J Mol Evol (1981) 17:273–284

Journal of Molecular Evolution

Reasons for the Occurrence of the Twenty Coded Protein Amino Acids

Arthur L. Weber[1] and Stanley L. Miller[2]

1 The Salk Institute for Biological Studies, San Diego, California 92138, USA
2 Department of Chemistry, University of California, San Diego, La Jolla, California 92093, USA

Summary. Factors involved in the selection of the 20 protein L-α-amino acids during chemical evolution and the early stages of Darwinian evolution are discussed. The selection is considered on the basis of the availability in the primitive ocean, function in proteins, the stability of the amino acid and its peptides, stability to racemization, and stability on the transfer RNA. We conclude that aspartic acid, glutamic acid, arginine, lysine, serine and possibly threonine are the best choices for acidic, basic and hydroxy amino acids. The hydrophobic amino acids are reasonable choices, except for the puzzling absences of α-amino-n-butyric acid, norvaline and norleucine. The choices of the sulfur and aromatic amino acids seem reasonable, but are not compelling. Asparagine and glutamine are apparently not primitive. If life were to arise on another planet, we would expect that the catalysts would be poly-α-amino acids and that about 75% of the amino acids would be the same as on the earth.

Key words: Amino acids–Molecular evolution–Genetic Code–Protein synthesis–Prebiotic synthesis

There are only twenty amino acids that are coded for in protein synthesis, along with about 120 that occur by post-translational modifications (Uy and Wold 1977). Yet there are over 300 naturally-occurring amino acids known (Mooz 1976), and thousands of amino acids are possible. The question then is – why were these particular 20 amino acids selected during the process that led to the origin of the most primitive organism and during the early stages of Darwinian evolution. This problem has not been explicitly discussed previously although there are extensive discussions of the related problem of the origin of the genetic code (Crick 1968; Crick et al. 1976; Woese 1967; Eigen 1978; Jungck 1978; Weber and Lacey 1978). Wong and Bronskill (1979) have pointed out that prebiotic synthesis experiments do not yield equal amounts to the protein amino acids, and that in some cases, e. g. asparagine and glutamine, their concentration in the primitive ocean would have been too low for incorporation into primitive proteins. Although simple abundances in the primitive ocean must have been an important factor, it is clear that the problem is more complex.

The following questions need to be considered.

(1) Why are amino acids used for the construction of biochemical catalysts rather than the corresponding hydroxy acids or other organic compounds?

(2) Why are α-amino acids and not β-, γ-, δ-amino acids incorporated into proteins?

(3) Why do the 20 amino acids have at least one α hydrogen?

(4) Why are N-alkyl amino acids, excluding proline, not among the 20?

(5) Why are only L-amino acids used?

(6) On what basis have the side chains been selcted?

(7) Why are there only 20 amino acids when the triplet genetic code has 64 codons available? Similarly, could the system work effectively with less than 20?

We can think of a number of reasons to explain the selection process.

(a) The heterotrophic hypothesis of the origin of life states that the first organisms used the organic material in the environment (Horowitz 1945). Therefore, the more abundant amino acids obtained by prebiotic synthesis or reactions in the primitive milieu would have a favored incorporation into the first proteins. However, abundance cannot be the only criterion, and selection on

Offprint requests to: Stanley L. Miller

Dedicated to the memory of Harold C. Urey

0022–2844/81/0017/0273/$ 02.40

the basis of abundance in the primitive environment would apply only within classes of amino acids.

(b) A compelling reason to include or exclude an amino acid from proteins would be its suitability or unsuitability for protein structure and function. An example would be the ability to form ordered secondary and tertiary structures. Unfortunately, there is insufficient information on the relationship between structure and function to deal adequately with this possibility. It is clear that the 20 protein amino acids are not uniquely suitable for proteins since norleucine can substitute for methionine in some proteins (Anfinsen and Corely 1969; Old and Jones 1975; Cowie et al. 1959).

(c) The free amino acid and its side chain in a peptide should be adequately stable in solution. If an amino acid decomposes too rapidly on the geological time-scale, its steady state concentration would be too low for effective use in primitive peptides and in the most primitive organisms. An example is aminomalonic acid which decarboxylates irreversibly to glycine + CO_2.

The side chain of an amino acid residue in a protein should not irreversibly react during the lifetime of the peptide backbone. An example would be phosphoserine in a peptide, which decomposes by hydrolysis and β-elimination. At the present time, stability of the amino acids in peptides is not so critical because of rapid synthesis and turnover, but stability would have been more important early in evolution.

(d)The free amino acid or its peptide should be stable to racemization. α-Phenylglycine and α,β-diaminopropionic acid racemize relatively rapidly (Jacobson et al. 1974) and might be unsuitable for this reason.

(e) The peptide of an amino acid should be relatively stable. Thus ornithine peptides are unstable because internal lactamization by the δ-amino group breaks the peptide backbone.

(f) Related to the stability of the peptide is the stability of the aminoacyl ester of tRNA or related primitive activated form. Again the ornithine ester would close rapidly to the lactam.

(g) It is possible that some physical-chemical process separated the amino acids to be incorporated into primitive protein from the non-protein amino acids. Examples of possible processes are solubilities, thermal stability, photochemical stability, selective absorption on clays (Lawless and Levi 1979) or selective binding to particular nucleotide sequences of polynucleotides (Jungck 1978; Reuben and Polk 1980). Although these factors may have played a role in the selection process, it is unlikely that a single process could account for the selection of the 20 amino acids.

(h) It is possible that some of the amino acids were never abundant in the primitive ocean before life arose, but became abundant after self-replicating organisms started to develop biosynthetic pathways.

(i) Some amino acids may not have been selected because they were too difficult to synthesize by early organisms. It is difficult to give a convincing example of this, since the biosynthesis of the aromatic amino acids is sufficiently complex that obvious amino acids such as norvaline cannot be excluded on this basis. However, it may be that the biosynthesis of valine is simpler than norvaline, and that valine was selected on this basis.

(j) There is always the possiblity that the incorporation or exclusion of an amino acid is based on chance or "frozen accident" (Crick 1968). The entire list of 20 protein amino acids could be claimed to be the result of chance, but we would like to appeal to this reason only as the last resort, and as we shall demonstrate, this rationale needs to be applied, if at all, in only a few cases.

In this paper we will discuss the reasons for the presence and absence of a variety of amino acids in addition to the 20 protein amino acids. We will show that the absences can be accounted for in terms of one or more of the above reasons, and that those in proteins are the ones that would be expected, with a few exceptions.

1. Why Are Amino Acids Used for the Construction of Biochemical Catalysts?

In our opinion, the basic reason that amino acids were used was their abundance in the primitive ocean. Most prebiotic experiments produce good yields of amino acids relative to other classes of organic compounds. The relevance ot these experiments to the primitive earth is supported by the amino acid abundances found in the Murchison and other carbonaceous chondrites (see Table 1).

