J Mol Evol (1995) 40:541–544

JOURNAL OF MOLECULAR EVOLUTION

Speculations on the Origin of the Genetic Code

Hyman Hartman

Institute for Advanced Studies in Biology, 880 Spruce St., Berkeley, CA 94707, USA

Received: 18 July 1994

Abstract. The most primitive code is assumed to be a GC code: GG coding for glycine, CC coding for proline, GC coding for alanine, CG coding for "arginine." The genetic code is assumed to have originated with the coupling of glycine to its anticodon CC mediated by a copper-montmorillonite. The polymerization of polyproline followed when it was coupled to its anticodon GG. In this case the aminoacyl-tRNA synthetase was a copper-montmorillonite. The first membrane is considered to be a β sheet formed from polyglycine. As the code grew more complicated, the alternative hydrophobic-hydrophilic polypeptide (alanine-"arginine") was coded for by the alternating CG copolymer. This alternating polypeptide (ala-"arg") began to function as both a primitive membrane and as an aminoacyl-tRNA synthetase. The evolution of protein structure is tightly coupled to the evolution of the membrane. The α helix was evolved as lipids became part of the structure of biological membranes. The membrane finally became the fluid mosaic structure that is now universal.

Key words: Genetic code — Aminoacyl-tRNA synthetase — Protein structure — Membrane — Clay

Introduction

The synthesis of proteins by present-day cells is a very complex process involving ribosomes, messenger RNA, tRNAs, and aminoacyl-tRNA synthetases. How could such a complex system have originated? Perhaps the best example of this, which typifies the problem of the origin of protein synthesis, is the interaction between the amino acid and the tRNA which must be mediated by an aminoacyl-tRNA synthetase. This necessity is the greatest obstacle that must be overcome in theorizing about the origin of the genetic code. The need for a complex protein to start the code obviously leads to great difficulties. This paper will attempt to overcome these difficulties.

Aminoacyl-tRNA Synthetases

Recent advances in the determination of the structures and of the amino acid sequences of the aminoacyl-tRNA synthetases has stimulated new interest in the evolution of the genetic code. A classification of the synthetases into two classes has been made (Eriani et al. 1990). The two classes are:

1. Class I, which includes valine (val), leucine (leu), isoleucine (ile), methionine (met), tyrosine (tyr), tryptophan (trp), glutamic acid (glu), glutamine (gln), arginine (arg), and cysteine (cys)
2. Class II, which includes proline (pro), threonine (thr), serine (ser), phenylalanine (phe), aspartic acid (asp), asparagine (asn), histidine (his), and lysine (lys); and class IIa, which includes glycine (gly) and alanine (ala)

Class II synthetases are considered to be the more primitive of the synthetases, and class IIa seems the most primitive of all.

Based on a presentation made at a workshop—"Aminoacyl-tRNA Synthetases and the Evolution of the Genetic Code"—held at Berkeley, CA, July 17–20, 1994

In the evolution of the genetic code, it would appear that one reason for the separation of the aminoacyl-tRNA into two classes was to make distinctions possible between similar amino acids which share the first two bases of a codon (e.g., glu and asp, codon {GA-}). Seven out of ten of the aminoacyl-tRNA synthetases in class I are related to hydrophobic amino acids (val, leu, ile, met, tyr, trp, cys), which are thus presumed to be later entries into the genetic code since class II is considered to have preceded class I synthetases. The question which then arises is, what is the evolutionary relationship between the other three synthetases (glu, gln, arg)? One might assume that the assignment of glutamyl-tRNA synthetase to class I was made as the aspartyl-tRNA synthetase was already in class II. The real problem is with arginine. How is it related to the rest of the class I synthetases? In order for the distinctions between glu and asp to have been made the distinction between class I and class II synthetases had to have already been made. The possibility thus exists that the first class I synthetase was assigned to arg or its precursor.

The recent structural studies on aminoacyl-tRNA synthetases have shown a conserved module which contains the binding site for ATP, the amino acid and the CCA of the amino-accepting arm of the t-RNA, and a nonconserved module which contains the anticodon binding site (Schimmel et al. 1993). It is a general view that proteins have evolved from simpler modules (Gilbert 1987). The search for even simpler modules in the aminoacyl t-RNA synthetases has just begun.

The Membrane

What was the origin of the aminoacyl-tRNA synthetases? One possibility is that the membrane, a lipid bilayer, acted as the earliest aminoacyl-tRNA synthetase. The lipid bilayer could distinguish between amino acids and nucleosides on the basis of their hydrophobicity and hydrophilicity: "Since proteins have in general their hydrophobic groups on the inside and hydrophilic groups on the outside, they could have evolved to take over the recognition function played by the lipid bilayer" (Hartman 1975a). This suggestion placed the burden for originating the code on the cell membrane. The time has come to look more closely at the origin and evolution of the cell membrane.

