids and the physicochemical determinism of amino acids, which together structured the genetic code. It seems to me that only by satisfactorily joining together these two forces to form a single picture will it be possible to form a new and more comprehensive theory on the origin of the organization of the genetic code.

References

Aboderin, A. A.: 1971, *Int. J. Biochem.* **2**, 537–544.
Alff-Steinberger, C.: 1669, *Proc. Natl. Acad. Sci.*, USA 64, 584–591.
Balaam, L.N.: 1972, 'Fundamentals of Biometry', Chapter 6. Allen, G, Uniwin M (eds.) London.
Balasubramanian, R., Seetharamulu, P. and Raghunathan, G.: 1980, *Origins Life Evol. Biosphere* **10**, 15–30.
Barzilay, I., Sussman, J.L. and Lapidot, Y.: 1973, *J. Chromat.* **79**, 139–146.
Baumann, U. and Oro, J.: 1993, *BioSystems* **29**, 133–141.
Bloch, D. P., McArthur, B. and Mirrop, S.: 1985, *BioSystems* **17**, 209–225.
Brack, A. and Orgel, L. E.: 1975, *Nature* **256**, 383–387.
Bull, H. B. and Breese, K.: 1974, *Arch. Biochem. Biophys.* **161**, 665–670.
Chou, P. Y. and Fasman, G. D.: 1974, *Biochemistry* **13**, 211–222.
Cothia, C.: 1976, *J. Mol. Biol.* **105**, 1–14.
Cornette, J. L., Cease, K. B., Margalit, H., Spouge, J. L., Berzofsky, J. A. and DeLisi, C.: 1987, *J. Mol. Biol.* **195**, 659–685.
Danchin, A.: 1989, *Prog. Biophys. Mol. Biol.* **54**, 81–86.
de Duve, C.: 1988, *Nature* **333**, 117–118.
Di Giulio, M.: 1989a, *J. Mol. Evol.* **29**, 191–201.
Di Giulio, M.: 1989b, *J. Mol. Evol.* **29**, 288–293.
Di Giulio, M.: 1991, *Z. Naturforsch.* **46C**, 305–312.
Di Giulio, M.: 1992a, *Origins Life Evol. Biosphere* **22**, 309–319.
Di Giulio, M.: 1992b, *J. Theor. Biol.* **159**, 199–214.
Di Giulio, M.: 1993, *J. Mol. Evol.* **37**, 5–10.
Di Giulio, M.: 1994a, *Origins Life Evol. Biosphere* **24**, 425–434.
Di Giulio, M.: 1994b, *J. Theor. Biol.* **171**, 303–308.
Di Giulio, M.: 1995a, *Origins Life Evol. Biosphere* **25**, 549–564.
Di Giulio, M.: 1995b, *J. Theor. Biol.* **177**, 95–101.
Di Giulio, M., Capobianco, M. R. and Medugno, M.: 1994, *J. Theor. Biol.* **168**, 43–51.
Dick, T. P. and Schamel, W. W. A.: 1995, *J. Mol. Evol.* **41**, 1–9.
Dillon, L. S.: 1973, *Bot. Rev.* **39**, 301–345.
Dunnill, P.: 1966, *Nature* **215**, 355–359.
Egelman, D. M., Steitz, T. A. and Goldman, A.: 1986, *Ann. Rev. Biophys. Chem.* **15**, 321–353.
Eigen, M. and Winkler-Oswatitsch, R.: 1981, *Naturwissenschaften* **68**, 282–292.
Epstein, C. J.: 1966, *Nature* **210**, 25–28.
Fitch, W. M.: 1966, *J. Mol. Biol.* **16**, 1–8.
Fitch, W. M. and Upper, K.: 1987, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 759–767.
Frommel, C.: 1984, *J. Theor. Biol.* **111**, 247–260.
Gamow, G.: 1954, *Nature* **173**, 318.
Garel, J. P., Filliol, D. and Mandel, P.: 1973, *J. Chromat.* **78**, 381–391.
Goldberg, A. L. and Wittes, R. E.: 1966, *Science* **153**, 420–424.
Goldman, N.: 1993, *J. Mol. Evol.* **37**, 662–664.
Goodman, M. and Moore, G. W.: 1977, *J. Mol. Evol.* **10**, 7–47.