Hydroxy acids are produced along with amino acids in prebiotic experiments (Miller 1957; Peltzer and Bada 1978; Peltzer 1979) although they are not usually looked for. Hydroxy acids are polymerized by the ribosomal system to give polyesters (Fahnestock and Rich 1971; Rich 1971). However, these polyesters differ from peptides in that they are relatively unstable to basic hydrolysis. Since the ester linkage is not planar and cannot form intrastrand hydrogen bonds, some structures available to peptides would not form with polyesters.

Many other polymers are possible that could serve as catalysts. An example would be variations on Nylon 66--that is, copolymers of dicarboxylic acids and diamines. Dicarboxylic acids are produced in fair yield in electric discharge experiments (Miller 1957: Zeitman et al. 1974: Peltzer 1979) but diamines have been looked for and not found (Van Trump and Miller, unpublished). However, diamines would be expected in the primitive ocean from the slow decarboxylation of diamino acids.

The head-tail polymerization mechanism of amino acids appears to be intrinsically simpler than that for diamines and dicarboxylic acids. In head-tail polymerization, each addition of a new monomeric unit to the

growing end yields a new terminus that possesses the same functional group as the previous terminus. This constancy of structure at the growing end is ideally suited to enzymatic catalysis. In contrast, the chemical character of the growing end of a Nylon 66-like polymer alternates between two types of functional groups. The biological exploitation of the head-tail mechanism is not limited to polypeptides, but is also found in biosynthesis of polynucleotides, polysaccharides and fatty acids. Head to head condensations usually stop at "dimers", e.g. sucrose, nicotinamide adenine dinucleotide (NAD).

2. Why Are β, γ and δ Amino Acids Absent?

The selection of α-amino acids for protein synthesis and the exclusion of the β, γ and δ amino acids by evolutionary processes raises two questions. First, why does protein synthesis use only one type of amino acid and not a mixture of various α, β, γ and δ acids? Second, why were the α-amino acids selected? Although an early translation system may have polymerized mixtures of α, β, γ and δ amino acids, the present ribosomal peptidyl transferase has an evolved specificity for only α-amino acids. Compounds with a more remote amino group reportedly do not function in the peptidyl transferase reaction (Krayevsky and Kukhanova 1979). In order to polymerize optimally a mixture of α- and β-amino acids, four enzymes would probably be required. One enzyme would be needed for each of the four combinations of substrates: α, α; α, β; β, α and β, β. In a related manner the ribosomal peptidyl transferase has evolved a specificity for L-α-amino acids (Nathans and Neidle 1963; Rychlik et al. 1970; Krayevsky and Kukhanova 1979), which may account for the use of a single optical isomer in protein amino acids.

The chemical basis for the selection of α-amino acids can be understood by considering the deleterious properties that β, γ and δ-amino acids give to peptides or have for protein synthesis. A basis for the exclusion of β-amino acids is suggested by the reported absence of an ordered secondary structure for poly-β-alanine and poly-β,L-aspartic acid in water (Glickson and Applequist 1971; Balasubramanian 1974). The apparent instability of ordered structures for these poly-β-amino acids was attributed to an increase in the polymer's entropy that is caused by internal rotation about the C^{α}-C^{β}-bond. The failure of poly-γ,L-glutamic acid to form an ordered secondary structure has also been attributed to greater conformational freedom caused by internal rotations about the C^{α}-C^{β} and C^{β}-C^{γ} bonds (Balasubramanian et al. 1973; Kushwaha et al. 1980). Although there are no conformational studies of poly-δ-amino acids, poly-ϵ,L-lysine has been reported to form an ordered structure that is similar to the antiparallel pleated-sheet conformation of proteins (Kushwaha et al. 1980). Taken

Table 1. Molar ratios of amino acids found in an electric discharge synthesis and the Murchison meteorite (glycine = 100)

	Electric discharge[a]	Murchison meteorite[b]
Glycine	100	100
Alanine	180	36
α-Amino-n-butyric acid	61	19
Norvaline	14	14
Valine	4.4	19
Norleucine	1,4	
Leucine	2.6	
Isoleucine	1.1	
Alloisoleucine	1.2	
t-Leucine	< 0.005	
α-Amino-n-heptanoic acid	~ 0.3[c]	
Proline	0.3	22
Pipecolic acid	0.01	11
α,β-diaminopropionic acid	1.5	
α,γ-diaminobutyric acid	7.6	
Ornithine	< 0.01	
Lysine	< 0.01	
Aspartic acid	7.7	13
Glutamic acid	1.7	20
Serine	1.1	
Threonine	~ 0.2	
Allothreonine	~ 0.2	
Methionine	0.1[d]	
Homocysteine	~ 0.5[e]	
Homoserine	~ 0.5[e]	
β-Alanine	4.3	10
β-Amino-n-butyric acid	~ 0.1	5
β-Aminoisobutyric acid	~ 0.1	9
γ-Aminobutyric acid	0.5	7
α-Aminoisobutyric acid	$\sim 7.$	33
α-Methyl-α-amino-n-butyric acid (isovaline)	$\sim 1.$	11
Sarcosine	12.5	7
N-ethyl glycine	6.8	6
N-propyl glycine	~ 0.5	
N-isopropyl glycine	~ 0.5	
N-methyl alanine	~ 3.4	3
N-ethyl alanine	< 0.05	
N-methyl-β-alanine	~ 1.0	
N-ethyl-β-alanine	~ 0.5	
Isoserine	1.2	
α-Hydroxy-γ-amino butyric acid	17.	

[a] From Ring et al. 1972 and Wolman et al. 1972

[b] From Kvenvolden et al. 1970 and Kvenvolden et al. 1971

[c] R.H. White and S.L. Miller, unpublished data. The following isomers were found in comparable amounts and were also identified by GC-MS: 5-Methylnorleucine, 4-Methylnorleucine (both diastereoisomers), 3-Methylnorleucine (both diastereoisomers), 4,4-Dimethylnorvaline and 3-Ethylnorvaline. 3,4-Dimethylnorvaline and 3,3-dimethylnorvaline could not be found probably because they were masked by leucine peaks.

[d] Obtained in a separate experiment (Van Trump and Miller 1972)

[e] Not analyzed for in the spark discharge experiment. The values given are based on the yields expected from the additions of H_2S and H_2O to acrolein which is produced by electric discharges (Schlesinger 1978)

together these reports suggest that polyamides of β, γ, δ, and ϵ amino acids might be able to form ordered secondary structures, but not as easily as poly-α-amino acids.

The γ and δ-amino acids are poor substrates for activation as tRNA-like esters, since their esters undergo lactamization (Hay et al. 1966; Hay and Porter 1967, Hay and Morris 1970, 1972; Martin et al. 1964). The rate of lactamization of methyl γ-aminobutyrate is not unusually rapid compared to hydrolysis, since at pH 9 the yield of lactam is 50%. In the case of δ-amino acid esters the rate of lactamization is much greater than hydrolysis and so its tRNA ester would lactamize before peptide bond formation occurs.

It is an interesting aspect of prebiotic syntheses that α-amino acids are produced in much greater yield than β, γ or δ amino acids (Table 1), and that this factor alone would favor the selection of α-amino acids.

