On the Relationships between the Genetic Code Coevolution Hypothesis and the Physicochemical Hypothesis

Massimo Di Giulio

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

Z. Naturforsch. **46c,** 305–312 (1991); received September 28, 1990

Genetic Code Theories, Coevolution, Precursor-Product Amino Acids, Amino Acids Properties, Hypergeometric Distribution

This paper analyzes the relationships between the genetic code coevolution hypothesis and the physicochemical hypothesis by means of a comparative study of the precursor-product amino acid pairs on which the former hypothesis is based. Even if the coevolution between the biosynthetic relationships of amino acids and the organization of the genetic code is not questioned in this paper, the results and the arguments used lead us to believe that the selective pressures considered essential by the physicochemical postulates, played a more active role than that of the precursor-product relationships in defining the allocation of these amino acids in the genetic code. It is furthermore pointed out that the two evolutionary hypothesis might be aspects of the same selective pressure, and thus difficult to differentiate.

Introduction

Some recent findings, such as self-splicing rRNA intron of ciliate *Tetrahymena* and other RNA molecules having catalytic activity [1], have given more credibility to the suggestion [2] that when life started on this planet, RNA was used as both a genome and a catalyst of its own replication [3]. Weiner and Maizels [4] discussed a model for the replication of RNA molecules that seems to have some implications for protein synthesis. They [4] suggest that tRNA-like structures which were important for the initiation of RNA synthesis, might have been specifically aminoacylated with an amino acid through an abarrant activity of the replicase. In this model the association of an amino acid to a specific codon should merely reflect the fortuitous affinity of the amino acid for the active site in a specific tRNA synthase: RNA replicase in Weiner and Maizels's model.

The association of an amino acid to a specific codon can also be explained by two theories proposed several years ago.

The first of these, the coevolution hypothesis of the genetic code [5, 6], suggests that codon distribution between amino acids was mainly determined by the concession of codons from the precursor amino acid to the product amino acid formed from the precursor. The organization of the genetic code should, therefore, reflect the biosynthetic relationships between amino acids. This evolutionary hypothesis seems to agree with the model discussed by Weiner and Maizels [4]: once the tRNA-like structure was charged by an amino acid through the aberrant activity of the RNA replicase, the amino acid was converted into the product amino acid while it was still attached to the tRNA-like molecule, as discussed by Wong [5]. There have been several direct [5], indirect [6, 7] and compatible [8–14] confirmations of the coevolution mechanism.

On the other hand, the physicochemical hypothesis of the genetic code [15, 16], postulates that amino acid allocation in the genetic code reflects the physicochemical properties of the amino acids, since the organization of the genetic code was determined by selective pressures tending either to minimize the deleterious effects of mutations [15] or to reduce translation errors [16]. This evolutionary hypothesis also seems to agree with Weiner and Maizels's [4] model: we need only to postulate that the tRNA-like structure was charged by tRNA synthase, through a generic discrimination between amino acids based on their physicochemical properties. The successive improvement of this apparatus through the above selective pressures probably determined amino acid allocation in the genetic code.

The physicochemical selective mechanism also appears to have received several corroborations

---

Reprint requests to Dr. M. Di Giulio.

Verlag der Zeitschrift für Naturforschung, D-7400 Tübingen
0939–5075/91/0300–0305 $ 01.30/0

binding the physicochemical properties of the amino acids to the organization of the genetic code [17–26].

In this paper I analyze the relationships between the coevolution hypothesis [5] and the physicochemical one [15, 16] by means of a comparative study on the precursor-product amino acid pairs on which the former hypothesis is based, so as to help identify which of the selective pressures contributed most in determining the allocation of the amino acids in a precursor-product relationship in the genetic code.

Results

The coevolution hypothesis of the genetic code [5, 6] is essentially based on the study of the contiguity of the precursor and product amino acid codons. Wong [5] uses the hypergeometric distribution to establish the correlation between codon allocation in the genetic code and biosynthetic pathways of the amino acids. In particular, Wong [5] considers that for every precursor codon, there will be *a* other triplets in the genetic code, which are contiguous to the amino acid codifying codons and *b* other triplets that are not. If a product of a given amino acid is codified by *n* codons, the random probability, *P*, that as many as *x* of these *n* codons will be contiguous to some of the precursor codons, can be determined according to the hypergeometric distribution [27]:

$$P = \sum_{x}^{n} \frac{a!}{(a-x)!x!} \cdot \frac{b!}{(b-n+x)!(n-x)!} \cdot \frac{(a+b-n)!n!}{(a+b)!} . \quad (1)$$