Cellular membranes are mainly composed of two types of molecules: (1) phospholipids and (2) proteins. An early and very influential model of cellular membranes was developed by Danielli and Davson (1935). In this model the phospholipids formed a lipid bilayer and the proteins were thought to be globular and associated with the charged surface of the phospholipid bilayer. Danielli later modified the association of proteins to the lipid bilayer by considering the protein to be in a extended β-form and some of the hydrophobic side chains of the amino acids of the protein to be buried in the lipid portion of the membrane. Physical chemical studies, however, showed that the dominant conformation of membrane proteins was largely α helical (Wallach and Zahler 1966). This necessitated a modification of the Davson-Danielli model for the cellular membrane. The new model was called the fluid-mosaic model. The lipid bilayer of phospholipids was retained but now there were two types of proteins associated with the membrane: (1) intrinsic proteins, which are proteins which are embedded in the lipid bilayer, and (2) extrinsic proteins, which are proteins which interact only with the charged surface of the phospholipid bilayer. The intrinsic proteins must interact with the phospholipids as they traverse the lipid bilayer (Singer and Nicolson 1972). The most common conformation of these proteins involves transmembrane α helices. The problem, then, with the origin of the membrane is: which came first—the phospholipids or the proteins?

A current model for the origin of membranes proposes that in the oceans of the early Earth an accumulation of fatty acids occurred, due, perhaps, to a Fischer-Tropsch reaction from carbon dioxide, carbon monoxide, and hydrogen. The possibility of other chemical syntheses of the fatty acids should not be ruled out as our knowledge of the early Earth is very limited.

These fatty acids would then, if the length of the alkyl side chain was long enough, spontaneously form lipid bilayers, and thus the primitive membrane was born. The formation of phospholipids was perhaps a later development.

The spontaneous formation of the lipid bilayers is a powerful motivation for this model of the primitive membrane. The major criticism of this proposal is that such a membrane is relatively impermeable to ions (e.g., potassium) and the monomers (e.g., amino acids). In fact, if such a membrane enclosed peptides or polynucleotides it would burst. The reason for such instability is the permeability of such lipid bilayers to water. The resulting osmotic pressure would cause the vesicles to burst.

These problems have led to a second model based on polypeptides for the formation of the primitive membrane. "Perhaps the earliest coded polypeptides acted only as a porous membrane or a glue to hold together the polynucleotides that had guided their synthesis" (Orgel 1987). The primitive membrane, according to Orgel, was made up of peptides which were alternating hydrophobic–hydrophilic sequences of amino acids coded for by some simple polynucleotides.

Experimentalists have taken up the study of alternating hydrophobic–hydrophilic oligomers of amino acids. The poly leucyl-lysine in salt has been shown to form "asymmetrical bilayers with a hydrophobic interior and a hydrophobic exterior because of hydrophobic side-chain clustering" (Brack and Orgel 1975).

The evolution of coded protein structure could have traversed the following route: first the alternating hydro-

phobic–hydrophilic polypeptides would form β sheets which would then form bilayers. The formation of β turns would then follow as the peptides grew in complexity. The appearance of helices would seem to coincide with the entry of lipids into the primitive membrane. The evolution of the membrane is clearly tied to the evolution of the genetic code and the complexity of the proteins synthesized.

The Genetic Code

Perhaps we can search for the origin and evolution of the code in the evolution of the tRNAs. It has been proposed that a precursor to transfer RNA was a small oligonucleotide subunit which had a loop and a double helical strand five bases long (Hartman 1984). This would suggest that the cloverleaf structure of t-RNA is a joining together of smaller precursors. If this was the case then the interactions between the aminoacyl-tRNA synthetases and tRNA should reflect this evolution of tRNA from its precursors. The recent work on microhelices and the minimalist tetraloop that are aminoacylated by glycine or alanine point to a smaller precursor of tRNA. It has been pointed out that "tRNA molecules may have developed from further differentiation and elaboration of RNA minihelices, which originally interacted with the minimalist active-site domain of the primordial synthetase. According to this viewpoint, the two domains of contemporary tRNA molecules had their origins in hairpin stem-loop oligonucleotides, which were eventually joined together to give the L-shaped tRNA structure" (Schimmel et al. 1993).