3. Why Do the 20 Amino Acids Have at Least One α-Hydrogen?

α-Amino isobutyric acid is a major prebiotic product as is isovaline (α-methyl-α-amino-n-butyric acid). These amino acids would have the advantage of being very difficult to racemize. However, their biosynthesis presents some difficulties. They cannot undergo transamination and would have to be synthesized by methylation of an amino acid with an α-hydrogen or by the carboxylation of an amine such as isopropylamine.

Replacement of the α-hydrogen by larger substituent, such as a methyl group, would also increase significantly steric hinderance around the amino and carboxyl groups. Steric difficulties have been encountered in the chemical synthesis of peptides with α-aminoisobutyric acid (Leplawy et al. 1960). Also, the severely constrained stereochemical behavior of peptides of α-methyl-α-amino-n-butyric acid (Burgess and Leach 1973; Nagaraj et al. 1979) suggests that peptides composed of α-methyl substituted amino acids would not have the structural versatility of peptides derived from α-hydrogen substituted amino acids. An additional factor involves the ribosomal peptidyl transferase, which would develop a specificity for either an α-hydrogen or an α-methyl of the amino acid. This would prevent the synthesis of proteins containing both types of amino acids.

4. The Absence of N-Alkyl Amino Acids

Amino acids such as sarcosine (N-methyl glycine) and N-methyl alanine are absent from proteins, although they are major prebiotic products. One reason for their absence is that peptide N-H hydrogen bonds are not possible, thereby limiting the types of protein structures. However, helical structures can be formed from poly-N-methyl-L-alanine (Goodman and Fried 1967; Mark and Goodman 1967), although they are different from an α-helix. Poly-N-alkyl peptides also cannot form β-sheets because of the absence of hydrogen bonding.

Proline can be considered as a N-alkyl amino acid (as is pipecolic acid). It has special functions in proteins, and the ring structure has the advantage over open chain "isomers" such as N-ethyl alanine in having less conformers.

Another possible disadvantage of N-alkyl amino acids is that they racemize much more rapidly than N-unsubstituted amino acids under some circumstances, e.g. diketopiperazines (Gund and Veber 1979). This applies to proline in diketopiperazines but apparently not in peptides. There are no data on whether N-alkyl amino acids in linear peptides racemize rapidly at neutral pH values, but if so, this could be a major reason for their absence from proteins.

5. Why Are Only L-Amino Acids Used in Proteins?

This is related to the question of the origin of optical activity in living organisms on which there is a very large literature (Bonner 1972; Norden 1978; Brack and Spack 1980). We do not propose to deal with this question here, except to note that arguments presented in this paper would apply to organisms constructed from either D or L amino acids.

6. On What Basis Were the Side Chains of the α-Amino Acids Selected?

It is assumed that different types of amino acids are needed to achieve the desired functions in proteins and that the selection from the primitive environment could not have been on the basis of abundance alone. Thus it is appropriate to discuss the selection of the amino acids by type – hydrophobic, acidic, basic, hydroxy, sulfur, aromatic and amide.

Hydrophobic Amino Acids

*As shown in Tab*le 1, these amino acids are synthesized under prebiotic conditions in good yield and they occur in carbonaceous chondrites. All the hydrophobic amino acids up to six carbons, except for t-leucine (3,3,3-trimethylalanine), are produced in electric discharge reactions of CH_4, N_2, H_2O and NH_3. There is a fall off in yield of about a factor of 6 on going from the 5 to 6 carbon amino acids and about a factor of 5 on going from the 6 to 7 carbon amino acids. It seems likely the 7-carbon aliphatic amino acids may have had too low an abundance relative to the 5 and 6 carbon amino acids in the primitive ocean to have been incorporated into primitive proteins. Of the 2 to 6 carbon aliphatic amino acids (with α hydrogens and an unsubstituted amino

group), glycine, alanine, valine, isoleucine and leucine occur in proteins, but α-amino-n-butyric, norvaline, norleucine, alloisoleucine and t-leucine are missing.

The absence of t-leucine is reasonable since it apparently was not abundant in the primitive environment. In addition, the non-reactive character of the carbonyl group of activated esters of t-leucine (Pospisek and Blaha 1976; Isumiya et al. 1953) suggests that at least one C^{β}-hydrogen is required for aminolysis of the carbonyl to occur without severe steric hindrance. Ribosomal peptidyl transferase, which catalyzes peptide-bond formation, may not be able to overcome the severe steric hindrance that results from three substituents on the β-carbon.

The absence of alloisoleucine seems plausible. There would be little to gain by using both isoleucine and alloisoleucine, since the differences conveyed to the properties of proteins would be small. The basis for the selection of isoleucine over alloisoleucine is not apparent.

The absence of α-amino-n-butyric acid, norvaline and norleucine is most striking and a major challenge to any attempt to account for the selection of the twenty protein amino acids. Their abundance in prebiotic experiments and carbonaceous chondrites is comparable to or greater than the hydrophobic amino acids used in proteins. This is the only place where we are tempted to appeal to reasons of chance or frozen accident. Indeed they may have been originally coded for and subsequently deleted. However, there are a number of reasons that could account for the absence of these three amino acids.

In the case of norleucine, it could be argued that its structural resemblance to methionine, which may be a more versatile amino acid, resulted in its deletion from proteins. It is interesting to note that norleucine can be incorporated into proteins in place of methionine without loss of enzymatic activity (Anfinsen and Corley 1969; Old and Jones 1975; Cowie et al. 1959).

The selection of branched aliphatic amino acids and the exclusion of linear aliphatic amino acids may be due to the ability of peptides of branched amino acids to form a more ordered tertiary structure. As shown in Fig. 1 the linear side chain of norvaline has an additional axis of rotation relative to valine and this would be expected to increase the mobility of the side chain of norvaline relative to valine. Similar considerations apply to norleucine, leucine and isoleucine. This concept is supported by the spin-lattice relaxation times from ^{13}C-NMR which indicate restricted motion of the γ- and δ-methyl groups of isoleucine and leucine relative to the ε-methyl group of norleucine (Deslauriers et al. 1973). It also seems likely that the movement of branched side chains within a protein would be more easily controlled by steric interference of neighboring groups than movements of linear side chains.

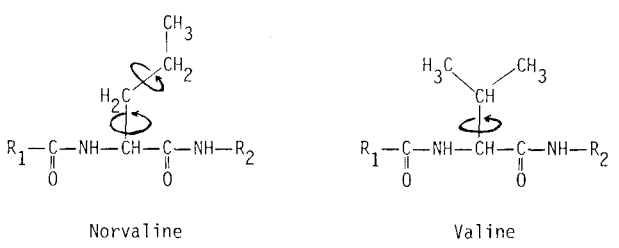


Fig. 1. Movement of hydrocarbon groups about the rotational axes of norvaline and valine

We conclude that the absence of α-amino-n-butyric acid, norvaline and norleucine might be accounted for in terms of protein structure and function, but the case is by no means compelling, and their absence may have another basis such as chance or a frozen accident.

