Polarity at onset of genetic coding. II, Primary recognition of amino acids by base doublets of prenucleic sugarless polymers secondarily taken over by ribonucleic acids

TURIAN, Gilbert

Abstract

According to our model of pregenetic molecular coding, nucleobase doublets anticodons were, by stereochemical affinities, the primary codons for amino acid letters of the primitive peptide sequences. Such doublets, anticipating the base sequences of the «modern» triplet codons of the reading strand of DNA would have been initially stabilized along polyphosphate chains by phosphoamidic dehydrating condensation with tripolyphosphate into riboseless prenucleic acids.

Reference


DOI : 10.5169/SEALS-740272
POLARITY AT ONSET OF GENETIC CODING.
II. PRIMARY RECOGNITION OF AMINO ACIDS BY BASE DOUBLETS OF PRENUCLEIC SUGARLESS POLYMERS SECONDARILY TAKEN OVER BY RIBONUCLEIC ACIDS

BY

Gilbert TURIAN*

(Ms reçue le 18.6.1997, accepté le 8.9.1997)

ABSTRACT

Polarity at onset of genetic coding. II. Primary recognition of amino acids by base doublets of prenucleic sugarless polymers secondarily taken over by ribonucleic acids. - According to our model of pregenetic molecular coding, nucleobase doublets anticodons were, by stereochemical affinities, the primary codons for amino acid letters of the primitive peptide sequences. Such doublets, anticipating the base sequences of the «modern» triplet codons of the reading strand of DNA would have been initially stabilized along polyphosphate chains by phosphoamidic dehydrating condensation with tripolyphosphate into riboseless prenucleic acids.

Key-words: Polarity, genetic, coding, prenucleic polymers, ribonucleic acids.

INTRODUCTION

We have recently considered the onset of genetic coding as possibly resulting of a molecular 3-main step evolutionary scenario, from the 1-letter coding amino acid units of self-reproducing peptides (TURIAN, 1996a; LEE & al., 1996) to the 3-letter nitrogenous base sequence of nucleic acids, through an intermediate, transitory step of a 2-letter base sequence of prenucleic polymers (TURIAN, 1996b). However, this model leaves open the below discussed fundamental questions of (1) direct recognition and thereby primordial coding of amino acids by doublets of nucleobases corresponding to the two first letters of «modern» anticodons stabilized on riboseless prenucleic polybasephosphate chains, (2) D-ribose insertion into these polybasephosphate chains and, with the advent of transfer (t) RNAs, (3) transition to the «modern» 3-letter base code on the newly produced pentosephosphate backbone, paralleled by (4) the increasing distance of the selected amino acids and their anticodons separated by adaptors complemented by newly formed aminoacyl tRNA synthetases.

1. Proximal prenucleic coding of peptidic amino acids.

The major problem at onset of coding implies the central question “did codons antedate or post-date anticodons?” If we consider that the original interaction between primordial amino acids and the first bases prebiotically available was a primary coding