Grantham, R.: 1974, *Science* **185**, 862–864.
Haig, D. and Hurst, L. D.: 1991, *J. Mol. Evol.* **33**, 412–417.
Hartman, H.: 1995, *J. Mol. Evol.* **40**, 541–544.
Hendry, L. B., Bransome, E. D. Jr., Hutson, M. S. and Campbell, L. K.: 1981, *Proc. Natl. Acad. Sci.*, USA 78, 7440–7444.
Hopfield, J. J.: 1978, *Proc. Natl. Acad. Sci.*, USA **75**, 4334–4338.
Janin, J.: 1979, *Nature* **277**, 491–492.
Jukes, T. H.: 1966, 'Molecules and Evolution'. Columbia University Press, New York.
Jukes, T. H., Holmquist, R. and Moise, H.: 1975, *Science* **189**, 50–51.
Jungck, J. R.: 1978, *J. Mol. Evol.* **11**, 211–224.
Jurka, J., Kolosza, Z. and Roterman, I.: 1982, *J. Mol. Evol.* **19**, 20–27.
Jurka, J. and Smith, T. F.: 1987a, *J. Mol. Evol.* **25**, 15–19.

Jurka, J. and Smith, T. F.: 1987b, *Cold Spring Harbor Symp. Quant. Biol.* **52**, 407–410.
Kenny, D. A.: 1979, 'Correlation and Casuality'. J Wiley & Sons, New York, pp. 54–56.
Kim, C. A. and Berg, J. M.: 1993, *Nature* **362**, 267–270.
Kyte, J. and Doolittle, R. F.: 1982, *J. Mol. Biol.* **157**, 105–132.
Lacey, J. C. Jr. and Mullins, D. W. Jr.: 1983, *Origins Life Evol. Biosphere* **13, 3–42.**
Lacey, J. C. Jr., Wickramasinghe, N. S. M. D. and Cook, G.W.: 1992, *Origins Life Evol. Biosphere* **22**, 243–275.
Lacey, J. C. Jr., Wickramasinghe, N. S. M. D., Cook, G. W. and Anderson, G.: 1993, *J. Mol. Evol.* **37**, 233–239.
Levitt, M.: 1976, *J. Mol. Biol.* **104**, 59–107.
Liquori, A. M. and Sadun, C.: 1978, *Gazz. Chim. Ital.* **108**, 513–517.
Maizels, N. and Weiner, A. M.: 1993, in: The RNA World, Gesteland RF, Atkins JF (eds.) Plainview NY: Cold Spring Harbor Laboratory Press, pp 577–602.
Maizels, N. and Weiner, A. M.: 1994, *Proc. Natl. Acad. Sci.* USA 91, 6729–6734.
Marlborough, D. I.: 1980, *Origins Life* **10**, 3–14.
McClendon, J. H.: 1986, *Origins Life* **16**, 260–270.
McClendon, J. H.: 1987, *Origins Life* **17**, 401–417.
Melcher, G.: 1974, *J. Mol. Evol.* **3**, 121–141.
Minor, D. L. Jr. and Kim, P. S.: 1994a, *Nature* **367**, 660–663.
Minor, D. L. Jr. and Kim, P. S.: 1994b, *Nature* **371**, 264–267.
Miseta, A.: 1989, *Physiol. Chem. Phys. Med. NMR* **21**, 237–242.
Moller, W. and Janssen, G. M. C.: 1990, *Biochimie* **72**, 361–368.
Moller, W. and Janssen, G. M. C.: 1992, *J. Mol. Evol.* **34**, 471–477.
Morowitz, H. J.: 1992, 'Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis', Yale University, Vail-Ballou Press, Binghamton, New York, pp. 160–171.
Mursier-Forsyth, K. and Schimmel, P.: 1993, *FASEB J.* **7**, 282–289.
Nagyvary, J. and Fendler, J. H.: 1974, *Origins Life* **5**, 357–362.
Nelsestuen, G. L.: 1978, *J. Mol. Evol.* **11**, 109–120.
Olsen, K. W.: 1980, *Biochim. Biophys. Acta* **622**, 259–267.
O'Neil, K. T. and DeGrado, W. F.: 1990, *Science* **250**, 646–651.
Orgel, L. E.: 1972, *Isr. J. Chem.* **10**, 287–292.
