J. theor. Biol. (1993) **164**, 291–305

Simulation Studies of Self-replicating Oligoribotides, with a Proposal for the Transition to a Peptide-assisted Stage†

RICARDO FERREIRA‡|| AND KALINE R. COUTINHO§

‡ *Department of Physics, University of California, San Diego, La Jolla, CA 92093-0354, U.S.A. and* § *Departamento de Fisica, Universidade Federal de Pernambuco, 50739 Recife, Pernambuco, Brasil*

(*Received on* 21 *April* 1992, *Accepted in revised form on* 19 *February* 1993)

A two-substrate Michaelis–Menten mechanism previously proposed for the self-replication of RNA-like oligomers is developed. Differential growth depends on the existence of two pairs of complementary monomers and leads to 2^n groups of 2^n components each (n is the oligomer size). As n increases the 2^n groups tend to overlap with one another, and the efficiency of the process to increase the information content of the strands decreases. In a second stage we suppose that randomly synthesized peptides with one predominant amino acid interacted with the ribotides, increasing the growth rate of some of them, and at the same time had their mean life increased by interactions with other ribotides of the same kinetic group. Natural selection could have preserved a favourable codon-anticodon–amino acid correlation, the precursor of the modern genetic code.

1. Introduction

The discovery of catalytic RNA (Kruger *et al.*, 1982; Guerrier-Takada *et al.*, 1983; Cech, 1986*a*, *b*, 1987; Gilbert, 1986; Westheimer, 1986; Zaug & Cech, 1986), in particular the announcement (Doudna & Szostak, 1989) of RNA-catalysed synthesis of complementary-strand RNA, has strengthened the proposals (Woese, 1967; Crick, 1968; Orgel, 1968) that self-replicating RNA-like oligomers were the first prebiotic systems to appear on the primitive Earth. According to recent estimates this RNA world started about 4×10^9 years ago, that is by the end of the first 500 million years of the Earth's history (Joyce, 1989). Joyce lists several lines of evidence in favour of this view and recalls that RNAs are the only molecules known to function both as genotypes and phenotypes, a fact which means that replication of RNA fragments enables Darwinian evolution to occur at a molecular level. As shown by Eigen (Eigen, 1971; Eigen & Schuster, 1978) some RNAs will replicate faster than others and will grow to dominate the population until some environmental change will give to another RNA a still greater selective advantage.

In Oparin's original proposal (1924) life began with self-replicating proteins. Polypeptides can be efficient catalysts, and it is probable that they were synthesized abiotically on the primitive Earth (Fox & Dose, 1977). Although Dyson (1982) and

† Work supported in part by CNPq (Brazilian Government Agency).

|| On leave of absence from the Departamento de Quimica Fundamental, Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil.

0022–5193/93/019291 + 15 $08.00/0

© 1993 Academic Press Limited

Kauffman (1986) showed that a sufficiently large assembly of random polypeptides could give rise to a self-sustained network of proteins, and in spite of model-building efforts (Root-Bernstein, 1982, 1983), there is no experimental evidence that random polypeptides can catalyse the formation of peptide chains with any significant degree of sequence conservation. On the other hand, nucleotides have inherent template properties and there is now a well-assessed body of experiments on non-enzymatic template-directed synthesis of informational oligonucleotides (Lohrmann & Orgel, 1977; Inoue & Orgel, 1982, 1983; Inoue *et al.*, 1984; Joyce *et al.*, 1984; Orgel, 1986, 1987; von Kiedrowski, 1986; Joyce, 1987; Zielinski & Orgel, 1987). These are landmarks in our understanding of biogenesis, but since template-free self-replication of informational molecules has not been achieved, there is a valid space for theoretical models, providing the studies do not conflict with the body of knowledge established by experimental work.

It is known, for example, that substrates must be activated through binding with "energy-rich" phosphoanhydride groups before condensation reactions can take place. Other information is that deoxyribonucleotide substrates are much less efficient than ribonucleotide ones (Lohrmann & Orgel, 1978). Effective template-directed syntheses are possible only if they are carried out at temperatures below the melting point of the template–substrate complex (Joyce, 1987; Orgel, 1987). It is also necessary that the stereochemistry of the reactive complex brings the template-bound substrates together in a favourable orientation. Other difficulties have been detected in the experimental demonstration of polyribotide self-replication. For example, molecules with regions of self-complementarity form intramolecular bonds (Joyce & Orgel, 1986). Another obstacle is the possibility that nucleotide derivatives of sugars other than D-ribose, formed in prebiotic conditions, would incorporate into the growing chains and act as chain terminators. Enantiomeric cross-inhibition supports the case for an ancestral genetic system involving more flexible acyclic analogues of the nucleotides (Joyce *et al.*, 1987).