* Pavillon des Isotopes, Microbiologie, Bd Ivoy 20, Université de Genève, 1211 Geneva 4, Switzerland.
by doublet then stabilized on polyphosphate, thereby behaving as primordial codon, the answer is ‘anticodons first’. We grounded our opinion on the fact that, in the ‘modern’ code, the first two nucleobases of their triplets rather than those of mRNA codons interact proximally in the closest affinity with specifically chosen amino acids and, therefore, would have been originally encoded by specific, mutual recognition with them. Such doublets of primordial codons would have been only primarily assigned to a given amino acid (CC for glycine, CG for alanine, etc.) and secondarily read in the anterotranslation processes as ‘modern’ anticodons after their stabilization, by phosphoramide bondings (N1/9-P) to polyphosphate chains (Fig. 2 in Turian, 1996b). Consequently, and admitting antedating of anticodons, also led us to pinpoint the common presence of C (cytosine) as first base coding the amino acid series glycine-alanine-valine (Fig. 3 in Turian, 1996b). As for the primary order in the tribase codons, it would have been found in a «grouping» pattern, namely with hydrophobic amino acids with a U (uracil) in median position while hydrophilic ones have a A (adenine) in this position. The earlier coding might thus have simply specified amino acids by their hydrophilic or hydrophobic characters or their acidity as already suggested by Crick in 1968. It is generally assumed that, during precellular evolution, the genetic code developed coordinately with a translation mechanism and that the early evolving or pregenetic code arose gradually and not all at once (see Darnell & al., 1990). It should therefore present adaptative features, the main one being that similar amino acids tend to be coded by similar anticodon doublets, for example CU for both aspartic and glutamic acids. In this respect, the coding process appears far from having been random and chemical specificity would have played a significant role in its onset. When we consider this fact that initial assignments of bases to amino acids resulted from chemical affinities and specific physical-chemical interactions between them, logically leads to admit that the closest fitting bases should have been the closest known to the selected amino acids, namely the two first bases of anticodons. This opinion has however been controverted since the deciphering of the «modern» genetic code (see § 3) and mechanistic models for the evolutionary basis of coding have primarily fallen into two classes: a first model implicating a direct interaction between the codon and the respective amino acid (Pelc & Welton, 1966; Woese & al., 1967; Lacey & Pruitt, 1969); however, if the present code retains the origins of its evolution, the results of Chapeville & al. (1962) which demonstrated that amino acids lose their identity once attached to tRNA, already made an amino acid-codon relationship unlikely as an amino acid selection mechanism. This led to the second model which suggested an interaction, either directly or in a common environment, between an amino acid and its anticodonic oligonucleotides (Dunnill, 1966; Ralph, 1966; Nagyvary & Fendler, 1974; Lacey & Weber, 1976). Such «anticodon first» models gained more recent support from the work by Hendry & al. (1981) who provided impressive evidence that anticodons and their corresponding amino acids show good stereochemical fits. Consequently, an amino acid - anticodon intermolecular recognition seems to have been the most likely in the primordial encoding process of amino acids by nucleobases.
Finally, it should be pinpointed out that, in the evolutive perspective provided by our pregenetical model, the primordial coding base doublets were those of tRNA anticodons. Such doublets are identical to the present-day base sequence of the leading strand—for a given gene—of DNA codons which, therefore, have been anticipated as primary code by the base sequence of tRNA anticodons but not by the only complementary ones of mRNA codons. For example, the doublet CC initially encoded glycine before becoming the first two letters of its tRNA anticodon as well as those of the corresponding DNA triplet codon in the «modern» primary code.

2. Insertion of ribose in to prenucleic polymers.

Primordial onset of a «RNA world» would have entitled the exploitation of ribozyme molecules which can both self-replicate and function as enzymes (see Gesteland and Atkins, 1993). However, attempts by Breaker & Joyce (1994) to produce a retroviral-like amplification of ribozyme (type II) were negative. Only recently, Wright & Joyce (1997) could positively turn the difficulty by ligation-dependent replication of a ribozyme in a continuous reaction of significant evolutionary potential. Nevertheless the weakest link in the whole RNA world hypothesis remains the question of its origin.

Prebiotic synthesis of RNA indeed raises questions because of its main requisite for ribose and secondly ribonucleotides (Shapiro, 1988; Oro, 1995; other ref. in Turian, 1996b). The pentose can be synthesized in the formose reaction in which polycondensation of formaldehyde (see Yuasa & al. 1995) will yield glycolaldehyde and a further triose condensation form ribose. The difficulty is that an uncontrolled use of this reaction would have produced many different sugars, not just ribose. Chemists have then been led to search for some structurally similar but simpler riboselike compounds, among which glycerol-derived acyclonucleosides (see Joyce, 1989; Orgel, 1992). Pyranosyl-RNA has also been substituted to the normal furanosyl form of ribose (Eschenmoser’s group, 1992, see Orgel, 1994). In our recently proposed scenario of the pregenetic takeover of primordial self-replicating peptide sequence (Turian, 1996b), we have also taken in account the generally admitted scarce prebiotic availability of ribose and envisaged the use of sugarless prenucleic acids to orderly line up on polynucleotides the first formed nitrogenous bases, by phosphoramidate (P-N) bonds of the creatine-phosphate type into pregenetic coding chains according to the following chemical reaction.

\[ n \text{ bases} + n \text{ tripolyphosphate} \rightarrow n \text{ base diphosphates} + n\Pi \]

as recently evidenced by hypochromic UV spectral changes and by acid-splitting of the complexes formed (Turian & Schönenberger-Solà, 1997). According to this scheme, each dehydrating N-P condensation would be driven by energetic coupling with the breaking of the first pyrophosphate bond of the tripolyphosphate.