Wong [5], therefore, associates a probability, *P*, to every pair of precursor-product amino acids that this particular association was determined by chance. Finally, he [5] calculates the $-2 \cdot \ln P$ quantities and, uses Fisher's [28] method, which established that the $-2 \cdot \ln P$ quantity is distributed according to the χ^2 law with two degrees of freedom (d*f*). He then sums these quantities ($-2 \cdot \ln P$), and from the total χ^2 establishes the probability of the aggregate for the set of the precursor-product pairs that he considers.

Wong [5] examines eight pairs of amino acids for which the precursor-product relationship is certain and finds a χ^2 value of 45.01 ($df = 16$) and, therefore, a probability of 1.4×10^{-4} that for these eight pairs the set of precursor-product contiguities was determined by chance. (The calculation of the probability associated to a given χ^2 value was carried out using equation 4.2 shown in Lancaster [29].)

In this analysis Wong [5] does not include pairs of precursor-product amino acids for which the relationship is less certain although very plausible. I have extended Wong's analysis so as to include a further seven pairs for which the contiguity between the precursor-product amino acid codons can be seen in the actual genetic code. The results are shown in Table I. The first eight pairs are those taken from Wong [5], while the last seven pairs are the ones that I have added. The χ^2 value for the seven additional pairs is 25.73 ($df = 14$, $P = 0.0280$). Considering the set of all fifteen precursor-product pairs (Table I), we get a χ^2 value of 70.74 ($df = 30$) and a probability of 3.9×10^{-5} that the precursor-product relationships were generated by chance.

The physicochemical hypothesis [15, 16] attributes a fundamental role to the physicochemical properties of the amino acids in the organization of the genetic code. It could be asked, therefore, as in the precursor-product relationship, what the probability is that the choice of that particular pair of amino acids was determined according to the

Table I. Random probability of precursor-product codon contiguities. The parameters *a*, *b*, *n*, *x* and *P* are defined by Eqn. (1) and in the text. For the precursor-product pairs that are biosynthetically interconvertible, as can be seen in Wong's [5] figure 1, the precursor-product direction has been established on the basis of this figure, shifting from the central amino acids (Asp, Glu) towards the amino acids on the outside (products).

Precursor-product	*a*	*b*	*n*	*x*	*P*	$-2 \cdot \ln P$
Ser–Trp	34	24	1	1	0.586	1.07
Ser–Cys	34	24	2	2	0.339	2.16
Val–Leu	24	36	6	6	0.00268	11.84
Thr–Ile	24	36	3	3	0.0591	5.66
Gln–His	14	48	2	2	0.0481	6.07
Phe–Tyr	14	48	2	2	0.0481	6.07
Glu–Gln	14	48	2	2	0.0481	6.07
Asp–Asn	14	48	2	2	0.0481	6.07
Thr–Met	24	36	1	1	0.400	1.83
Ala–Ser	24	36	6	4	0.167	3.58
Gly–Ser	24	36	6	2	0.780	0.50
Ala–Gly	24	36	4	4	0.0218	7.65
Glu–Asp	14	48	2	2	0.0481	6.07
Asp–Ala	14	48	4	2	0.217	3.05
Glu–Ala	14	48	4	2	0.217	3.05

physicochemical properties of amino acids. In order to calculate this probability, we first of all need an index to measure the similarity or dissimilarity between the amino acids. Since we can expect that such an index must correlate with some measure of evolutionary acceptability [7], *i.e.* the substitution of one amino acid with another, the choice seems to fall on Grantham's [30] index, because it correlates well with the relative substitution frequencies of amino acids. Since this index is the combination of three amino acid properties (polarity, molecular volume and composition), two of which have been recognized as being important in the organization of the genetic code [21, 25], it seems at present to be the best index available in literature and could potentially be used in the below analysis.