On the basis of the highly specific template properties of tetranucleoside triphosphoramidates Zielinski & Orgel (1989) have stated that "theories that treat systems of replicating polynucleotides by statistical approximations that fail to recognize the individuality of particular sequences are unlikely to provide an adequate description of their evolution". It is true that parts of the detailed mechanism of oligomer formation, such as the need for the presence of ions like Mg^{++} for initial nucleation, have been studied by Felsenfeld & Miles (1967), Ts'o (1974) and Porschke (1978) among others. Our justification for not trying to incorporate such relevant work in our study is that not all significant early prebiotic conditions are known at the present time. For example, prebiotic condensation reactions could have been favoured by non-specific conditions such as low-activity water, present in aqueous solutions of ethylene–glycol and similar substances (de Meis, 1989). In our model, starting with monomers and short oligomers, we make the assumption that dimers and larger fragments can act as templates as well as increase the growth rate of oligomers. We introduce no more specificity than to envisage the activated complexes as small helices involving no more than three base pairs. This is not in contradiction with the fact that the first double helical complexes that live long

enough to allow for further ligation are hexamers, since our complexes represent the transition-states of the condensation reactions.

2. Self-replicating Ribonucleotide Oligomers

2.1. THE MODEL

A preliminary version of our model has been published (Ferreira, 1987, 1988). We proposed that the condensation reactions between two small random fragments followed a two-substrate Michaelis–Menten mechanism, with a third fragment acting as a catalyst. As known, the Michaelis–Menten model is an equilibrium approximation to a steady-state process first described by Briggs & Haldane (1925). Given the idealized nature of our system we feel justified in making this approximation.

We start with a mixture of ribotide-like molecules with a phosphate group in a position we label 5′. These molecules contain another reactive centre which we label 3′ and consisting of a free OH group. Condensation reactions occur between the 5′ PO_3OH group of one molecule and the 3′ OH group of another molecule. For reactions between unequal sized molecules it is necessary to assign in the condensation products the origin of the remaining free PO_3OH group, i.e. whether it belonged to the larger or the smaller fragment. Scheme (1) below corresponds to leaving the 5′PO_3OH group of the *larger* fragment free (we call this a 5′ → 3′ growth). The other possibility, shown in scheme (2) below, corresponds to leaving free the 5′PO_3OH group of the *smaller* reactant (we call this a 3′ → 5′ growth). One example of a 5′ → 3′ growth is the following synthetic route for the pentamer 5′pUAAGC;

$$\begin{aligned} &5'\text{pU}-\text{A}-\text{A}+5'\text{pG-C} \xrightarrow[5'\text{pG-C-U}]{} (5'\text{pU-A-A}).(5'\text{pG-C}) \longrightarrow \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \dot{\text{U}}\cdots\dot{\text{C}}\cdots\dot{\text{G}}\text{p}5' \\ &\longrightarrow (5'\text{pU-A-A-G-C}) \longrightarrow 5'\text{pU-A-A-G-C}+5'\text{pG-C-U}. \\ &\qquad\qquad\quad \dot{\text{U}}\text{-}\dot{\text{C}}\text{-}\dot{\text{G}}\text{p}5' \end{aligned} \tag{1}$$

In our notation, A, C, G and U are intended to represent monomers of the class of *the activated ribonucleotides. We consider all possible routes leading to the chosen* pentamer (in this case 5′pUAAGC) with definite restrictions in the upper size of the catalytic fragments (see below). For example, one of the other possible routes to 5′pUAAGC is a 3′ → 5′ growth:

$$\begin{aligned} &5'\text{pU}-\text{A}+5'\text{pA-G-C} \xrightarrow[5'\text{pC-U-U}]{} (5'\text{pU-A}).(5'\text{pA-G-C}) \longrightarrow \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \dot{\text{U}}\cdots\dot{\text{U}}\cdots\dot{\text{C}}\text{p}5' \\ &\longrightarrow (5'\text{pU-A-A-G-C}) \longrightarrow 5'\text{pU-A-A-G-C}+5'\text{pC-U-U}. \\ &\qquad\qquad\quad \dot{\text{U}}\text{-}\dot{\text{U}}\text{-}\dot{\text{C}}\text{p}5' \end{aligned} \tag{2}$$

We can make these two routes to 5′pUAAGC equally probable or we can put a directional bias in the simulation. Although ribozymes act contrary to the polymerase-assisted direction of growth (Westheimer, 1986), in the following we will

discuss results obtained by privileging the 5′→3′ direction; it is equally easy, of course, to privilege the opposite direction. Note that the growing and the catalytic fragments interact in the anti-parallel (Watson–Crick) form.