The takeover by RNA would only then have implicated the acquisition of D-ribose by the condensed polynucleobasephosphate framework, thereby maintaining the original
informational ordering of nitrogenous bases on the newly synthesized and themselves sequentially ordered ribose molecules. To achieve such a RNA takeover, molecular topology of the prenucleic type of polynucleobasephosphate chains offers convenient possibilities because of a presumed facilitated intercalation of D-ribose on angle of each monomer by phosphoramide bond acid hydrolysis permitting N(1/9)-glycosyl ester bonding with the nucleobase, on one side, and phosphoester (5') bonding, on the other.

3. From base doublets to trinucleotide triplet codons.

The present genetic code has been largely regarded as some form of a modified doublet code. According to a landmark paper by Jukes (1965 and 1966), amino acids could be arranged into two groups: one of 15 "primordial" amino acids originally coded for by one of the 16 doublets of anticodons available in the "archetypal" code (including stop signal). This doublet code later became modified to the current triplet code proposed in 1961 by Crick & Brenner by the accretions of A/G and/or U/C; the second group of 5 «new» amino acids all coded for by triplets are supposed to have entered evolutionary history more recently. As for the discrepancy between the 64 triplet codons available and the only 20 translatable amino acids known naturally to occur, it was attributed by Crick & Brenner to a so-called code degeneracy rather than to the alternative of superfluous codons meaning non-sense. Evidently, the first two letters of each codon are primary determinants of its specificity; the 3rd position is less specific and is occupied by a base designated as "wobble" by Crick in his well known 1968 hypothesis. Nevertheless, he considered that the genetic code is and must always have been a triplet code because a change in the base would always have meant the mistranslation of essential proteins and thereby a loss of darwinian fitness. According to Eigen & al. (1981), a change from a doublet codon to a triplet one is not chemically simple because in such a transition all existing messages would be nonsense until they were completely recoded and «the genetic code must therefore have been based on a triplet frame from the start». However, Crick conceded that the primitive genetic code would have been less specific than that which now holds way. Indeed, there are reasons to admit that the initial two «letters» of the code may have been decided first, such as its non randomness suggested by the fact that chemically similar amino acids such as those defined by their acidity (see § 1) have related codons, suggesting that the recognition of classes of amino acids may have preceded the recognition of individual amino acids.

In his evolutionary scheme in which the bases enter the genetic code in a temporal sequence, Hartman (1975) later attempted to explain the transition from doublet to a triplet code by postulating a role of spacer for the 3rd base and proposed a mechanism of block polymerization of nucleotides rather than that of base pairing on the original template strand. Then, when Jungck (1978) proposed his model of archetypal genetic coding in which an amino acid can recognize only two bases of a doublet, he unfortunately did not attempt to explain further how the code became triplet from such doublet origin.
The archaic replicative mechanism further proposed by Dounce (1981) was based on the assumption that there are «fits» between the residue groups of many of the amino acids and specific nucleotide base pairs of RNA-type polymers. Amino acids would have meaningful structural relationships to the anticodons of the Jukes dyads rather than to the codons themselves, leading Dounce to the conclusion that the original coding polynucleotide strands must have given rise to the present-day adapter molecules or tRNAs if the present-day system developed stepwise from Jukes archaic coding system.

4. From proximal to distal amino acid-anticodon recognition in tRNAs.

If we assume that the code arose stepwise suggests that initial assignments of bases to amino acids resulted from chemical affinities and specific physico-chemical interactions between them. Consequently, the closest fitting bases should have been those of anticodons unless there was no direct relationship between the amino acid and the anticodon or the codon, an opinion which led Crick to invent in 1960 an adapter, later demonstrated to be tRNA, connecting the amino acid through an activating enzyme.