The probability associated to a given pair of precursor-product amino acids which is the expression of their physicochemical properties, must be affected by the number of times that this particular pair of amino acids is selected in the genetic code, as this number informs us of the importance that selection has attributed to a given precursor-product pair.* The coevolution hypothesis is verified on the notion of contiguous codons between the precursor and the product [5] and not on the number of times that a given precursor changes into a product, when its codons undergo a single base change. It is clear that in order to compare the two evolutionary hypotheses, we must also consider the case in which the weight to be attributed to the physicochemical distance of the precursor-product pair is based on the number of contiguous codons between the precursor and its product, and not only on the number of times that two interchange.**

---

* As I have said above from a general point of view the probability associated to a given pair of precursor-product amino acids must reflect an aspect of the genetic code, indicating the importance that the choice of that pair has had in the organization of the code. For instance, for the precursor-product pairs Val-Leu and Ser-Trp it seems natural to believe that this probability must be affected by the fact that in the genetic code the Val-Leu pair is chosen 6 times while the Ser-Trp is chosen only once. Therefore, in the evolution of the code the Val-Leu pair must have been more important than Ser-Trp, and this must be reflected in the probability. Moreover, it does not seem relevant to think that the number of times that the code chooses a given pair of amino acids (or equivalently, the number of contiguous codons; see footnote **), must be normalized compared to an expected value. (This expected value could be determined: (i) on the basis of the total number of times that a given amino acid interchanges with other amino acids, or (ii) compared to other and not better specified characteristics of the genetic code.) In this case the choice made during the evolution of the code would be replaced with another number that, even taking other aspects of the code into account, could not be more correct than the one given by the number of times that an amino acid interchanges with another. This is because it is this number that acted during the evolution of the code, and it is to this number that a probability, which must in some way be an expression of the physicochemical properties of the amino acids, has to refer. This number represents a link between the pair of amino acids that was determined during the evolution of the code and that conditioned its organization. Moreover, its substitution with another normalized number, which is abstract to some extent, cannot be considered because in this context it would not reflect the actual organization of the genetic code. Finally, it must be stressed that the probabilities calculated by Wong [5] depend in the final analysis, almost exclusively on the number of contiguous codons between the two amino acids. Therefore, a probability reflecting the physicochemical properties of the amino acids and which also incorporates an aspect of the genetic code, must also be affected by the same weights, at least for comparative purposes as in this paper.

** In actual fact all the calculations have been carried out in conditions that are meaningful for the coevolution hypothesis. The use of hypergeometric distribution (Epn. (1)) on the set of codons contiguous to a given precursor, is meaningful for providing a statistical base for the coevolution hypothesis, but has no real biological meaning. On the other hand, the mutational or translational structure of the genetic code does have a biological meaning. In other words, for the Ser-Cys pair for instance, the important thing from a biological point of view is not the fact that 2 Cys codons are contiguous to the Ser codons, but the fact that Ser (Cys) transforms into Cys (Ser) 4 times, when its codons undergo single mutations or translation errors. The two evolutionary hypotheses should, therefore, be compared in the more appropriate field to the mutational (translational) structure of the genetic code. I have performed these calculations, *i.e.* I have extended the use of hypergeometric distribution to the set of all 392 changes of sense defined in the genetic code after redefining the parameters. And in parallel, in the Z test (see below) I have used the number of times that a given precursor transforms into the product (on the base of the genetic code) as a weight to be attributed to the distance between the precursor and its product. I have obtained a similar result to the ones referred to below, even if the differences in the probability values associated to the χ^2 values are more pronounced. These results are not given because they reach the same conclusions, albeit more markedly, as those shown in the paper.

The probability associated to a given precursor-product pair is calculated by Z test [31]. In order to understand why this test is used, we must first make some considerations.

Regardless of the mechanism determining the allocation of amino acids in a precursor-product relationship, if in that mechanism the physicochemical properties of the amino acids were involved (here measured using Grantham's [30] distances), then the only distances involved in this allocation were the 37 distances between the precursor and its product. If the contiguity between the codons of the precursor and those of the product was, for example, determined by a competition mechanism between these two amino acids so as to be charged onto primordial tRNAs [5, 10], then the distances in which the precursor and product amino acids are not involved would not have been at all able to affect the contiguity between these two amino acids. Hence the question we must answer is: what is the probability that the characteristic distance of the precursor-product pair (when weight with the number of contiguous codons between these two amino acids) was extracted from the population of the 37 distances in which the precursor and the product amino acid participate, and not from the population of the 190 distances in which all twenty amino acids are involved?