For each of the synthetic pathways to a given oligomer we write the corresponding Michaelis–Menten rate expression. Thus, in the case of 5′pUAAGC and according to scheme (1) we will have:

$$v = \frac{k_5[5'\text{pUAA}].[5'\text{pGC}].[5'\text{pGCU}]}{K_{M_1}.K_{M_2}} \tag{3}$$

where K_{M_1} and K_{M_2} are the Michaelis constants for the process:

$$5'\text{pGCU} + 5'\text{pUAA} + 5'\text{pGC} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} 5'\text{pUAA} + 5'\text{pGC} \underset{k_4}{\overset{k_3}{\rightleftharpoons}} \underset{\dot{\text{U}}\text{CGp5}'}{\dot{}} \tag{4}$$

$$(5'\text{pUAA}).(5'\text{pGC}) \overset{k_5}{\rightarrow} 5'\text{pUAAGC} + 5'\text{pGCU}.$$
$$\dot{\text{U}} \text{---} \dot{\text{C}}\text{Gp5}'$$

That is:

$$K_{M_1} = \frac{k_2}{k_1} = \frac{[5'\text{pUAA}].[5'\text{pGCU}]}{[5'\text{pUAA}.5'\text{pGCU}]} \tag{5}$$

and

$$K_{M_2} = \frac{k_4}{k_3} = \frac{[5'\text{pUAA}.5'\text{pGCU}].[5'\text{pGC}]}{[5'\text{pUAA}.5'\text{pGCU}.5'\text{pGC}]}. \tag{6}$$

We make the assumption that the Michaelis constants are related to the A::U and G⋮⋮C interaction constants in Watson–Crick complementary pairs. Thus, from eqn (4) we write $K_{M_2} = (K_{\text{AU}})^{-1}$, and $K_{M_2} = (K_{\text{CG}})^{-2}$, where K_{AU} and K_{CG} are the formation constants of the corresponding complementary pairs *under the conditions of the reaction.* In particular we assume that the ratio of the Watson–Crick pairing constants, $K_{\text{CG}}/K_{\text{AU}}$ is conserved under the conditions of the reaction.

Equation (3) can be written as:

$$v = k_5 K_{\text{AU}} K_{\text{CG}}^2 [5'\text{pUAA}].[5'\text{pGC}].[5'\text{pGCU}]. \tag{7}$$

We make the further assumption that k_5, the rate constant for the (almost) irreversible step of the Michaelis–Menten mechanism, is independent of the specific nature of the ribotides involved. We know that in experiments involving template polymers base-specific differences are observed (Zielinski & Orgel, 1989). But we think our assumption of constant k_5 is a reasonable approximation for ribonucleotides, since the different bases (A, C, G and U) are about 5 Å away from the 0(3′) and 0(5′) atoms of D-ribose. Thus, in the case of the small helical strands which form the transition states in our model all specificity arises from differences in the interaction constants K_{AU} and K_{CG}. Further base discrimination, the result of differences in the value of k_5, will appear only in a second stage, involving nucleotide–peptide interactions.

Adding all possible synthetic routes for each of the oligomers (64 trimers, 256 tetramers, 1024 pentamers; in general, 4^n oligomers) we obtain their growth rates as a function of all possible products of the type $k_5 K_{AU}^m K_{CG}^q$, as well as of the concentrations of the reactant and catalytic fragments. If we are restricted to relatively small oligomers ($n \leq 10$) and even allowing k_5 to vary by a factor of 2, the growth rate of the various strands depends critically on the bonding constants K_{AU} and K_{CG}. If they were equal, all possible oligomers of a given size would grow at the same rate and *selection could not occur*. In his famous paper on the origin of the genetic code, Crick (1968), answering the question "Why 4?", argued that if originally there were only two bases in the nucleic acid (a suggestion retaken by others such as Hartman, 1984), the code did not survive possibly because 2 was too restrictive a number. If our proposed autocatalytic stage did occur, the minimum number of bases required was 4 from the very start (four bases correspond to two distinct interaction constants).

2.2. RESULTS

A very simple program allows for the simulation of the growth-process according to the proposed mechanism. Given a fragment of size n we privilege one direction of growth making $k_5(5' \rightarrow 3') > k_5(3' \rightarrow 5')$. We have only a few clues as to the numerical values of K_{AU}, K_{CG} and k_5. In complementary chain interactions it is known that the free-energy of formation of a GC pair is approximately 1·5 times greater than that of an AU pair. We assume that this ratio holds in the conditions of the reactions. The constants are considered to be independent from one another, that is, the corresponding free-energies are additive. The products of the form $K_{AU}^m K_{CG}^q$ are the inverse of the Michaelis constants for our reactions. For most enzymes the Michaelis constants lie between 10^{-6} and 10^{-2}. Since our catalytic fragments must be poor catalysts we put 10^2 as the higher limit for the products $K_{AU}^m K_{CG}^q$. For example, we made sets of calculations using $K_{AU} = 3$, $K_{CG} = 6$, up to $K_{AU} = 3$, $K_{CG} = 15$. The rate constant k_5 of the irreversible step in enzymes varies between 10^2 and 10^6 sec^{-1}, whereas for non-catalytic aqueous processes it is as low as 10^{-6} sec^{-1}. For our systems we used the intermediate value 10^{-2} sec^{-1}. In fact, for each pair of K_{AU}, K_{CG} constants we varied k_5 from 2×10^{-2} sec^{-1} (for the purine–purine condensations) to 4×10^{-2} sec^{-1} (pyrimidine–pyrimidine condensations). We found that within those limits the 4^n possible oligomers are split in 2^n groups, each containing 2^n members.