The role of tRNA in the translation of the genetic code also remains in need of further clarification, particularly the fact that the amino acid binds tRNA at a site far from anticodon. To account for such distant affinity preference between RNA triplets and specific amino acids, a still elusive fact as reemphasized in 1993 by Kauffman, Bedian (1982) inquisitely assumed that a given population of random peptides would be synthesized and non coded colinear polymerization could occur via something like tRNA molecules. These peptides would charge amino acids to the tRNA molecules and, with selective coevolution of the peptides, underlying coding mRNA to a consistent coded state could then occur. Finally, and particularly relevant to this persistent problem of the long intermolecular distance between the amino acid and its anticodon is the humorous remark by Maddox (1994) that «although tRNA molecules are small, there is no obvious way in which the amino acid molecules they carry can interact with their signatures, the anticodons». This signature would only have been at the closest during the prenucleic takeover when our postulated nucleobase doublets of primordial anticodons bilaterally pinched and “caged” amino acids to first recognize and then assemble them!

In the kind of molecular data tape which would direct how amino acids are incorporated into the new protein, tRNA molecules bring in these amino acids to line them up orderly along mRNA. It is in these interactions that the nucleotide triplet of the anticodon plays a determining role either by some physical chemical recognition of the amino acid it specifies or by the cavernous active site of the tRNA-associated synthetase. There would necessarily have 20 of such aminoacyl synthetases (AAS), each one providing by the folding of its suitably sequenced polypeptide a cavity specific for the side group of one specific amino acid (see Pauling’s proposals in 1960).

Moreover, how these 20 different enzymes specifically recognize amino acids and their compatible, or cognate, tRNAs? In fact, the basis for the identification of each
tRNA by its cognate AAS has not yet been completely solved, even though «there are cases in which the anticodons are the most important part of the tRNA» (see Darnell & coll., 1990).

Another puzzling fact about tRNA which remains to be resolved is that of the specificity and accuracy of interaction between mRNA codon and tRNA anticodon which depend on numerous factors that transcend the rules of base pairing (see Soll & Rajbhandary, 1994). At this level, also remains the mystery of «how do a tRNA and its synthetase recognize each other and does this amino acid recognition involve an editing step or its tRNA recognition completely idiosyncratic?» (see T. A. Steitz in Waldrop, 1989). The evolution of assignment enzymes remains the crucial step that has to be explained (Maynard Smith & Szathmary, 1995), as attempted by Szathmary (1993) in his scenario involving ribozymes recruiting amino acids cofactors. However, such a scenario could only have occurred after that the evolutionary difficulties encountered in the synthesis of oligoribonucleotides (see p. 97) would have been overcome.

DISCUSSION

The search for a physical basis of genetic coding dates back to Gamow (1954) who, in one of the earliest models, suggested that amino acids literally fit into the rhombic cleft between three contiguous base pairs in a DNA double helix. Since these pioneer times, two opposing views have been proposed on this origin: 1) the purely stochastic view which envisages a code selected by circumstance and that, by virtue of its «workability» would be the best code; as no sensible stereochemical or other interaction between amino acids and codons or polynucleotides has been discerned to date (Zubay & Doty, 1958; Crick, 1958), Crick in 1968 privileged this view; as many other scientists, he believed that nucleic-amino acid interaction was the result of an inexplicable, «frozen» accident; 2) the alternate view which envisages a code resulting ultimately from specific interactions between amino acids and nucleotides or polynucleotides, with Woese and associates (1967) as his major proponent. Nowadays, there is still little convincing evidence for either the hazardous or the specific interaction hypotheses. In favor of some type of stereochemical relationship between a polynucleotide and its corresponding amino acid, it can be mentioned the work of Woese (1966-67) who could marshall a case for some form of amino acid-nucleotide interaction as the basis for the genetic code but the search of available evidence for the site recognizing the amino acid on tRNA remained contradictory. Pelc & Welton and (1966) suggested that it was the codon - a possibility found stereochemically unacceptable by Crick (1966) while, parallely, Dunnill (1966) sought to find some relationship between the anticodon trinucleotides and amino acids and proposed that «the anti-codon triplet for an amino-acid in a transfer RNA molecule provides a cage which will bind amino-acids». Unfortunately, as commented by Ralph in 1968, no further evidence for amino acid-anticodon fit has been forthcoming. Nevertheless, in 1978, Jungck further
established correlations between the properties of amino acids and their respective anticodon dinucleotides and concluded from such correlations that interactions between amino acids and their anticodons were the basis of the genetic code. Nelsestuen (1978) also presented a rationale based on an amino acid-directed nucleic acid synthesis. He proposed that the amino acid R groups are inserted between nucleotide base pairs adjacent to the ester linkage between the α-carboxyl groups of the amino acids and the C₂-OH groups of the ribose components of the nucleotide. He also assumed the occurrence of trinucleotide adapter molecules and also rationally stated that ‘any hypothesis that explains the origin of life from the prebiotic molecular environment must provide the interactive link between amino acid and nucleic acid structure which led to nucleic acid-directed protein synthesis’. Consequently, Nelsestuen theorized a plausible mechanism for the copolymerization of these molecules implicating as basic structural unit a nucleotide substituted with an amino acid at the 2’ position of the ribose component of the nucleotide. He expected it to provide a concrete rationale opposed to the belief of many scientists that nucleic-amino acid interaction was the result of a stochastic event such as a ‘frozen accident’. Also of interest is the case of a pentanucleotide which turned out to be a double-sided template for a primitive decoding system able to nestle an amino acid via hydrogen bonds in a complex found to be an energetically favourable conformation (Balasubramanian & Seetharamulu, 1985). These works confort us in our proposal of a primordial selective recognition of amino acids from primitive peptides by nucleobases stabilized by phosphoramid bonds on polyphosphate chains (Turian, 1996b).