What was the evolution of the code? It is usually assumed that the code began simply and evolved into its present complexity. The vocabulary of the primitive code must have been limited to a few amino acids and nucleotides, and the vocabulary of nucleotides and amino acids expanded as the code evolved.

A vocabulary expansion model, which was proposed in 1975, postulated that the earliest code was a (GC) code. The amino acids coded for were glycine (GG), proline (CC), alanine (GC), and "arginine" (CG). The code then evolved by adding A to form a (GCA) code. The amino acids which were added were glutamic acid, aspartic acid, glutamine, asparagine, lysine, histidine, threonine, and serine. Finally U was added to the code to form a (GCAU) code. The amino acids which were added were valine, leucine, isoleucine, methionine, phenylalanine, tyrosine, tryptophan, and cysteine (Hartman 1975a).

The evolution of the genetic code then was from a GC code which coded for structural polypeptides and only later evolved into the complex enzymes as adenosine and finally uridine entered the code.

In a previous paper on the evolution of tRNA, the following conclusions were reached:

> The primitive genetic code was a doublet, G, C code. The code consisted of GG coding for glycine, CC coding for proline, CG coding for arginine and GC coding for alanine. . . . The remnants of this primitive code exist in the loops of the present day tRNA. If one considers the cloverleaf structure of tRNA then aside from the anticodon loop, the major invariances of the loops or open loops are 5′UGGU3″, 5′TψCGA3′, 5′*CCA3′. The GG, CG, CC are considered to be remnants of the ancient doublet G, C code. (Hartman 1984).

More recently in a discussion of β turns in early evolution, a similar conclusion reached was that the earliest code was a GC code to be followed by a GCA code. Finally, "the 'youngest' amino acids are those included after the appendage of uracil to the genetic code. These 'young' amino acids are mostly nonpolar and crucial for the formation of tertiary structures of globular proteins" (Jurka and Smith 1987). What is of interest is that this model of the evolution of the code was based on the evolution of β turns.

It has been proposed that arginine was preceded by a simpler basic amino acid, e.g., ornithine, in the early genetic code (Jukes 1973; Hartman 1975b). In a recent paper an even simpler candidate, β amino alanine, was proposed as the primitive basic amino acid precursor to arginine (Hartman 1994). Another possible precursor to arginine would be guanidioacetic acid.

If the primitive peptide was an alternating hydrophobic–hydrophilic polypeptide, then the simplest alternating hydrophobic–hydrophilic polypeptide coded for by a GC code would be a polymer of alanine and "arginine." Alanine would be coded for by GC- and "arginine" by CG-. This could have been the first membrane and it might have performed a number of functions. It might have performed the functions of an aminoacyl-tRNA synthetase as alternating hydrophobic–hydrophilic polypeptides have recently been shown to have catalytic properties (i.e., hydrolysis of oligoribonucleotides) (Brack 1993). It has been suggested that the polyalanine-arginine was the first polymerase. "Primitive nucleic acid replication is postulated to involve a particular type of peptide that establishes a 'growing point'—a polypeptide in which *alternate* amino acid residues are basic" (Woese 1973). The polypeptide suggested was a copolymer of alanine and arginine which would have been coded for by an alternating copolymer of C and G. It was also suggested that the copolymer of alanine and arginine was polymerized on some mineral surface (Woese 1973).

The problem here is, what was the primary function of the first membrane? Was it to form an enclosed space with an inside and outside or was it to coat a surface? If it was to coat a surface—for example, a clay particle—then this membrane could have been preceded by a yet-simpler precursor. Since, as discussed earlier, the early code included GG, which coded for glycine, and CC,

which would be the anticodon for glycine, a possible candidate would be glycine oligomers. This coupling between the anticodon CC and glycine would have to be mediated by what can be called the ultimate aminoacyl-tRNA synthetase. Where would such interactions between the polymerization of amino acids and polynucleotides take place that would lead to the origin of the code? The answer is, on the surface and edges of clays.

Experimentally the polymerization of glycine has been demonstrated on the surface of homoionic clays (e.g., montmorillonite). In particular, copper-montmorillonite has been shown to polymerize glycine into oligomers up to pentaglycine in length. This was done by a cycling procedure of drying, warming, and wetting. It was concluded from these experiments that, "the metal ions in the clay matrix could have played a key role in the processes believed to have been important for the origin of life on earth" (Lawless and Levi 1979). Experimental results have also shown that polypeptides are formed from aminoacyladenylates catalyzed by montmorillonites (Paect-Horowitz and Eirich 1988).

Experimentally, the polymerization of activated nucleotides to form polynucleotides catalyzed by montmorillonites is now a very active research topic (Ferris 1993).