This is an unusual but favorable statistical situation. The above arguments help us understand that the 37 physicochemical distances, in which the precursor and its product are involved, do not represent the sample but are the population of all the interactions (at least of those that have been significant in the definition of the genetic code) between the precursor-product and the molecules, such as tRNA, which may have determined the position occupied by these amino acids in the genetic code. Since we can identify the main statistical indices of this population, we are in the most advantageous position to apply the Z test [31]. More specifically, by calculating the mean and variance of the 37 distances, we can ask with what probability the distance of the precursor-product pair was extracted from this population of distances, when its numerousness is given by the number of contiguous codons between these two amino acids.

Thus, the value of the precursor-product distance is considered as a sample extracted from the population of the 37 distances. We can, therefore, measure whether or not the precursor-product distance can be regarded as being indicative of the population of the 37 distances representing its natural control. For instance, in the case of the Val-Leu pair, the mean and standard deviation of the 37 distances [30] in which Val and Leu participate, are equal to 93.784 and 48.448 respectively, and since the number of contiguous codons is 6, the value of the standardized normal variable (Z) is $Z = (D_{\text{Val, Leu}} - \mu)/(\sigma/(x)^{1/2}) = (32 - 93.784)/(48.448/(6)^{1/2}) = -3.1237$, where 32 is the value of the Grantham's distance between Val and Leu. From the value $Z = -3.1237$ we get the required probability: $P(Z \leqslant -3.1237) = 8.929 \times 10^{-4}$. This probability indicates that the Val-Leu distance when weighted with a frequency of 6, cannot be regarded as being representative of the population of the 37 distances, and hence physicochemical selective pressures must have acted on the choice of this pair.

The results of this analysis are shown in Table II. For the eight pairs analyzed by Wong [5] we get $P = 1.7 \times 10^{-6} (\chi^2 = 56.89$, $df = 16)$ and for the remaining seven pairs we get $P = 0.00894$ ($\chi^2 = 29.50$, $df = 14$); while the probability for the set of all fifteen amino acid pairs is 2.3×10^{-7} ($\chi^2 = 86.39$, $df = 30$; Table II).

Discussion

In discussing these results it become immediately clear that the coevolution hypothesis [5] does not appear to be based on sufficiently solid foundations, since all the χ^2 values that can be associated to this theory are systematically below those derived from the physicochemical postulates [15, 16]. Specifically, the χ^2 probability of the set of the fifteen pairs of amino acids derived from the physicochemical postulates is about 170 times smaller (about 80 times smaller for the eight main pairs) than the one calculated for the coevolution hypothesis. And even if with the coevolution hypothesis there should be a certain extension, albeit limited, of the minimization of the physicochemical distances in the genetic code [7], may be relevant to ask the following questions. How can a subsidiary selective mechanism, the physicochemical one, as specified in the coevolution hypothesis [7], give probability lower than those obtained from the assumed main theory? Or are the two evolutionary hypotheses perhaps aspects of the same phenome-

Table II. μ and σ are the mean and the standard deviation, respectively, of the 37 distances [30] in which the precursor and its product are involved. D_{ij} is the value of the Grantham's [30] distance between the two amino acids; x indicates the number of contiguous codons between the precursor and the product and this number represents numerousness in the Z test [31]. $Z_{obs.} = (D_{ij} - \mu)/(\sigma/(x)^{1/2})$ and $P(Z \leqslant Z_{obs.})$ is the probability of observing that particular Z value, representing the normal standardized variable: N(0,1).

Amino acid pairs	μ	σ	D_{ij}	x	$Z_{obs.}$	$P(Z \leqslant Z_{obs.})$	$-2 \cdot \ln P$
Ser−Trp	111.865	45.145	177	1	+1.4428	0.9255	0.15
Ser−Cys	141.703	48.510	112	2	−0.8659	0.1933	3.29
Val−Leu	93.784	48.448	32	6	−3.1237	8.929×10^{-4}	14.04
Thr−Ile	85.000	43.820	89	3	+0.1581	0.5628	1.15
Gln−His	81.649	33.918	24	2	−2.4037	0.0081	9.63
Phe−Tyr	99.486	51.207	22	2	−2.1400	0.0162	8.25
Glu−Gln	89.757	38.750	29	2	−2.2174	0.0133	8.64
Asp−Asn	111.594	45.254	23	2	−2.7686	0.00282	11.74
Thr−Met	83.162	41.377	81	1	−0.0522	0.4792	1.47
Ala−Ser	102.432	36.853	99	4	−0.1862	0.4261	1.71
Gly−Ser	108.703	37.514	56	2	−1.9868	0.0235	7.50
Ala−Gly	105.594	36.785	60	4	−2.4789	0.0066	10.04
Glu−Asp	107.459	45.716	45	2	−1.9322	0.0267	7.25
Asp−Ala	107.351	43.110	126	2	+0.6118	0.7296	0.63
Glu−Ala	97.135	39.286	107	2	+0.3551	0.6387	0.90

non? However, Wong's [5] results and those mentioned so far in this paper need to be clarified.