Proline and its Homologs

Of the cyclic amino acids only proline and pipecolic acid, together with methylated derivatives, are reasonable possibilities. The three, four and seven membered rings are unstable and would not be expected in the primitive ocean. In both the carbonaceous chondrites and electric discharge products, proline is more abundant than pipecolic acid (see Table 1). The methylated prolines have not been found in either of these sources. Proline has an advantage over pipecolic acid in that its ring is more rigid and so there would be less structural flexibility in proteins with proline.

Acidic Amino Acids

The selection of aspartic and glutamic acids for utilization in proteins is very logical on a number of grounds, assuming acidic amino acids are needed. These are found in the Murchison meteorite and are formed in good yield in most prebiotic syntheses.

The other choices for acidic amino acids of six carbons or less would be aminomalonic acid, β-methylaspartic acid, α-aminoadipic acid and β- or γ-methylglutamic acids. Aminomalonic acid is a poor choice, since it decarboxylates readily to glycine. The maximum rate is at pH 2 where $H_3N^+-CH(COO^-)COOH$ is the predominate species (Thanassi 1970). The rate constant is 2×10^{-3} min^{-1} ($\tau_{1/2}$ = 0.24 days at 45°C). Assuming no decarboxylation of the anion, which may well occur, the half life at pH 8 and 45°C would be 66 years. Decarboxylation of aminomalonic acid in peptides would also be expected to be rather rapid.

β-Methylaspartic acid has not been found in either the Murchison meteorite or among the electric discharge products. A small amount would be expected

in the primitive ocean, but at levels far below aspartic or glutamic acids. β-Methylaspartic also decomposes to methylfumaric acid to a greater extent than aspartic acid, thereby further lowering its concentration (Bada and Miller 1968).

α-Aminoadipic acid has been found neither in the Murchison meteorite nor among the electric discharge products, so it would appear to have been of low abundance in the primitive ocean. The same applies to β- and γ-methyl glutamic acids.

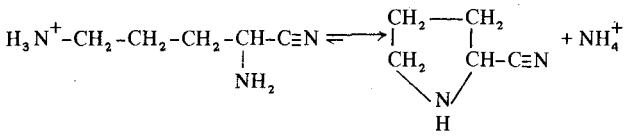
The only factor that would tend to select against glutamic acid is that it forms an equilibrium mixture of glutamic and pyrrolidonecarboxylic acid, containing only 5% glutamic acid (Meister and Bukenberger 1962). The equilibrium is only slowly attained at pH 8 (Wilson and Cannan 1937), but the pyrrolidonecarboxylic acid would probably have been formed first from the nitrile of glutamic acid. The 5% free glutamic acid was probably more than adequate for incorporation into prebiotic peptides, since yields of glutamic acid in prebiotic experiments are quite good.

The Basic Amino Acids

Lysine, arginine and histidine are treated separately.

Lysine. A good case can be made for the incorporation of lysine into proteins if it is assumed that free amino groups (other than the N-terminal amino acid) are needed. The homologs of lysine include α,β-diaminopropionic acid, α,γ-diaminobutyric acid, and α,δ-diaminovaleric acid (ornithine). Diaminopropionic acid and diaminobutyric acid are major products of the electric discharge synthesis. There are no clearly established prebiotic syntheses of ornithine or lysine, nor were any basic amino acids found in the Murchison meteorite (Kvenvolden et al. 1970 and 1971).

The reason that ornithine and lysine are not found in the typical prebiotic synthesis is that these amino acids, if formed from the nitrile, tend to close the ring to give proline and pipecolic nitriles (Miller, unpublished data).

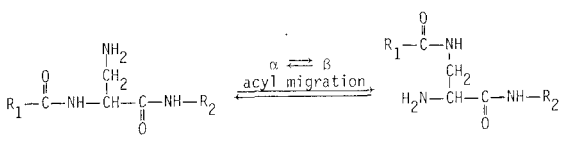


Ring formation is avoided in the biosynthesis of ornithine and lysine by acylation at the appropriate step. Although prebiotic syntheses might be worked out for ornithine and lysine it would seem, on the basis of the present prebiotic synthesis results, that diaminopropionic acid and diaminobutyric acid would have been more abundant than ornithine, and ornithine in turn would have been more abundant than lysine. Why then was lysine selected? A number of reasons can be cited.

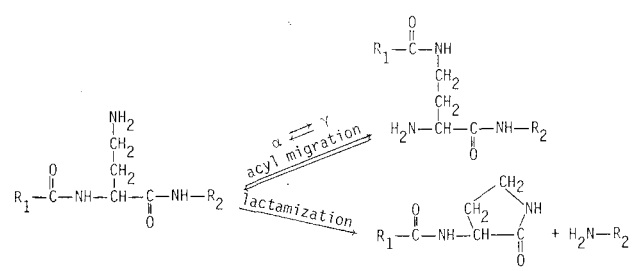
In the case of α,β-diaminopropionic acid, the most serious problem is α ⇌ β acyl migration in polypeptides as shown in Fig. 2. Reversible α ⇌ β acyl migration has been shown for the synthetic peptide, N^{α}-glycyl and N^{β}-glycyl-α,β-diaminopropionic acid and for the natural antibiotics Edeine A and B (Poduska et al. 1965; Hettinger and Craig 1970). The rapid racemication of diaminopropionic acid (Jacobson et al. 1974) and its relatively rapid deamination by β-elimination (Van Trump et al. 1981) are additional factors that may have excluded it from proteins.

As shown in Fig. 2, α,γ-diaminobutyric acid undergoes deleterious α ⇌ γ acyl migration and lactamization in polypeptides. Activated esters of α,γ-diaminobutyric acid would also undergo lactamization. Reversible α ⇌ γ acyl migration has been shown for the synthetic peptides N^{α}-glycyl and N^{γ}-glycyl-α,γ-diaminobutyric acid (Poduska et al. 1965). Lactamization of α,γ-diaminobutyric acid in peptides has been shown for α,γ-diaminobutyric acid amide, α,γ-diaminobutyrylglycine and α,γ-diaminobutyrylleucine (Poduska et al. 1965). Esters of α,γ-diaminobutyric acid lactamize in the same way as γ-aminobutyric acid (Hay and Morris 1972). A greater problem with diaminobutyric acid is that a peptidyl tRNA with a diaminobutyric acid at the C-terminus can lactamize and terminate peptide synthesis.

Ornithine also undergoes lactamization in peptides and activated esters. Lactamization of ornithine in peptides has been shown in the case of ornithine amide and ornithinylglycine (Lipson and Sondheimer 1964; Poduska et al. 1965). The lactamization of ornithine methyl



α,β-diaminopropionic acid



α,γ-diaminobutyric acid

Fig. 2. Acyl migration and lactamization in peptides of α,β-diaminopropionic acid and α,γ-diaminobutyric acid

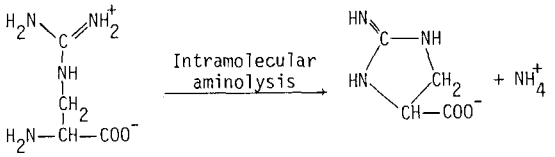
ester is very rapid and would not only lactamize on the tRNA, but would terminate peptide synthesis when positioned at the C-terminus of peptidyl tRNA.

