A WORKING HYPOTHESIS ON THE INTERDEPENDENT GENESIS OF NUCLEOTIDE BASES, PROTEIN AMINO ACIDS, AND PRIMITIVE GENETIC CODE*

FUJIO EGAMI

Mitsubishi-Kasei Institute of Life Sciences, Minamiooya, Machida-shi, Tokyo 194, Japan

(Received 15 December, 1980; in revised form 6 May, 1981)

Abstract. In the course of experimental approach to the chemical evolution in the primeval sea, we have found that the main products from formaldehyde and hydroxylamine are glycine, alanine, serine, aspartic acid etc., and the products from glycine and formaldehyde are serine and aspartic acid. Guanine is found in the two-letter genetic codons of all these amino acids.

Based upon the finding and taking into consideration the probable synthetic pathways of nucleotide bases and protein amino acids in the course of chemical evolution and a correlation between the two-letter codons and the number of carbon atoms in the carbon skeleton of amino acids, I have been led to a working hypothesis on the interdependent genesis of nucleotide bases, protein amino acids, and primitive genetic code as shown in Table I.

Protein amino acids can be classified into two groups: Purine Group amino acids and Pyrimidine Group amino acids. Purine bases and Pyrimidine bases are predominant in two-letter codons of amino acids belonging to the former and the latter group respectively.

Guanine, adenine, and amino acids of the Purine Group may be regarded as synthesized from C_1 and C_2 compounds and N_1 compounds (including C_1N_1 compunds such as HCN), probably through glycine, in the early stage of chemical evolution.

Uracil, cytosine, and amino acids of the Pyrimidine Group may be regarded as synthesized directly or indirectly from three-carbon chain compounds. This synthesis became possible after the accumulation of three-carbon chain compounds and their derivatives in the primeval sea.

The Purine Group can be further classified into a Guanine or (Gly + nC_1) Subgroup and an Adenine or (Gly + nC_2) Subgroup or simply nC_2 Subgroup. The Pyrimidine Group can be further classified into a Uracil or $C_3C_6C_9$ Subgroup and a Cytosine or C_5-chain Subgroup (Table I).

It is suggested that the primitive genetic code was established by a specific interaction between amino acids and their respective nucleotide bases. The interaction was dependent upon their concentration in the primeval environments and the binding constants between amino acids and their respective bases.

Based upon chemical evolution studies in the primeval sea (Hatanaka and Egami, 1977; Ochiai *et al.*, 1978; Kamaluddin *et al.*, 1979), chemical and biochemical synthetic pathways for the nucleotide bases and protein amino acids, and an arithmetic analysis of the correlation between genetic codons and number of carbon atoms in the carbon skeleton of amino acids (Egami, 1979a), I have been led to a working hypothesis on the genesis of the bases, amino acids, and the primitive genetic code. Here, for the sake of simplicity, the third letter in the codons is neglected. It was probably of less importance, if any, in the primitive code (Crick, 1968).

A preliminary note was published elsewhere (Egami, 1979b). The present one is a revised hypothesis.

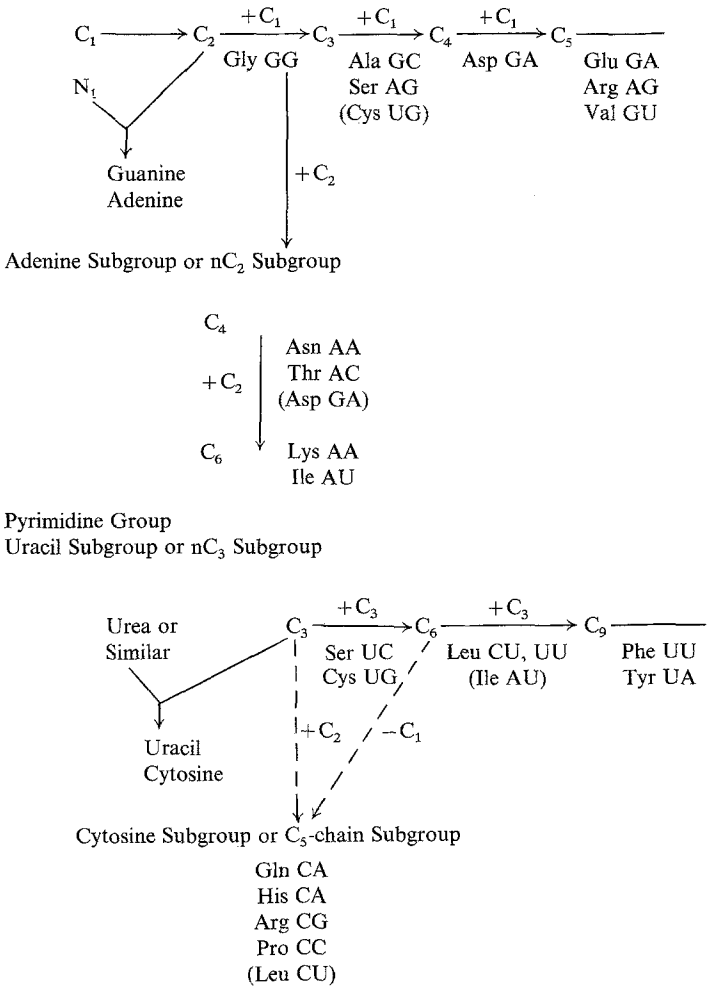
The hypothesis is summarized in a scheme shown in Table I. The background for the hypothesis is the following:

* Presented at the International Symposium (Lipmann Symposium) on 'The Concepts of Chemical Recognition in Biology' held in Grignon near Versailles (France) on July 18–20, 1979.

Origins of Life **11** (1981) 197–202. 0302–1688/81/0113–0197$00.90.
Copyright © 1981 by D. Reidel Publishing Co., Dordrecht, Holland, and Boston, U.S.A.

TABLE I

Classification of amino acids in relation to the genetic code



1. The purine bases (guanine and adenine) and amino acids of Purine Group may be regarded as derived from C_1 and C_2 compounds and N_1 compounds (including such compounds as HCN) probably through glycine as an intermediate. This could have occurred in the early stages of chemical evolution.

(a) We have observed that a series of amino acids were produced from formaldehyde and hydroxylamine in modified sea medium enriched with transition element ions (Hatanaka and Egami, 1977). The predominant amino accids produced from formaldehyde and hydroxylamine in the medium were glycine, alanine, serine, as-

partic acid, glutamic acid, β-alanine and α-aminobutyric acid, (Ochiai *et al.* 1978; Kamaluddin *et al.* 1979). Serine and aspartic acid were produced from glycine and formaldehyde in the same medium (Kamaluddin *et al.*, 1979; Yanagawa *et al.* unpublished). C_1 compounds and N_1 compounds should have been abundant in the early primeval sea and glycine would have been produced at first, and from it Ala, Asp, etc were produced by stepwise addition of one carbon atom. This group may contain Gly GG, Ala GC, Ser AG, Asp GA, Glu GA, etc. We see that G and A, especially G. predominate in these codons. I designate this group the Guanine or (Gly + nC_1) Subgroup of Purine Group.

(b) Probably in the relatively early course of chemical evolution a C_4-amino acid precursor was synthesized through the addition of a C_2 compound to glycine and it initiated the synthesis of α-aminobutyric acid and amino acids of Adenine or (Gly + nC_2) Subgroup of Purine Group. This subgroup may contain Thr AC, Asn AA, Lys AA, and Ile AU. Ile has a C_6 chain composed of a C_4 main chain and a C_2 side chain.

(c) As stated, G and A predominate in the codons of the amino acids of the Purine Group. The first letters are almost exclusively purine bases. This may relate to the similarity between the primitive synthetic pathways for these amino acids and that for the purine bases. As is wel-known, purine bases are synthesized in extant organisms from glycine and C_1 components. As Oró has found, adenine can be chemically synthesized easily from HCN (Oró, 1961), and guanine was from HCN tetramer (Sanchez *et al.*, 1966). In any case the synthesis of purine bases does not require a ready-made C_3-chain. Thus it is highly probable that these purine bases could be synthesized from C_1 compounds and N_1 compounds and accumulated in the early primeval sea together with the amino acids of Purine Group. The concentration of pyrimidine bases must have been much lower than that of purine bases in the early primeval sea.

2. Pyrimidine bases (uracil and cytosine) and amino acids of the Pyrimidine Group may be regarded as derived from compounds with C_3 – or longer chains and N_1 compounds. The accumulation of such compounds in the primeval sea occurred either later than the predominant accumulation of C_1 and C_2 compounds in the primeval sea (chronological necessity) or in a specific part, where the concentration ratio C/N was much higher and such chain compounds were accumulated (topographic necessity). In other words, the accumulation of pyrimidine bases and amino acids of Pyrimidine Group depended upon the accumulation of compounds with longer carbon chains.

(a) We have observed that amino acids with C_3-chains can be produced in higher yields from C_3 compounds (unpublished observation). I believe that in the course of chemical evolution, after the accumulation of C_3 compounds, C_6 and C_9 compounds were produced from the C_3 compounds. Thus $C_3C_6C_9$ amino acids were produced.

These amino acids are Ser UC, Cys UG, Leu UU, CU, Phe UU, and Tyr UA. I designate this group the Uracil or $C_3C_6C_9$ Subgroup of Pyrimidine Group.