In the renormalization group approach to this problem (Ferreira & Tsallis, 1985; Tsallis & Ferreira, 1986) we proposed that the catalytic fragments could be no larger than the sum $(m+q) = n$ of the growing fragments. Although the growth of stable double helices requires the formation of nucleation stages with a minimum of five or six monomers, from our knowledge of enzymatic catalysis the active sites must be restricted to a maximum of three ribotides. To take fair account of the difference between K_{AU} and K_{CG} we substitute $K_{AU}^{m'} K_{CG}^{q'}$ in the various synthetic routes for $K_{AU}^m K_{CG}^q$, and define:

$$m' = \frac{3}{(m+q)} m \tag{8}$$

and:

$$q' = \frac{3}{(m+q)}\, q. \tag{9}$$

In the first type of calculations made the "reaction chamber" is supposedly infinite in size and the concentrations of the monomers are taken as equal, that is [A+U] = [C+G]. The concentrations remain constant throughout the process. Table 1 shows the distribution of the groups in the case of pentamers. We have also made simulations for hexamers, etc up to dodecamers, but the numbers become too large for convenient handling in tables.

Clearly, from Table 1, the 2^n groups correspond to the possible sequences of strong (s) and weak (w) interactors: (1) 5'p-sssss, (2) 5'p-ssssw, (3) 5'p-wssss, . . ., (31)

TABLE 1

Differential growth of pentamers†

Rank	Group representative	Growth rate (relative)	Rank	Group representative	Growth rate (relative)
1	5'pCGCCC	1125	17	5'pAACGA	120
2	5'pCGCCA	698	18	5'pAUCCA	119
3	5'pAGCCC	690	19	5'pAGACA	94
4	5'pCGCAC	437	20	5'pCUCAA	81
5	5'pCUCCC	431	21	5'pCGAAA	80
6	5'pAGCCA	369	22	5'pAUCAC	79
7	5'pCGACC	323	23	5'pAUACC	78
8	5'pCGCAA	306	24	5'pCUACA	59
9	5'pAUCCC	298	25	5'pAGAAC	58
10	5'pCUCCA	211	26	5'pCUAAC	40
11	5'pAGCAC	210	27	5'pAUCAA	35
12	5'pGCACA	190	28	5'pAGAAA	32
13	5'pUCACC	188	29	5'pAUACA	31
14	5'pGACAC	141	30	5'pCUUAA	22
15	5'pGCAAC	123	31	5'pAUAAC	21
16	5'pGAACC	121	32	5'pAUAUA	10

Pentamers of rank 10 (*5'pswssw*)

5'pCACCA	5'pCUCCA	5'pCAGCA	5'pCUGCA
5'pCACGA	5'pCUCGA	5'pCAGGA	5'pCUGGA
5'pCACCU	5'pCUCCU	5'pCAGCU	5'pCUGCU
5'pCACGU	5'pCUCGU	5'pCAGGU	5'pCUGGU
5'pGACCA	5'pGUCCA	5'pGAGCA	5'pGUGCA
5'pGACGA	5'pGUCGA	5'pGAGGA	5'pGUGGA
5'pGACCU	5'pGUCCU	5'pGAGCU	5'pGUGCU
5'pGACGU	5'pGUCGU	5'pGAGGU	5'pGUGGU

† $K_{AU} = 3{\cdot}00\ k_5(3'\text{OH} \rightarrow 5'\text{P}) = 0{\cdot}020$.
$K_{CG} = 15{\cdot}00\ k_5(5'\text{P} \rightarrow 3'\text{OH}) = 0{\cdot}021$.
Catalytic fragments normalized to 3.

5'p-wwwws, and (32) t'p-wwwww. In other words, each group is composed of sequences which correspond to different hydrogen-bond patterns. Because K_{CG} is larger than K_{AU}, for equal initial concentrations of monomers the C, G-rich oligomers dominate the A, U-rich ones. However, we have also used a Monte Carlo method to include the variation in the initial concentrations of (C + G) and (A + U). In equations such as (7) the concentrations of the fragments are expressed in differing powers of (C + G) and (A + U), which appear with different probabilities. It is easily shown that the growth rates change; for example, if the initial concentration of (A + U) is 1·5 times that of (C + G), the group of rank 10 in Table 1 will become the fastest growing one. Exhaustion of monomer concentration, hydrolysis of the oligomers, and the formation of intra-chain duplexes can be introduced in our model to limit the chain size.