Diaminobutyric acid and ornithine would make excellent chain terminating codons. At first sight ornithine would seem superior to diaminobutyric acid, but the rapid lactamization at the aminoacyl-tRNA stage might preclude its incorporation into peptidyl-tRNA.

Jukes (1974) has proposed that the codons now coding for arginine originally coded for ornithine. We believe that the above factors make this unlikely.

The above discussion shows that lysine is the shortest diamino acid that is suitable for proteins. The only problem with lysine is that an efficient prebiotic synthesis is still to be worked out. The next higher homolog, 2,7-diaminoheptanoic acid would also be suitable, but the availability in the primitive ocean would have been less.

Arginine. A good case can also be made for the incorporation of arginine into proteins if it is assumed that a permanent positively charged side chain is needed. The pK of the guanidino group is about 13 and so very little is uncharged around pH 7.



α-Amino-β-guanidinopropionic acid

Fig. 3. Cyclization of a guanidino amino acid

There is no clearly established prebiotic synthesis of arginine or its homologs. It would be expected, however, that the diamino acids would be guanidinated easily in the primitive ocean if cyanamide were present. The lower homologs, α-amino-β-guanidino propionic acid and α-amino-γ-guanidino butyric acid, would therefore be expected to have been reasonably abundant in the primitive ocean. In the same way, arginine and homoarginine (α-amino-ϵ-guanidino caproic acid) would have been present to the extent that ornithine and lysine were present. The two lower homologs are not present in proteins and to our knowledge have not been found in nature (Mooz 1976). Yet they were probably more abundant in the primitive ocean.

The absence of the two lower homologs can be accounted for by their tendency to cyclize as shown in Fig. 3 for α-amino-β-guanidino propionic acid. The cyclization of α-amino-β-guanidino propionic acid would yield a substituted imidazoline, and cyclization of α-amino-γ-guanidino butyric acid would yield a substituted tetrahydropyrimidine. The related cyclization of aminoethylguanidine has been shown to occur under weakly basic conditions (Abramson et al. 1969; Adcock et al 1961). Also, the inductive effect of the positively charged guanidino group of the lower arginine homologs may destabilize peptide bonds of these amino acids.

Homoarginine would appear to be as suitable as arginine for use as a positively charged amino acid. The absence of homoarginine in proteins can be attributed to reasons such as its abundance was lower than arginine, and that one guanidino amino acid is apparently adequate.

The above discussion shows that arginine is the shortest guanidino amino acid that is suitable for proteins. The only problem is the absence of a well worked out prebiotic synthesis.

Histidine. It is generally thought that histidine is present in proteins because of the special properties of the imidazole group. There is no established prebiotic synthesis of histidine, although one will probably be worked out before too long. If the special properties of the imidazole group are accepted as the reason for the presence of histidine, then we need only discuss whether other homologs of histidine might be used in proteins.

The structure of histidine suggests that its tRNA ester could possibly form a lactam, via a five-membered cyclic intermediate. However, lactam formation with histidine is energetically uphill, and apparently does not form during the hydrolysis of histidyl tRNA (Gatica et al. 1966) and methyl histidiante (Hay et al. 1966).

It is likely that the tRNA ester of the higher homolog of histidine, γ-(4'-imidazolyl)-α-amino-butyric acid, is unstable, since the rate of hydrolysis of γ-(4'-imidazolyl) -butyrate has been found to be 20 times faster than ethyl butyrate (Bruice and Sturtevant 1959). The lower homolog of histidine, α-(4'-imidazolyl)-α-aminoacetic acid is probably unstable to deamination (Bruice and Herz 1964).

Histidine, then appears to be the most suitable imidazole containing amino acid. The case would be improved if a prebiotic synthesis were available.

Hydroxy Amino Acids

Serine and threonine are the simplist hydroxy amino acids. Serine has been found among the electric discharge products, but not in the Murchison meteorite. Serine decomposes by reversible dealdolization to glycine and formaldehyde (Friedmann et al. 1971), dehydration by β-elimination to pyruvate and NH_3 (Schroeder and Bada 1977; Bada et al. 1978) and decarboxylation to ethanolamine (Vallentyne 1964). The rates in solution are dehydration > dealdolization > decarboxylation, but the order is reversed in carbonate shells. In solution the half-life for dehydration is 2 ×

10^5 years at 4°C, so serine is a relatively unstable amino acid. However, the synthesis of serine would be rapid by the reverse of the dealdolization reaction if the formaldehyde concentration in the primitive ocean was relatevely high,

$$\text{Glycine} + H_2CO \rightleftharpoons \text{Serine}$$

The equilibrium constant is not known, but it is likely to be between 10^2 and 10^4. Thus serine may have been much more abundant in the primitive ocean than indicated by the electric discharge yields and its relatively rapid decomposition.

The yields of threonine and allothreonine are substantially less in the electric discharge than serine, and none has been found in the Murchison meteorite. As with serine, threonine and allothreonine decompose by β-elimination, dealdolization and decarboxylation. The synthesis of threonine from the reaction

$$\text{Glycine} + \text{Acetaldehyde} \rightleftharpoons \text{Threonine} + \text{Allothreonine}$$

would be expected, although it has not been investigated except using Cu^{+2} and pyridoxal as a catalyst (Sato et al. 1957; Metzler et al. 1954). This reaction would probably be less favorable than the serine synthesis, so threonine would have been considerably less abundant than serine in the primitive ocean.

From a structural standpoint there would appear to be no advantage of using threonine over allothreonine. However, the rate of β-elimination with allothreonine is at least twice that with threonine (Schroeder and Bada 1977). This instability therefore may account for the use of threonine rather than allothreonine in proteins.

It is to be noted that threonine is biosynthesized in two easy steps from homoserine, and this may have been one of the first biosynthetic amino acid transformations.

Homoserine may have been relatively abundant on the primitive earth since it would have been produced from acrolein along with glutamic acid, methionine and diaminobutyric acid. The relative amounts would depend on the concentrations of nucleophiles in the primitive ocean (Schlesinger 1978). No homoserine was found in the electric discharge, but this is probably due to the high concentrations of reactants in these experiments.

The major disadvantage of homoserine is the rapid lactonization of its activated esters (aminoacyl or peptidyl tRNA) that proceeds via a five-membered cyclic intermediate (Fig. 4). Fickel and Gilvarg (1973) suggested that the lactonization of homoseryl tRNA would interfere with the use of homoserine in protein synthesis. The next homolog of homoserine, α-amino-δ-hydroxy valeric acid would also be expected to undergo lactonization of its activated tRNA ester. α-Amino-ε-hydroxy caproic acid would not lactonize. Neither of these homologs have been found in prebiotic experiments.