It should be added that the astrophysical role of C_3 compounds, especially tricarbon dioxide, was recently discussed by Shimizu (1979) in connection with the genetic code.

(b) The group designated as the Cytosine of C_5-chain Subgroup contains Gln CA, His CA, Arg CG, and Pro CC. It is rather remarkable that all these amino acids contain C as the first letter of codons in spite of the quite different structures and properties. Although the origin of the C_5-chain is obscure, it may be derived from a C_6-chain by the loss of one carbon or from a C_3-chain by the addition of two carbons. Leu with a C_6-chain can be regarded either as composed of two C_3-chains (UU) or as a C_5 main chain with a one-carbon side-chain (CU).

(c) U and C predominate in the codons of amino acids with C_3, C_6, and C_9-chains and these with C_5-chain respectively. The first letters are almost exclusively pyrimidine bases. These amino acids can be classified as amino acids of the Pyrimidine Group. Again there is a similarity between the synthetic pathway of these amino acids and that of the corresponding pyrimidine bases. Pyrimidine bases are generally synthesized from urea and C_3-chain compounds in organic chemistry. In their biosynthesis the C_3-chain in pyrimidines derives from aspartic acid by subsequent loss of CO_2. It is reasonable to assume that during chemical evolution pyrimidines were synthesized from urea or something like urea such as guanidine and C_3-chain compounds, as reported by Ferris *et al.* (1974).

3. It is suggested that the primitive genetic code was established through specific interaction between amino acids and respective nucleotide bases. The interaction was dependent not only upon the binding constants between amino acids and their respective nucleotide bases, but also upon the concentration of these in the primeval environments (Mikelsaar, 1975). Thus purine rich codons and pyrimidine rich codons were established.

Additional Comments

1. The codon for Val (GU) suggests that it was synthesized from glycine not by the stepwide addition of one carbon atom but by the addition of a C_3 compound and has a conjugated carbon chain belonging to both Guanine and Uracil Subgroups.

$$\text{Cly (G)} + C_3\text{ (U)} \rightarrow \text{Val (GU)}$$

The codon of Cys (UG) might be related to its intimate relation to both the Uracil Subgroup and the Guanine Subgroup.

Threonine belongs to the Adenine or (Gly + nC_2) Subgroup. It is consistent with the finding that it can be synthesized from glycine and acetaldehyde through copper complexes (Sato *et al.*, 1957). We have also observed the formation of threonine from glycine and acetaldehyde in our modified sea medium (unpublished). It means that threonine should be regarded not as a product of Gly + $2C_1$, but as a product of Gly + C_2. The route is different from that suggested by Dillon (1978).

2. It should be pointed out that Trp and Met are not included in the present scheme. They might be incorporated much later into the primitive protein, probably in the course of early biological evolution. However, their two-letter codons are not inconsistent with the present hypothesis: Met with a C_4-chain AU, Trp with a conjugated chain.

$$[C_6\text{ (U)} + \text{Gly (G)}] + C_3\text{ (U)} \rightarrow \text{Trp (UG)}$$

3. His and Gln, Asn and Lys, and Asp and Glu have the same two-letter codons, CA, AA, and GA respectively. This may be due to the similar 'polar requirement' (Woese *et al.*, 1966) and the probable common prebiotic synthetic pathways – as may be predicted from the present hypothesis – of two amino acids of each group. They could not be distinguished because of similar physico-chemical properties.

4. Many speculations have been published on the genesis of the genetic code (Reviews to be consulted: Woese, 1967: Yčas, 1969; Jukes, 1978; Dillon, 1978). Woese (1967) classified the various hypotheses into stochastic models and mechanistic models. In mechanistic models the affinity or binding constant between amino acids and respective nucleotide codons or anticodons was taken into consideration for the interaction between them (Woese, *et al.* 1966; Woese, 1967; Lacey and Pruitt, 1969; Saxinger and Ponnamperuma, 1974; Weber and Lacey, 1978; Jungck, 1978). However the interaction depended not only on the affinity, but also on the concentration of both amino acids and respective nucleotides. The physico-chemical experiments so far carried out in order to find out the relationship between amino acid-nucleotide interaction and genetic code have to be reevaluated taking into consideration the probable concentration of amino acids and nucleotides in the primeval sea.

The present hypothesis has the advantage of taking into consideration both chronological or topographic necessity and chemical necessity for the interaction between nucleotide bases and their respective amino acids. It explains the existence of two quite different codons, UC and AG, for serine: Serine could have been produced by the amination of a C_3 compound (UC) and by the addition of a C_1 compound, probably formaldehyde, to glycine (AG). It may be explained in a similar way that arginine has two codons, AG and CG, and that glutamic acid belongs to (Gly + nC_1) Subgroup and glutamine belongs to C_5-chain Subgroup.