As astutely commented by Maddox in 1994 «an understanding of how the code evolved must certainly be a pointer to the origin of itself... so it is disappointing, but not surprising, that the origin of the genetic code is still as obscure as the origin of life itself». Its «central dogma» enunciated by Crick asserts the one-way transfer of information from DNA-to-RNA-to protein. However, it has been challenged, apart from reverse transcriptase, by the proposition by Mekler (1967) of reverse translation related to antigene-antibody synthesis, a proposal further advocated by Mehta in 1986. Cook (1977) also further discussed such a reversal process in its relevance to adaptative evolution by generating complexity by responsiveness to foreign environmental molecules rather than by neo-darwinian random mutations-selection. The inversion of the code by a backwards mechanism to its present state was suggested by Root-Bernstein (1982) to occur as «a result of the evolution of tRNA molecules which supplanted parallel codon-amino acid interactions with antiparallel codon-anticodon ones». This so-called «reverse genetic code» has been deciphered by Davydov (1983, see Poglazov & al., 1995) according to which one-atom different amino acids are encoded by codons with one-base different doublets.

Our stepwise model of molecular coding evolution also involves in its pregenetic takeover the retrotranslation of amino acids from the primordial coding peptidic chains by coding doublets of nucleobases stabilized on polyphosphate chains as prenucleic polymers which then functioned as anticodons when anterotranslation occurred in the
«modern» direction. These still «nude» anticodons of nucleobase doublets postulated to bilaterally pinch and «cage» amino acids to recognize them thus act as their most proximal «signature». However, it would have been only after ribosylation of the pre-nucleic polybasephosphate chains during their takeover by ribonucleic acids and the advent of tRNA aminoacyl synthetases that intervened a progressive distancy between the amino acids and their anticodons. Only that later availability of ribose allowed the second, prebiotic takeover of the nascent prenucleic code by that of the RNA instruction manual which, as already defended by several, origin of life scientists led by Carl WOESE was therefore not the first spark of life but, according to the provoking comments by Phil COHEN (1996), «a mere evolutionary afterthought that helped fan its flames».

We thank Sandrine Girard-Turian for her help in typing the manuscript.

RÉSUMÉ

Polarité à l’origine du codage génétique. II. Reconnaissance primaire des amino acides par les doublets de bases de polymères dépouvvus de sucres puis en relève par les acides ribonucléïques. - Selon notre modèle de codage moléculaire prégénétique, les doublets de bases d’anticodons ont été, par affinités stéréochimiques, les codons primaires pour les lettres d’amino acides des séquences peptidiques primitives. De tels doublets, anticipant les séquences de bases des codons «modernes» à triplets des brins lecteurs de l’ADN, auraient été originellement stabilisés le long de chaînes polyphosphates par condensation déshydratante phosphoamidique avec des tri-polyphosphates en acides sans ribose prénucleïques.

Mots-clés: Polarité, Génétique, Codage, Polymères prénucleïques, Acides ribonucléïques.

REFERENCES