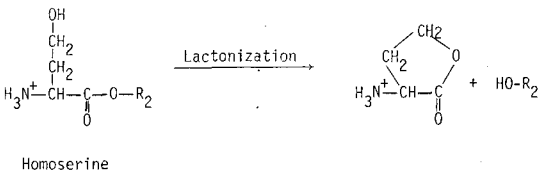


Fig. 4. Lactonization of an activated ester of homoserine

Cysteine. Of the thiol amino acids, cysteine is the most likely to be incorporated into proteins. It has not been found in electric discharge syntheses or in the Murchison meteorite. It is reported to be synthesized from the photochemical reaction of H_2S, CH_4, NH_3 and H_2O (Sagan and Khare 1971; Khare and Sagan 1971). Substantial amounts of cysteine were probably synthesized in the primitive ocean from dehydroalanine, which is formed by β-elimination from serine and diaminopropionic acid, since H_2S is a good nucleophile relative to H_2O and NH_3. The problem with cysteine is its decomposition, both by β-elimination and oxidation. In the absence of data on these processes, it is difficult to estimate how much was present in the primitve ocean.

It is possible that homocysteine was more abundant than cysteine in the primitive ocean, since it is produced from acrolein and is probably more stable than cysteine. However, the activated tRNA esters of homocysteine would be particularly susceptible to thiolactone formation.

Methionine. This amino acid is synthesized in fair yield by electric discharges via acrolein (Van Trump and Miller 1972) and appreciable concentrations may have built up in the primitive ocean (Schlesinger 1978). Methionine is generally thought to be unstable, but this instability is due to oxidation by O_2.

Anaerobic heating experiments show that methionine is relatively stable (Van Trump and Miller 1972). In the primitive ocean, oxidation by Fe^{+3} or other oxidizing agents may have occurred, and cyanogen would have destroyed the methionine. However, the balance between synthesis and destruction is not clear. It is possible that homocysteine was more abundant than methionine in the primitive ocean and that methionine was derived from homocysteine in one biosynthetic reaction. The argument that methionine is not a primitive amino acid because of its single codon is not a strong one, and there are two codons for methionine in mitochondria (Barrell et al. 1979).

The lower homolog of methionine, S-methyl cysteine, has not been identified in any prebiotic synthesis, but would be expected from the Michael addition of CH_3SH to dehydroalanine. S-methyl cysteine is probably less stable than methionine, but data are needed on this point. The higher homologs of methionine have not been found in any prebiotic experiments.

Since the function of methionine in proteins is not understood, it is difficult to judge its suitability relative to S-methyl cysteine or higher homologs. If a CH_3-group is needed, then methionine would probably be favored over S-methyl cysteine on the grounds of abundance and stability.

Aromatic Amino Acids

The number of possible aromatic amino acids is very large. Except for phenylalanine, tyrosine and tryptophan, there are no prebiotic syntheses of aromatic amino acids. Other aromatic amino acids may be present in prebiotic syntheses but not looked for during analysis.

There is a good prebiotic synthesis of phenylalanine from phenylacetylene, which in turn is produced in good yield by pyrolysis of hydrocarbons (Friedmann and Miller 1969). The concentration of phenylalanine in the primitive ocean could have been substantial, possibly as high as the leucines, depending on the exact conditions.

Tyrosine would have been synthesized by hydroxylation of phenylalanine using traces of OH radicals in the primitive ocean. Ortho and meta tyrosine would also have been produced in this reaction. Activated esters of ortho tyrosine would be especially susceptible to lactone formation. The problem with tyrosine, being a phenol, is that it is relatively reactive. There are no data available to estimate how fast and by which path tyrosine would decompose in the primitive ocean.

Tryptophan is sometimes thought not to have been a primitive amino acid because of its apparent complexity and single codon (mitochondria have two codons for tryptophan (Macino et al. 1979; Barrell et al. 1979)). However there is a reasonable prebiotic synthesis of tryptophan from indole and dehydroalanine, the indole being produced in good yield from the pyrolysis of hydrocarbons and NH_3 (Friedmann et al. 1971). The weak part of this synthesis is the slow addition of indole to the dehydroalanine. Nevertheless a small amount of tryptophan might have been present in the primitve ocean.

The biosyntheses of these aromatic amino acids start with erythrose-4-phosphate + phosphoenolpyruvate and branch off at chorismic acid, seven steps later, with three additional steps for phenylalanine and tyrosine and four additional for tryptophan. Several of the steps beyond chorismic acid can occur non-enzymatically (Andrews et al. 1973; Danishefsky et al. 1979). This situation suggests that only one of the aromatic amino acids was incorporated initially into proteins of early organisms and that the shikimic acid pathway was developed in reverse order, as suggested by Horowtiz (1945). The other two aromatic amino acids would have appeared as by-products of this process, and being useful in proteins, would have been incorporated permanently. Phenylalanine may have been the first aromatic amino acid used, since it was more abundant in the primitive ocean. However, arguments can be made that tyrosine or tryptophan were used first, with the other two arising as useful by-products of the biosynthetic pathway.

Asparagine and Glutamine. These two amino acids would have been synthesized prior to the formation of aspartic and glutamic acids by the hydrolysis of their nitriles. However it is doubtful whether significant amounts of these two amino acids could have occured in the primitive ocean, a point emphasized by Miller and Orgel (1974) and by Wong and Bronskill (1979). Glutamine is particularly unfavorable since it is converted rapidly to pyrrolidone-α-carboxylic acid (Hamilton 1945; Gilbert et al. 1949, Martin et al. 1964). Glutamine is much more stable in peptides than free in solution, but both glutamine and asparagine in peptides would have half lives for hydrolysis that are short on the geological time scale (Robinson et al. 1973). However, under polymerizing conditions in the presence of ammonia considerable steady state concentrations of asparagine and glutamine could have been formed in peptides. Nevertheless, these two amino acids are likely to be late additions.

It is possible that the early substitute for glutamine was albizzine (α-amino-β-ureidopropionic acid). This would be produced from diaminopropionic acid and cyanate. It is an analog of glutamine and inhibits a number of glutamine enzymes. Kinetic data indicate that albizzine is several hundred times slower than glutamine in its decomposition (Van Trump et al. 1981).

Functional Groups Missing from Coded Amino Acids

There are many functional groups in proteins that are synthesized by post-translational modifications (Uy and Wold 1977). These amino acids are presumably present because of their special functions, which are not useful for a typical enzyme, but which are required for a limited number of proteins.

Many possible functional groups are unstable to hydrolysis. These unstable groups include acetals, ketals, phosphates, acid anhydrides, thiocyanates, imides, amidines, thioamides, chlorides, bromides, iodides, and esters. In addition, some of these groups react rapidly with the amino or other groups in proteins.

Nitriles (e.g., β-cyanoalanine) are unstable to hydrolysis and although not a very rapid reaction, the hydrolysis would probably be fast enough for nitriles to be unsuitable. The hydrolysis of nitriles differs from the hydrolysis of amides (e.g. asparagine and glutamine) in that polymerizing conditions, which can resynthesize asparagine and glutamine, cannot do so with nitriles in substantial yield. Amides of the form ϵ-N-acetyl lysine are missing, presumably due to hydrolytic instability. Also missing are ureas (e.g. citrulline). These are

relatively stable to hydrolysis and would be formed under prebiotic conditions. It is not clear what role they would play in protein function (however, see section on glutamine and asparagine).