If one of the groups of oligomers of size n obtained complete dominance, the gain in information content relative to a mixture of all possible sequences would be $\log_2 2^n = n$ bits. However, even in the simplified framework of our model a complication arises because as the size of the oligoribotides increases, the differences in the values of the constants renormalized according to (8) and (9) become small. Populations of closely related groups, i.e. those which differ only in a few positions of their base sequences, will tend to form clusters of groups growing at approximately the same rate. This is shown in Fig. 1, in which we have plotted the distribution curves of the self-growing oligomers for the cases of $n = 5$, 6 and 7. The clustering of groups of oligomers increases markedly with the size of the chains.

In modern RNAs, with unique sequences of four distinct monomers, the gain in information content is $\log_2 4^n = 2n$ bits. As shown in the next section, we propose that this further gain in information content resulted from the interaction of oligoribotides with randomly synthesized peptides.

One may ask if there are some "intellectual" vestiges of this past auto-catalytic stage in the modern DNA → RNA → protein replication system. In our opinion one such vestige lies in the very fact that the code for transcription and translation requires four ribonucleotides. In principle, sequences of only two monomers can support any information content, and, in fact, two is the required minimum in Anderson's (Anderson, 1983; Anderson & Stein, 1985) and Eigen's (Eigen, 1971; Eigen & Schuster, 1978) models. Our scheme, on the other hand, requires two pairs of complementary units from the very start.

Another vestige should be a connection with a past in which cytosine and guanosine were more abundant constituents of ribonucleotide chains than adenosine and uracyl. This is not true for modern species, 60% of which contain more A, U than C, G pairs (Mandel, 1970). The very remote past predominance of C, G pairs should appear in the present codon–amino acid relationship. Now, of the 20 amino acids of proteins, eight are determined entirely by the first two bases of their codons. We think that it is significant that out of the privileged doublets, four are the strong interactor combinations CC, CG, GC and GG and *none* are combinations of the weak interactors AA, AU, UA and UU. We have calculated (Ferreira, 1988) that the probability that this distribution be fortuitous is only 0·0032 (as against a most probable distribution of 0·1152).

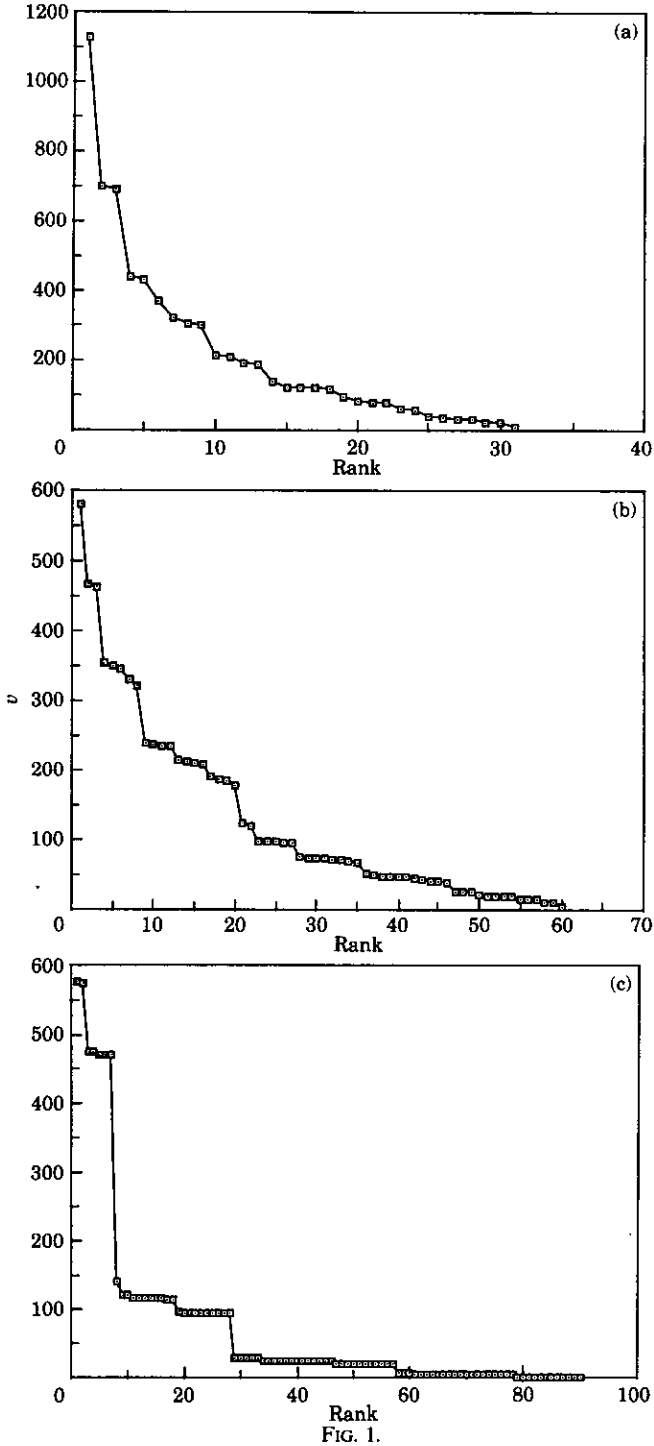
(a)
1200
1000
800
600
400
200
0
10
20
30
40
Rank
(b)
600
500
400
300
200
100
v
0
10
20
30
40
50
60
70
Rank
(c)
600
500
400
300
200
100
0
20
40
60
80
100
Rank