One possible interpretation may be given by the fact that the physicochemical properties of the precursor-product amino acids, measured using Grantham's index are preserved along the biosynthetic pathways linking them, even if a priori there is nothing that states that this should occur. I have, therefore, calculated the partial (linear) correlation coefficient [32] for the fifteen pairs of amino acids between the $-2 \cdot \ln P$ quantities (Table I and II), while keeping constant, and thus eliminating, the influence of the x parameter values that link the $-2 \cdot \ln P$ variables. This gives a partial correlation coefficient of $R = +0.676$ ($P < 0.005$, $t = 3.177$, $df = 12$). The correlation is significant even when we consider only the eight main amino acids pairs. This proves that Grantham's distances are, at least partly, preserved along the biosynthetic pathways linking the precursor to the product amino acid.

Wong's [5] observations might, therefore, be due to the fact that the physicochemical properties of the amino acids are preserved along the biosynthetic pathways linking them. Furthermore, if the main selective pressure determining the organization of the genetic code, was the one invoked by the physicochemical postulates [15, 16], then this pressure might not only have affected the organization of the code, but also have determined the choice of the biosynthetic pathways linking the various pairs of amino acids.* Moreover, even if coevolution between the biosynthetic pathways and the genetic code is not questioned here, we are trying to identify the selective pressure that was mainly responsible for the organization of the genetic code.

The coevolution hypothesis identifies in the extension of the chemical varieties of the amino acid lateral chains the main selective pressure determining the organization of the genetic code [6, 33]. If, therefore, the introduction of the new amino acids occurred through the concession of codons from the precursor to the product amino acid then (and so that there is a selectable extension of the chemical varieties of the protein molecule lateral chains),

---

* The correlation found between the $-2 \cdot \ln P$ quantities ($R = +0.676$, $P < 0.005$) may be interpreted differently. The pairs of amino acids for which there is a low probability that they are involved by chance (on the basis of the genetic code), are characterized by having very similar physicochemical properties. Therefore, this association might have little to do with the precursor-product relationships, but could be a manifestation of the non-random distribution of the physicochemical properties in the genetic code.

we can expect that the amino acids in the precursor-product relationship must have dissimilar physicochemical properties. Let us consider two developing biosynthetic pathways that link the same precursor to two products, one similar and the other dissimilar to the precursor. The selective advantage of fixing the biosynthetic pathway that links the precursor to its dissimilar product, must have been greater than that fixing the competing pathway that links the precursor to a similar product. The reason for this is, if the main selective pressure was indeed the one that increased the variety of the amino acid lateral chains, then the introduction of a product that was dissimilar to its precursor would certainly have caused a greater increase in the chemical varieties of the amino acid lateral chains, compared to that made by a product similar to its precursor. This view helps us understand that in order to fully verify the coevolution hypothesis, the correlation coefficient between the $-2 \cdot \ln P$ quantities must have been significant and negative, *i.e.* contrary to observations. This, implies the selection of dissimilar precursor-product pairs in the physicochemical properties, which would have rapidly expanded the chemical varieties of the lateral chains of the proteins, thus improving their catalytic efficiency.

Table III shows a code obtained from the genetic code through permutations of amino acids, for which the above conditions hold true. This code has the same fifteen precursor product amino acid pairs of the actual code. If we apply the hypergeometric distribution (Eqn. (1)) we get, for the fifteen pairs examined in this paper, an χ^2 value of 87.46 which is greater than the corresponding value $\chi^2 = 75.11$, obtained by applying the Z test (data not shown). Moreover, this code (Table III) shows a partial correlation coefficient between the $-2 \cdot \ln P$ quantities (data not shown) that is negative and significant ($R = -0.443$, $t = 1.712$, $df = 12$, $P = 0.05$). The construction of this code merely aims to show the possibility that both the χ^2 value and the sign and significance of the correlation coefficient between the $-2 \cdot \ln P$ quantities could fully verify the coevolution hypothesis. When this code (Table III) is compared to the real code, it seems to highlight the fact that other selective pressures must have been in play. Moreover, although the biosynthetic relationships between amino acids are certainly reflected in the organization of the genetic code, they do not seem to confirm the predictions of the coevolution hypothesis. In fact, these observations seem to show that the selective pressure invoked by this hypothesis (even if so general and hence undoubtedly in play) should have produced a genetic code for the type shown in Table III. This table shows that the concession of codons from precursor to product amino acids is more manifest compared to the physicochemical postulates.