Sulfonic acids, e.g. cysteic acid, are missing. These might have been synthesized under primitive earth conditions, but they may not be particularly useful in proteins. Although sulfonic acids have a negative charge, it is too strong an acid to act as a general acid or the anion as a general base. Phosphonic acids which contain a C-P bond have suitable pK values for general acid-base reactions. They occur in some marine organisms (Kittredge and Roberts 1969) but not in their proteins. These amino acids might be suitable for proteins but there is no prebiotic synthesis of them. The absence of phosphate ester side chains, e.g. phosphoserine, is probably due to their being unstable in hydrolysis and β-elimination (Samuel and Silver 1963; Belec and Jenness 1962).

Quaternary ammonium ions are also missing (e.g. ϵ-N-trimethyl lysine), except for post-translational modifications. These compounds cannot act as general acids or bases or form imines as can lysine. There is a permanent positive charge on the side chain, and so they could substitute for arginine. It is not clear why arginine was selected over quaternary amines of this type, although the ability of the guanidino group to hydrogen bond may be responsible. Secondary and tertiary amines (e.g. ϵ-N-methyl lysine and ϵ-N-dimethyl lysine) are missing from proteins. These might substitute for lysine, but there does not seem to be a clear advantage in doing so. In addition the prebiotic synthesis would probably be less favorable than for lysine.

Aldehydes and ketones are missing, which is expected since these functional groups are too reactive. When these are synthesized after translation, they tend to form crosslinks. Ethers such as O-methyl serine or aminoethyl serine (which could replace lysine) do not occur. It is difficult to see what function an ether side chain would play.

Side Chains With Several Functional Groups. These do not occur except for post-translational modification, e.g. γ-carboxy glutamic acid. In general, one would expect that the spatial adjustments of mono-functional amino acids permits better fine tuning of properties than does the use of polyfunctional amino acids. Exceptions to this might be those cases where metal ion chelation is involved (e.g. calcium chelation by γ-carboxy glutamic acid). It is noted that many of the bifunctional amino acids (e.g. δ-hydroxylysine and γ-aminoornithine) are unfavorable because of lactone or lactam formation or other features that exclude compounds which are monofunctional.

7. Why Just 20 Amino Acids?

This question is related to the origin of the genetic code of which there have been many discussions. The usual concern has been to decide which amino acids are late additions rather than primitive with the one codon amino acids, methionine and tryptophan, being the favorite choices and the two codon amino acids being the second choice. For example Wong (1976) has proposed a coevolution theory of the origin of the genetic code in which the code started out with 7 amino acids, which grew to 20 by replacing adjacent codons with metabolically related amino acids. Eigen and Schuster (1978) propose that 8 amino acids were coded for originally, based on codon-anticodon stabilities. Crick (1967, 1968) has proposed that the present number of amino acids and their codon assignments are at least in part due to an historical accident and by the time the 20 were fixed any change in the genetic code would be strongly selected against or lethal. Lagerkvist (1978) has discussed the possibility that the present genetic code can handle no more than 24 things (23 amino acids and a terminator) and still maintain adequate fidelity of translation because ambiguity in reading the third base of the codon is greater than predicted by "wobble" theory (Crick 1966) for some codons.

Restrictions in the coding capacity of the present code do not necessarily apply to the genetic code at an early stage in its development. The early list of coded amino acids may have been larger than the present 20. The basis for this is that many different amino acids were present in the primitive ocean and it is difficult to see how some of them, present in large amounts, were not used, e.g. α-amino-n-butyric acid, norvaline, and norleucine. It can be argued that norvaline is difficult to distinguish from leucine, but the fidelity in the activating enzymes was probably low at this time, so having the extra amino acids would not have been a great disadvantage in this regard. However, as translational precision improved, the extra amino acids would have been deleted in order to improve the fidelity of aminoacylation of tRNA and to reduce translation errors that result from having two amino acids assigned to codon families that are read by the "two out of three" method (Lagerkvist 1978).

Conclusions

The above discussion shows that the twenty amino acids that occur in proteins are by and large the most suitable ones within each class, i.e. acidic, basic, etc. There is no single reason, such as abundance in primitive ocean, to account for the selection, and it was necessary to apply a variety of considerations. The arguments seem compelling for the acidic amino acids and for lysine and arginine. Histidine may be present because of the special

properties of the imidazole group. The hydroxy amino acids are the reasonable set, as well as the hydrophobic amino acids including proline, except for the absence of α-amino-n-butyric acid, norvaline and norleucine. The two sulfur amino acids are quite plausible from many standpoints. The aromatic amino acids are also plausible, but there are so many other possibilities that the basis for their selection becomes uncertain. Asparagine and glutamine do not seem primitive because of their instability.

We can express the plausibility of our arguments quantitatively by asking the question: Assuming an independent origin of life on another planet, with this life having protein catalysts and assuming there are 20 amino acids in these proteins, how many of the 20 would be the same as on earth? Based on the above discussion, we believe that 15 of the 20 would be the same. We believe that our arguments are compelling enough to say that only 10 out of the 20 being the same would be surprising. On the other hand, the uncertainties and puzzling absences would make it equally surprising to find all 20 amino acids the same as on the earth.

It might be said that this discussion is based on after the fact knowledge, and that such theories are not testable. The ideas, of course, can be tested when an independent origin and evolution of life is found on another planet. Lacking that possibility at present, we can still examine the problem by investigation prebiotic synthetic processes, the stability of amino acids, their peptides and esters, the structure of the ribosome and its precursor, and comparative studies of protein and RNA sequences.

Acknowledgments. This work was supported by NASA grants NSG 7627 (AW) and NAGW-20 (SM).

References

Abramson FB, Furst CI, McMartin C, Wade R (1969) Biochem J 113:143–156

Adcock B, Lawson A, Miles DH (1961) J Chem Soc:5120–5127

Andrews PR, Smith GD, Young IG (1973) Biochemistry 12: 3492–3498

Anfinsen CB, Corley LG (1969) J Biol Chem 244:5149–5152

Bada JL, Miller SL (1968) Science 159:423–425

Bada JL, Shou M, Man EH, Schroeder RA (1978) Earth Planet Sci Lett 41:67–76

Balasubramanian D (1974) Biopolymers 13:407–410

Balasubramanian D, Kalita CC, Kovacs J (1973) Biopolymers 12:1089–1098

Barrell BG, Bankier AT, Drouin J (1979) Nature 282:189–194

Belec J, Jenness R (1962) Biochim Biophys Acta 63:512–514

Bonner WA (1972) Origins of molecular chirality. In: Ponnamperuma C (ed) Exobiology. North-Holland, Amsterdam, p 170

Brack A, Spach G (1980) J Mol Evol 15:231–238

Bruice TC, Herz JL (1964) J Am Chem Soc 86:4109–4116

Bruice TC, Sturtevant JM (1959) J Am Chem Soc 81:2860–2870

Burgess AW, Leach SJ (1973) Biopolymers 12:2599–2605

Cowie DB, Cohen GN, Bolton ET, Robichon-Szulmajster HDe (1959) Biochim Biophys Acta 34:39–46