FIG. 1.

3. Peptide-assisted Growth of Ribonucleotides

We have seen that with increasing size of the oligoribonucleotides the overlap between the groups increases and the gain in information content of the strands is reduced. Self-replication tends to become an inefficient process of information storage and transfer. How did the RNA world overcome this difficulty to the extent that co-operation was established between nucleotides and peptides, leading to the modern forms of life? Wong (1991) has pointed out that this had to be accomplished through an unbroken chain of advantageous steps.

Our proposal is that some of the oligoribotides interacted in solution in a definite way with peptides synthesized at random from a mixture of activated amino acids. The interaction was such that some of the peptides increased the growth rate of some of the self-replicating oligoribotides and, simultaneously had their mean-life (with respect to hydrolysis) increased in the presence of oligoribotides. We will show in the following how such nucleotide–peptide interaction could lead to a correlation between diribonucleotides and amino acids.

To be efficient catalysts the original peptides must have been from the beginning stereo-specific like their modern descendants, the enzymes. The reason is that the polynucleotides have very strict stereochemical requirements (Wald, 1957): no helix is formed if C(3′) and C(4′) of the ribose moiety have opposite chiralities, and no base pairing is possible if, in addition, the C(1′)s are not all of the same chirality (β-glycosidic bonds). In contrast, the monochirality of amino acids is not so demanding, and the growth and stability of small sequences of protein secondary structures, such as the α-helices, can be achieved from mixtures of D and L amino acids. It is only with respect to the tertiary structures of modern proteins that complete monochirality is demanded (Wald, 1957, 1964). It seems natural to suppose that the strict monochirality of modern amino acids was selected through the interaction of ancestor peptides with pre-existing stereo-specific RNA chains.

We suppose that the pertinent peptides at this stage were small (typically with $n = 4$), more likely on random synthesis conditions to have a predominant amino acid than larger fragments. Since, in the course of evolution, these oligo-peptides became proteins, they should have been of a size sufficient for them to form important pieces of protein secondary structures. In this sense tetra-peptides are good candidates because they allow for a complete turn of the α-helix (3·6 residues per turn). Also, the vertical distance between two consecutive bases (in DNA) is 3·4 Å, and four amino acids, either in a fully extended chain or in an α-helix can easily span two bases along an RNA chain. Enzymatically assisted RNA chains grow in the $5' \rightarrow 3'$ direction, hence we suppose that these early catalytic peptides interacted with the 3′-end of some of the growing ribonucleotides, stabilizing the transition state of the chain-growth reactions and privileging the modern $5' \rightarrow 3'$ direction.

FIG. 1. (a) Curve showing the growth rate of penta-ribotides (in arbitrary units) as a function of the growth-rate rank of the groups; each points corresponds to 32 pentamers. (b) Same for hexa-ribotides; each point corresponds to 64 hexamers. (c) Same for hepta-ribotides; each point corresponds to 128 heptamers.

We state now the basic feature for the beginning of a new stage in the evolution of macromolecules: tetrapeptides built from a single (or predominant) amino acid increased the growth-rate of RNA-like oligomers at their 3'-ends, and, simultaneously, these peptides bound non-catalytically with the chemically different 3'-ends of other ribonucleotide molecules in the same or closely related kinetic groups.

Suppose we start with a solution containing one of our pentaribotides, for example, those of rank 7 on Table 1. This corresponds to the pattern 5'p-sswss. Eight of the 32 members of this group will have a GG(3')-end: 5'pCCAGG, 5'pCCUGG, 5'pGGAGG, 5'pGGUGG, 5'pCGAGG, 5'pCGUGG, 5'pGCAGG and 5'pGCUGG. Suppose now that tetraglycine, in the presence of *reacting* fragments such as 5'pCCAGG and 5'pAUC and *template* fragments such as 5'pUCC (these being interchangeable roles) catalyses the condensation reaction:

$$\begin{matrix} 5'\text{pCCAGG} + 5'\text{pAUC} \overset{\text{GlyGlyGlyGly}}{\rightleftharpoons\!\!=\!\!=\!\!=\!\!=\!\!=\!\!=\!\!=} 5'\text{pCCAGGAUC} + \text{H}_2\text{O} \\ (+\text{CCUp5}') \qquad\qquad\qquad\qquad (+\text{CCUp5}'). \end{matrix} \tag{10}$$

In this way the concentration of the octamer 5'pCCAGGAUC will be larger than in the absence of tetraglycine. This chain-lengthening process will also occur in the cases of the other pentaribotides with GG(3')-ends.