Table III. The code obtained from the actual genetic code by simple permutation of the amino acids, that is keeping the blocks of synonymous codons unchanged as in the actual code. See text for the meaning of this code. Ter-chain termination signal.

UUU Glu	UCU Ile	UAU Phe	UGU Gln
UUC Glu	UCC Ile	UAC Phe	UGC Gln
UUA Ala	UCA Ile	UAA Ter	UGA Ter
UUG Ala	UCG Ile	UAG Ter	UGG Lys
CUU Ala	CCU Met	CAU Tyr	CGU Gly
CUC Ala	CCC Met	CAC Tyr	CGC Gly
CUA Ala	CCA Met	CAA Val	CGA Gly
CUG Ala	CCG Met	CAG Val	CGG Gly
AUU Trp	ACU Thr	AAU Pro	AGU Ile
AUC Trp	ACC Thr	AAC Pro	AGC Ile
AUA Trp	ACA Thr	AAA Cys	AGA Gly
AUG Ser	ACG Thr	AAG Cys	AGG Gly
GUU Asp	GCU Arg	GAU Asn	GGU His
GUC Asp	GCC Arg	GAC Asn	GGC His
GUA Asp	GCA Arg	GAA Leu	GGA His
GUG Asp	GCG Arg	GAG Leu	GGG His

Finally, the low probability ($P = 1.7 \times 10^{-6}$, $\chi^2 = 56.89$; Table II) associated to the eight main pairs of precursor product amino acids could indicate that for these pairs there might have been a competition mechanism between the precursor and the product to be charged onto primordial tRNA [5, 10]. This would entail the lack of pairs of amino acids in the genetic code in which the precursor amino acid is involved but has no precursor product relationships and, for which, a lower probability than 1.7×10^{-6} cannot be calculated. This does not agree with the following result. In the genetic code we can choose eight pairs of amino acids in which the precursor amino acid is involved (Ser-Asn, Ser-Thr, Val-Ile, Thr-Pro, Gln-Lys, Phe-Leu, Glu-Lys, Asp-His) and for which we get $P = 1.8 \times 10^{-7}$ ($\chi^2 = 62.70$, $df = 16$; data not shown). This probability is about 9 times lower than the one referring to the eight main pairs of precursor-product amino acids (Table II). However, even if these probabilities are about the same magnitude and, therefore, such as not to definitively exclude the competition mechanism between the precursor product amino acids to be charged on the primordial tRNA; this competition mechanism would have been based not on an amino acid distance index but rather on an amino acid competition index. There is some evidence that this amino acid competition mechanism was not one of the main factors determining the organization of the genetic code [34].

In my opinion the analysis made in this paper better highlights the relationships between the coevolution and the physicochemical hypothesis. Furthermore, this analysis seems to have identified in the preservation of the physicochemical properties of the amino acids in a precursor-product relationship, a statistical significance factor higher than the one that can be associated to the precursor-product relationships.

In conclusion, Wong's [5] observation and those shown here can be regarded as being two aspects of the same phenomenon. In this case, since the theories discussed above are the results of the same selective pressure, they cannot be easily differentiated even with the comparison used in this paper, and therefore the evaluation of the relative importance of the two evolutionary hypotheses [7] in determining the organization of the genetic code, could lose some of its significance. This seems to be confirmed by the study performed by Jurka and Smith [35]. These authors stress that in the prebiotic environment the β-turns have become the object of selection and have affected both the evolution of the genetic code and the biosynthetic pathways for amino acids.

The authors [35] also note that the simple codon-ambiguity-reduction model of the genetic code [14] (which does not seem to be essentially different from what has been here reported as the physicochemical hypothesis) may produce, a code, in only two reduction steps that accounts for β-turns, β-sheets and hydrophilic-hydrophobic structures. Thus, the coevolution between the biosynthetic pathways and the organization of the genetic code could be interpreted as a manifestation of the importance of the preservation of amino acid physicochemical properties, which agrees with the main result shown in this paper. This gives a higher statistical significance in the case in which the organization of the genetic code is interpreted from the point of view of physicochemical properties, than in the case where this is seen from the point of view of precursor-product relationships.