Crick FHC (1966) J Mol Biol 19:548–555

Crick FHC (1967) Nature 213:119

Crick FHC (1968) J Mol Biol 38:367–379

Crick FHC, Brenner S, Klug A, Pieczenik G (1976) Origins of Life 7:389–397

Cronin JR, Moore CB (1971) Science 172:1327–1329

Danishefsky S, Hirama M, Fritsch N, Clardy J (1979) J Am Chem Soc 101:7013–7018

Deslauriers R, Walter R, Smith CP (1973) FEBS Lett 37:27–32

Eigen M, Schuster P (1978) Naturwissenschaften 65:341–369

Fahnestock S, Rich A (1971) Science 173:340–343

Fickel TE, Gilvarg C (1973) J Org Chem 38:1421–1423

Friedmann N, Haverland WJ, Miller SL (1971) Prebiotic synthesis of aromatic and other amino acids. In: Buvet R, Ponnamperuma C (eds) Chemical evolution and the origin of life. North-Holland, Amsterdam, p 123

Friedmann N, Miller SL (1969) Science 166:766–767

Gatica M, Allende CC, Mora G, Allende JE, Medina J (1966) Biochim Biophys Acta 129:201–203

Gilbert JB, Price VE, Greenstein JP (1949) J Biol Chem 180: 209–218

Glickson JD, Applequist J (1971) J Am Chem Soc 93:3276–3281

Goodman M, Fried M (1967) J Am Chem Soc 89:1264–1267

Gund P, Veber DF (1979) J Am Chem Soc 101:1885–1887

Hamilton PB (1945) J Biol Chem 158:375–395

Hay RW, Morris PJ (1970) J Chem Soc (B):1577–1582

Hay RW, Morris PJ (1972) J Chem Soc Perkin II:1021–1029

Hay RW, Porter LJ (1967) J Chem Soc (B):1261–1264

Hay RW, Porter LJ, Morris PJ (1966) Aust J Chem 19:1197–1205

Hettinger TP, Craig LC (1970) Biochemistry 9:1224–1232

Horowitz NH (1945) Proc Natl Acad Sci USA 31:153–157

Isumiya N, Fu SJ, Birnbaum SM, Greenstein JP (1953) J Biol Chem 205:221–230

Jacobson SJ, Wilson CG, Rapoport H (1974) J Org Chem 39: 1074–1077

Jukes TH (1974) Origins of Life 5:331–350

Jungck JR (1978) J Mol Evol 11:211–224

Khare BN, Sagan C (1971) Nature 232:577–579

Kittredge JS, Roberts E (1969) Science 164:37–42

Krayevsky AA, Kukhanova MK (1979) The peptidyltransferase center of ribosomes. In: Cohn WE (ed) Progress in nucleic acid research and molecular biology. Vol. 23. Academic Press, New York, p 1

Kushwaha DRS, Mathur KB, Balasubramanian D (1980) Biopolymers 19:219–229

Kvenvolden K, Lawless J, Pering K, Peterson E, Flores J, Ponnamperuma C, Kaplan IR, Moore C (1970) Nature 228:923–926

Kvenvolden KA Lawless JG, Ponnamperuma C (1971) Proc Natl Acad Sci USA 68:486–490

Lagerkvist U (1978) Proc Natl Acad Sci USA 75:1759–1762

Lawless JG, Levi N (1979) J Mol Evol 13:281–286

Leplawy MT, Jones DS, Kenner GW, Sheppard RC (1960) Tetrahedron 11:39–51

Lipson MA, Sondheimer E (1964) J Org Chem 29:2392–2394

Macino G, Coruzzi G, Nobrega FG, Li M, Tzagoloff A (1979) Proc Natl Acad Sci USA 76:3784–3785

Mark JE, Goodman M (1967) J Am Chem Soc 89:1267–1268

Martin RB, Parcell A, Hedrick RI (1964) J Am Chem Soc 86: 2406–2413

Meister A, Bukenberger MW (1962) Nature 194:557–559

Metzler DE, Longenecker JB, Snell EE (1954) J Am Chem Soc 76:639–644
Miller SL (1957) Biochim Biophys Acta 23:480–489
Miller Sl, Orgel LE (1974) The origins of life on the earth. Prentice-Hall, Englewood Cliffs, New Jersey, p 121
Mooz ED (1976) Data on the naturally occuring amino acids. In: Fasman GD (ed) Handbook of biochemistry and molecular biology. proteins, Vol. 1. Chemical Rubber Co. Press, Cleveland, p 111
Nagaraj R, Shamala N, Balaram P (1979) J Am Chem Soc 101: 16–20
Nathans D, Neidle A (1963) Nature 197:1076–1077
Norden B (1978) J Mol Evol 11:313–332
Old JM, Jones DS (1975) Biochem Soc Trans 3:659–660
Peltzer ET (1979) Thesis, University of California, San Diego
Peltzer ET, Bada JL (1978) Nature 272:443–444
Poduska K, Katrukha GS, Silaev AB, Rudinger J (1965) Collect Czech Chem Commun 30:2410–2433
Pospisek J, Blaha K (1976) Syntheses of peptides containing a tert-leucine residue. In: Loffet A (ed) Peptides 1976. Editions Universitaires, Brussels, p 95
Reuben J, Polk FE (1980) J Mol Evol 15:103–112
Rich A (1971) The possible participation of esters as well as amides in prebiotic polymers. In: Buvet R, Ponnamperuma C (eds) Chemical evolution and the origin of life. North-Holland, Amsterdam, p 180
Ring D, Wolman Y, Friedmann N, Miller SL (1972) Proc Natl Acad Sci USA 69:765–768
Robinson AB, Scotchler JW, McKerrow JH (1973) J Am Chem Soc 95:8156–8159
Rychlik I, Cerna J, Chladek S, Pulkrabek P, Zemlicka J (1970) Eur J Biochem 16:136–142
Sagan C, Khare BN (1971) Science 173:417–420
Samuel D, Silver BL (1963) J Chem Soc 289–296
Sato M, Okawa K, Akabori S (1957) Bull Chem Soc Japan 30: 937–938
Schlesinger G (1968) Dissertation, University of California, San Diego
Schroeder RA, Bada JL (1977) Geochim Cosmochim Acta 41: 1087–1095
Thanassi JW (1970) Biochemistry 9:525–532
Uy R, Wold F (1977) Science 198:890–896
Van Trump JE, Miller SL (1972) Science 178:859–860
Van Trump JE, White R, Miller SL (1981) in press
Vallentyne JR (1964) Geochim Cosmochim Acta 28:157–188
Weber AL, Lacey JC Jr (1978) J Mol Evol 11:199–210
Wilson H, Cannan RK (1937) J Biol Chem 119:309–331
Woese CR (1967) The genetic code: The molecular basis for genetic expression. Harper and Row, New York
Wolman Y, Haverland WJ, Miller SL (1972) Proc Natl Acad Sci USA 69:809–811
Wong JT (1976) Proc Natl Acad Sci USA 73:2336–2340
Wong JT, Bronskill PM (1979) J Mol Evol 13:115–125
Zeitman B, Chang S, Lawless JG (1974) Nature 251:42–43

Received October 29, 1980 / Revised February 23, 1981