Now, to this same group of pentaribotides of Table 1 belong eight other ribotides ending with CC(3'), such as 5'pCCACC, 5'pCCUCC, etc. Because dicytidine is different from diguanosine, it is conceivable that tetraglycine is capable of binding to the 3'-end of these ribotides to form stable complexes such as (5'pCCACC).(GlyGlyGlyGly). It is expected that the rate of hydrolysis of bound tetraglycine is much slower than that of the free peptide in solution. Hence, at the same time that tetraglycine (or other tetrapeptide in which the glycine residue predominates) gives a selective advantage to 5'pCCAGG and its companion pentamers with a GG(3') end, it would have its mean-life increased by being adsorbed by 5'pGGACC, 5'pGGUCC, etc. All peptides which in the presence of this group of ribonucleotides are not adsorbed will hydrolize faster and their end products, the amino acids themselves, will take part in the random synthesis of *de novo* peptide molecules.

This amounts to the proposal that the doublets at the 3'-ends of fragments that have their growth rate increased through interaction with specific peptides became the first two positions of modern codons, and the complementary doublets at the 3'-ends of fragments belonging to the same (or nearly the same) kinetic group of oligoribotides became the last two positions of the corresponding anticodons.

The main point of our model is, first, that ribotides such as 5'pCCAGG and 5'pCCACC belong to the same kinetic group because both are of the class 5'p-sswss. However, chemically their 3' ends are quite different: GG(3') are the first two positions of the glycine codon, and CC(3') are those of the proline codon; glycine is a polar amino acid and proline is a non-polar one. It is reasonable to assume that the same oligopeptide (tetraglycine in our example) can interact catalytically with ribotides ending in GG(3') and in a different way with ribotides ending in CC(3').

Thus we come to a picture in which randomly synthesized oligopeptides with a given predominating amino acid gave a selective advantage to oligoribotides with certain doublet 3′ terminals by lowering the activation energy (i.e. stabilizing the transition state) of condensation reactions involving these terminals; at the same time these peptides would be protectively bound to companion fragments ending with the anticodonic doublet 3′ terminal.

A causal relationship was in this way established between a given amino acid residue and a doublet of ribonucleotides. This established a concentration correlation, that is, an environmental break from randomness, which, through natural selection, lead to a codon–amino acid relationship that has survived in modern organisms.

4. A Classification of Oligoribotides

Inspection of the 16 possible doublet 3′ terminals shows that with respect to the proposed scheme the oligoribotides fall into three classes:

(i) Oligoribotides with the 3′ terminals GG, AA, CC and UU, which conform with the double requirement that a given peptide may act as a catalyst for their growth and simultaneously may be adsorbed by the 3′ terminals of other ribotides of the same kinetic group.

(ii) Oligoribotides with the 3′ terminals CU, UC, AC, CA, AG, GA, GU and UG, which conform with a less strict double requirement, namely, that a given peptide may act as a catalyst for their growth and simultaneously be adsorbed by the 3′ terminals of other ribotides of a closely related group. For example, eight pentamers with the sequence 5′p-swsGU (group 10 of Table 1) could have their growth rates increased by a tetrapeptide made solely or predominantly of L-threonine, which, by its turn, could be adsorbed and protected from hydrolysis by pentaribotides with the sequence 5′p-swsAC (group 14 of Table 1). We have shown that as the size of the ribotides grows, groups which differ only in the positions of two bases tend to belong to the same cluster.

(iii) Oligoribotides with the 3′ terminals CG, GC, AU and UA, which do not conform with the double requirement. This results from the fact that the bases of the first two positions of a codon are the same as the bases of the last two positions of the corresponding anticodon. Hence, if a given peptide could favour a chain-lengthening process through the "codonic" 3′-end of a given oligoribotide, it would act in the same way with respect to the "anticodonic" ribotides of the same group. The interaction would not increase the life expectancy of the peptides and it would be an evolutionary dead-end.

5. Discussion

It is assumed that these processes through which correlated concentrations of amino acids and diribotide residues could accumulate in the primitive medium belong to the pretranslational stage and that they also antedate the *inventive biosynthetic* stage (phase 2 of Wong; Wong, 1975, 1976, 1981, 1988).