Dillon's (1973, 1978) speculation is different from others. It depends on the probable biosynthetic pathways of different amino acids in primitive organisms. However, it is very difficult to explain the origin of codons for all protein amino acids on these sole grounds. The prebiotic synthetic pathways of amino acids such as valine, threonine etc. presented in the present hypothesis are quite different from the biosynthetic pathways suggested by Dillon. Moreover he takes into consideration only an inter-amino acid order, but not a codon-amino acid order, as defined by Woese (1967).

5. The present hypothesis deals only with the genesis of the primitive code. In the course of the evolutionary modification of the genetic code so as to minimize the frequencies of deleterious mutation, what Woese (1967) calls 'stochastic models' might come into play, especially in the determination of the third letter. But the universality of the genetic code leads us to consider that the essential feature of the first and second letters of the codons must have been fixed in the early stage of evolution. The mechanism of the fixation remains unknown. In the course of evolution before the fixation, the first and second letters seems to have had the tendency of sharing the predominant meanings in the codons: the first letter predominantly took charge of the nature of carbon skeleton of respective amino acids, or, in other words, the synthetic pathways at the genesis of amino acids, and the second letter pre-

dominantly took charge, in general, of the selective affinity of nucleotide bases to respective amino acids, or, in other words, the fitness to the structure and physicochemical properties of amino acids.

6. The present codon-anticodon interaction seems to have appeared later and to have initiated the development of the present protein synthesis mechanism. The codon-anticodon interaction had higher affinity than the codon-amino acid interaction and replaced the latter.

In conclusion, a working hypothesis is presented on the interdependent genesis of nucleotide bases, protein amino acids, and the primitive genetic code: the primitive genetic code was dependent upon the concentration of different nucleotide bases and amino acids coexisting in the primeval environments and upon the selective affinity between bases and amino acids. Although there remain some ambiguities on the assignment of several amino acids in the scheme (Table I), I hope that the ambiguities will be gradually eliminated by further studies on the prebiotic syntheses of amino acids and nucleotide bases and the present hypothesis will provide a basis for further studies on the genesis of the genetic code.

References

Crick, F. H. C.: 1968, *J. Mol. Biol.* **38,** 367.
Dillon, L. S.: 1973, *Botanical Review* **39,** 301.
Dillon, L. S.: 1978, *The Genetic Mechanism and The Origin of Life*, Plenum Press, New York and London, pp. 216-230.
Egami, F.: 1979a, *Kagaku (Science, Tokyo)* **49,** 527.
Egami, F.: 1979b, *Nippon Nŏgeikagaku Kaishi (J. Agr. Chem. Soc. Jpn)* **53,** 173.
Ferris, J. P., Zamek, O. S., Altbuch, A. M., and Freiman, H.: 1974, *J. Mol. Evol.* **3,** 301.
Hatanaka, H., and Egami, F.: 1977, *Bull. Chem. Soc. Jpn.* **50,** 1147.
Jukes, T. H.: 1978, *Adv. Enzym.* ed. by A. Meister, **47,** 375.
Jungck, J. R.: 1978, *J. Mol. Evol.* **11,** 211.
Kamaluddin, Yanagawa, H., and Egami, F.: 1979, *J. Biochem. (Tokyo)* **85,** 1503.
Lacey, J. C. and Pruitt, K. M.: 1969, *Nature* **223,** 799.
Mikelsaar, H. N.: 1975, *J. Theor. Biol.* **50,** 203.
Ochiai, T., Hatanaka, H., Ventilla, M., Yanagawa, H., Ogawa, Y., and Egami, F.: 1978, in H. Noda (ed.), *Origin of Life*, Proc. 5th Intern. Cong. Origin of Life, Center Acad. Publ. Jpn. pp. 135-139.
Oró, J.: 1961, *Nature* **191,** 1193.
Sanchez, R. A., Ferris, J. P., and Orgel, L. E.: 1966, *Science* **153,** 72.
Sato, M., Okawa, K., and Akabori, S.: 1957, *Bull. Chem. Soc. Jpn.* **30,** 937.
Saxinger, C. and Ponnamperuma, C.: 1974, *Origins of Life* **5,** 190.
Shimizu, M.: 1979, *Astrophys. Space Sci.* **62,** 509.
Weber, A. L. and Lacey, J. C.: 1978, *J. Mol. Evol.* **11,** 199.
Woese, C. R.: 1967, *The Genetic Code. The Mol. Basis for Gen. Exp.*, Harper and Row. Publ., New York, Evanston, and London.
Woese, C.R., Dugre, D. H., Saxinger, W. C., and Dugre, S. A.: 1966, *Proc. Nat. Acad. Sc. USA* **55,** 966.
Yčas, M.: 1969, *The Biological Code*, North-Holland Publ., Amsterdam.