The surface and edges of the clays are certainly good candidates for the catalysis of the polymerization of both nucleotides and amino acids. The next experimental frontier is to demonstrate their coupling.

Thus at some time the earliest code would have to have oligomers of guanine (poly G) coding for polyglycine. The code would thus have been trivial as poly G would code for oligomers of glycine. What is not trivial is that glycine would have to be recognized by a primitive tRNA (i.e., CC). How did this come about? One possibility is suggested by an interesting cocrystallizaton of a cytosine-glycyl-glycyl-copper(II) complex (Saito et al. 1974) and a cytidine-glycyl-glycyl-copper(II) complex (Szalda et al. 1975). Here a possible early "coding" is mediated by a copper ion and it is a diglycine which is "coding" for a cytosine. It is a diglycine which is specifically (with the help of a copper ion) interacting with the monomer cytosine. Thus it may have been diglycine which in the presence of copper ions coded for the polymerization of oligo C. The earliest code may have started from glycine oligomers and copper ions interacting with an activated cytosine which would line up on the edge of a copper-montmorillonite and be polymerized there. This poly C could then be a template for the polymerization of poly G. An interaction between GG and proline perhaps mediated by a copper or zinc ion and then polymerized on a poly C would also become possible on the edges and surface of a clay.

These sets of polymerizations on the edges and surface of a clay particle set the stage for the further evolution of both the code and the aminoacyl-tRNA synthetases.

In summary, the origin of the genetic code is tied up with the origin and evolution of the membrane.

References

Brack A (1993) Liquid water and the origin of life. Orig Life Evol Biosph 23:3–10

Brack A, Orgel L (1975) β structures of alternating polypeptides and their possible prebiotic significance. Nature 256:383–387

Danielli JF, Davson H (1935) A contribution to the theory of permeability of thin films. J Cell Physiol 5:495–508

Eriani G, Delarue M, Poch O, Gangloff J, Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347:203–206

Ferris, JP (1993) Catalysis and prebiotic synthesis. Orig Life Evol Biosph 23:307–315

Gilbert W (1987) The exon theory of genes. Cold Spring Harb Symp Quant Biol 52:901–905

Hartman H (1975a) Speculations on the evolution of the genetic code. Orig Life 6:423–427

Hartman, H. (1975b) Speculations on the origin and evolution of metabolism. J Mol Evol 4:359–368

Hartman H (1984) Speculations on the evolution of the genetic code III: the evolution of tRNA. Orig Life Evol Biosph 14:643–648

Hartman H (1995) Speculations on the evolution of the genetic code IV: the evolution of the aminoacyl-tRNA synthetases. Orig Life Evol Biosph 25:265–269

Jukes TH (1973) Arginine as an evolutionary intruder into protein synthesis. Biochem Biophys Res Commun 53:709–714

Jurka J, Smith TF (1987) β turns in early evolution: chirality, genetic code, and biosynthetic pathways. Cold Spring Harb Symp Quant Biol 52:407–410

Lawless JG, Levi N (1979) The role of metal ions in chemical evolution: polymerization of alanine and glycine in a cation-exchanged environment. J Mol Evol 13:281–286

Orgel L (1987) Evolution of the genetic apparatus: a review. Cold Spring Harb Symp Quant Biol 52:9–16

Paecht-Horowitz M, Eirich FR (1988) The polymerization of amino acid adenylates on sodium-montmorillonite with preadsorbed polypeptides. Orig Life Evol Biosph 18:359–387

Saito R, Terashima R, Sakai T, Tomita K (1974) The crystal structure of cytosine-glycyl-glycine-copper(II) complex, a biologically important ternary coordination complex. Biochem Biophys Res Commun 61:83–86

Schimmel P, Giege R, Moras D, Yokoyama S (1993) An operational RNA code for amino acids and possible relationship to genetic code. Proc Natl Acad Sci USA 90:8763–8768

Singer SJ, Nicolson GL (1972) The fluid mosaic model of the structure of cell membranes. Science 175:720–731

Szalda DJ, Marzilli LG, Kistenmacher TJ (1975) Dipeptide-metal-nucleoside complexes as models for enzyme-metal-nucleic acid ternary species. Synthesis and molecular structure of the cytidine complex of glycylglycinatocopper(II). Biochem Biophys Res Commun 63:601–605

Wallach DFH, Zahler PH (1966) Protein conformations in cellular membranes. Proc Natl Acad Sci USA 56:1552–1559

Woese C (1973) Evolution of nucleic acid replication: the possible role of simple repeating sequence polypeptides therein. J Mol Evol 2: 205–208