[1] T. R. Cech and B. L. Bass, Ann. Rev. Biochem. **55,** 599 (1986).
[2] F. H. C. Crick, J. Mol. Biol. **38,** 367 (1968).
[3] T. R. Cech, Proc. Natl. Acad. Sci. U.S.A. **83,** 4360 (1986).
[4] A. M. Weiner and N. Maizels, Proc. Natl. Acad. Sci. U.S.A. **84,** 7383 (1987).
[5] J. T. F. Wong, Proc. Natl. Acad. Sci. U.S.A. **72,** 1909 (1975).
[6] J. T. F. Wong, Proc. Natl. Acad. Sci. U.S.A. **73,** 2336 (1976).
[7] J. T. F. Wong, Proc. Natl. Acad. Sci. U.S.A. **77,** 1083 (1980).
[8] F. R. Salemme, M. D. Miller, and S. R. Jordan, Proc. Natl. Acad. Sci. U.S.A. **74,** 2820 (1977).
[9] J. T. F. Wong and P. M. Bronskill, J. Mol. Evol. **13,** 115 (1979).
[10] J. T. F. Wong, Proc. Natl. Acad. Sci. U.S.A. **80,** 6303 (1983).
[11] F. Zinoni, A. Birkmann, W. Leinfelder, and A. Boch, Proc. Natl. Acad. Sci. U.S.A. **84,** 3156 (1987).
[12] A. Schon, C. G. Kannagara, S. Gough, and D. Soll, Nature **331,** 187 (1988).
[13] D. Soll, Nature **331,** 662 (1988).
[14] W. M. Fitch and K. Upper, Cold Spring Harbor Sym. Quant. Biol. **52,** 759 (1987).
[15] T. M. Sonneborn, in: Evolving Genes and Proteins (V. Bryson and H. J. Vogel, eds.), pp. 377–397, Academic Press, New York 1965.
[16] C. R. Woese, D. H. Dugre, M. Kando, and W. C. Saxinger, Cold Spring Harbor Sym. Quant. Biol. **31,** 723 (1966).
[17] C. J. Epstein, Nature **210,** 25 (1966).
[18] A. L. Goldberg and R. E. Wittes, Science **153,** 420 (1966).
[19] M. V. Volkenstein, Biochim. Biophys. Acta **119,** 421 (1966).
[20] C. Allf-Steinberger, Proc. Natl. Acad. Sci. U.S.A. **64,** 584 (1969).
[21] J. R. Jungck, J. Mol. Evol. **11,** 211 (1978).
[22] A. L. Weber and J. C. Lacey, Jr., J. Mol. Evol. **11,** 199 (1978).
[23] R. V. Wolfenden, C. C. F. Cullis, and C. C. B. Southgate, Science **206,** 575 (1979).
[24] M. Sjostrom and S. Wold, J. Mol. Evol. **22,** 272 (1985).
[25] M. Di Giulio, J. Mol. Evol. **29,** 191 (1989).
[26] M. Di Giulio, J. Mol. Evol. **29,** 288 (1989).
[27] N. L. Johnson and S. Kotz, Urn Models and their Application, Chapter 2, pp. 79–83, John Wiley & Sons, New York, London, Sydney, Toronto 1977.
[28] R. A. Fisher, Statistical Methods for Research

Workers, 11th ed., p. 99, Oliver & Boyd, Edinburgh, London 1950.
[29] H. O. Lancaster, The chi-Squared Distribution, p. 21, John Wiley & Sons, New York, London, Sydney, Toronto 1969.
[30] R. Grantham, Science **185,** 862 (1974).
[31] L. N. Balaam, Fundamentals of Biometry, Chapter 6, George Allen & Unwin (eds.), London 1972.
[32] D. A. Kenny, Correlation and Casuality, pp. 54–56, John Wiley & Sons, New York, London, Sydney, Toronto 1979.
[33] J. T. F. Wong, TIBS **6,** 33 (1981).
[34] M. Di Giulio, The Evolution of the Genetic Code: A Comparison between some Theories (unpublished results, available on request).
[35] J. Jurka and T. F. Smith, J. Mol. Evol. **25,** 15 (1987).