Orgel, L. E.: 1975, *Isr. J. Chem.* **14**, 11–16.
Orgel, L. E.: 1977, in: Bradbury E. M., Javaherian, K. (eds.), 'The Organization and Expression of the Eukaryotic Genome: Proceedings of the International Symposium', 3–6 May 1976, Teheran. Academic Press, London, pp. 499–504.
Orgel, L. E.: 1989, *J. Mol. Evol.* **29**, 465–474.
Pelc, S. R.: 1965, *Nature* **207**, 597–599.
Pelc, S. R. and Welton, M. G. E.: 1966, *Nature* **209**, 868–870.
Salemme, F. R., Miller, M. D. and Jordan, S. R.: 1977, *Proc. Natl. Acad. Sci.*, USA 74, 2820–2824.
Schimmel, P., Giege, R., Morras, D. and Yokoyama, S.: 1993, *Proc. Natl. Acad. Sci.*, USA 90, 8763–8768.
Schimmel, P. and Henderson, B.: 1994, *Proc. Natl. Acad. Sci.*, USA 91, 11283–11286.
Shimizu, M.: 1982, *J. Mol. Evol.* **18**, 297–303.
Shimizu, M.: 1995, *J. Biochem.* **117**, 23–26.
Siemion, I. Z. and Stefanowicz, P.: 1992, *BioSystems* **27**, 77–84.
Sjostrom, M. and Wold, S.: 1985, *J. Mol. Evol.* **22**, 272–277.
Sonneborn, T. M.: 1965, In: *Evolving Genes and Proteins*, Bryson, V. and Vogel, H. J. (eds.), New York, Academic press, pp. 377–397.
Swanson, R.: 1984, *Bull. Math. Biol.* **46**, 187–203.
Szathmary, E.: 1991, *J. Mol. Evol.* **32**, 178–182.
Szathmary, E.: 1993, *Proc. Natl. Acad. Sci.* USA 90, 9916–9920.
Szathmary, E. and Zintzaras, E.: 1992, *J. Mol. Evol.* **35**, 185–189.
Taylor, F. J. R. and Coates, D.: 1989, *BioSystems* **22**, 177–187.
Tolstrup, N., Toftgard, J., Engelbrecht, J. and Brunak, S.: 1994, *J. Mol. Biol.* **243**, 816–820.
Volkenstein, M. V.: 1966, *Biochem. Biophys. Acta* **119**, 421–424.

von Heijne, G., Blomberg, C. and Baltscheffsky, H.: 1978, *Origins Life Evol. Biosphere* **9**, 27–37.
Wachtershauser, G.: 1988, *Microbiol. Rev.* **52**, 452–484.
Weber, A. L. and Lacey, J. C. Jr.: 1978, *J. Mol. Evol.* **11**, 199–210.
Wertz, D. H. and Scheraga, H. A.: 1978, *Macromolecules* **11**, 9–15.
Wetzel, R.: 1978, *Origins Life* **9**, 39–50.
Woese, C. R.: 1965, *Proc. Natl. Acad. Sci.* USA 54, 1546–1552.
Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M. and Saxinger, W. C.: 1966, *Cold Spring Harbor Symp. Quant. Biol.* **31**, 723–736.
Wojcik, J., Altmann, K. H. and Scheraga, H. A.: 1990, *Biopolymers* **30**, 12–20.
Wolfenden, R. V., Cullis, P. M. and Southgate, C. C. F.: 1979, *Science* **206**, 575–577.
Wong, J. T.: 1975, *Proc. Natl. Acad. Sci.* USA 72, 1909–1912.
Wong, J. T.: 1976, *Proc. Natl. Acad. Sci.* USA 73, 2336–2340.
Wong, J. T.: 1980, *Proc. Natl. Acad. Sci.* USA 77, 1083–1086.
Wong, J. T.: 1981, *Trends Biochem. Sci.* **6**, 33–36.
Wong, J. T.: 1988, *Microbiol. Sci.* **5**, 174–181.
Wong, J. T. and Bronskill, P. M.: 1979, *J. Mol. Evol.* **13**, 115–125.
Yarus, M.: 1991, *New Biologist* **3**, 183–189.
Zimmerman, J. M., Eliezer, N. and Simha, R.: 1968, *J. Theor. Biol.* **21**, 170–201.