In doublets of class (i) the correlation could have started at a stage in which there existed small oligoribotides. The corresponding very ancient amino acids are Gly, Pro, Lys, Asn, Leu and Phe. To class (ii) doublets correspond the amino acids Val, Ser, Thr, Arg, Asp, Glu, Cys, Trp, His and Gln. Given the uncertainties of the model it is impossible to say if these amino acids are somewhat more recent than the first ones, but the correlation process in this case could start only for oligoribotides of moderate size ($n \geq 10$). Finally, the proposed form of codon–amino acid relationship *could not* have originated by this mechanism in the cases of Tyr, Ile, Met and Ala (nor Arg through codons starting with CG).

The proposed model is, in principle, testable. Firstly, it should be possible to measure experimentally the interactions between a given peptide and oligoribotides, respectively containing and not containing the corresponding anticodonic 3'-end. A stronger attractive interaction is expected in the first case, if we can extrapolate from the known correlation between physical properties of amino acids and their anticodonic ribotides (Dunnill, 1966; Grantham, 1974; Lacey & Weber, 1976; Jungck, 1978; Weber & Lacey, 1978; Lacey & Mullins, 1983).

One can also look for clues in the modern code which support our mechanism. We believe such supporting evidence exists. Thus, the codons for methionine (AUU) and for two of the terminators (UAA and UAG) are among those which in our model could not have originated by the proposed mechanism. Now, methionine signals for the start of protein synthesis in modern organisms and must have been a relatively recent acquisition. The same is true for the releasing factors coded by UAA and UAG. It is true that the triplet UGA, which belongs to our class (ii) of doublets, normally codes for a third terminator, but it is known that in yeast mitochondria (and possibly in other cases) UGA is used as a tryptophan codon, like UGG (Macino *et al.*, 1979). This is considered an indication that the use of UGA as a terminator is relatively recent.

Homologies between the tRNAs of Ile and Tyr and of other amino acids (Mullins *et al.*, 1973; Staves *et al.*, 1987) indicate that these two amino acids [both belonging to our class (iii)] received anticodonic assignments late in evolution.

Another clue is the discovery that in 5S rRNAs the frequency of adenine dinucleotide decreases with increasing organismic complexity (Guimaraes & Erdmann, 1989). There are also indications of greater frequency of other homeodinucleotides in the more ancient 5S rRNAs (Subacius, 1990). Although ribosomal RNAs are not as much involved with information transfer as mRNAs and tRNAs, these findings could be significant to our model, since the homeodinucleotides belong to our class (i).

The only discordant case is that of alanine, which was presumably abundant in the prebiotic medium but belongs to our class (iii).

It should be pointed out that the peptides with a dominant amino acid residue were not the ancestors of modern DNA-directed RNA polymerases, nor of a previously existing RNA-directed RNA polymerase (Lazcano *et al.*, 1988). In fact, polymerases smooth the base differences, thereby making possible the growth of a very broad range of base sequences (Benner, 1989) whereas our "proto-enzymes" would alter specifically the k_5 rate constants. Our peptides more likely were

ancestors of modern *ligases*; modern ligases require a DNA template, but it is believed that they had an RNA-directed precursor (Weiner & Maizells, 1987).

Since Gamow's original paper (Gamow, 1954) stereochemical theories of the origin of the code suppose that there was a direct interaction between the amino acids and their anticodons. Dunnill (1966), Grantham (1974) and especially Lacey & Weber (1976), Weber & Lacey (1978) and Jungck (1978) have shown that, with the exceptions of Ile, Tyr, Trp and the GC anticodon of Arg, there is a strong correlation between properties such as polarity and hydrophobicity of a given amino acid and those of the corresponding anticodonic dinucleotide. These results point to an early causal relation between anticodons and amino acids. Still another kind of direct interaction theories of the origin of code are those proposed by Sonneborn (1965) and Woese *et al.* (1966), according to which there are selective advantages in assigning to amino acids that are similar in physical properties neighbouring codons that differ from one another in a single base (protection against damage owing to mutational errors). But the calculated spectrum of chemical distances between neighbouring amino acids in the genetic code has not shown any distance minimization (Salemme *et al.*, 1977; Wong, 1980).

Such difficulties led Crick (1968) to propose his "frozen accident" theory, implying that the earliest relation between codons and amino acids was a matter of chance; the code became universal (or almost; see Jukes, 1983) because any change would be harmful.

Although Crick's view is largely accepted today, models suggesting other forms of early ribonucleotide–amino acid interactions have been proposed. Orgel (1989), for example, suggested that an amino acid or a dipeptide could attach to the 3′ OH group of RNA fragments, providing an incentive to the synthesis of aminoacyl-tRNAs (or of dipeptidyl-tRNAs), essential intermediates in modern translation. The reaction of peptides with primordial RNAs has also been proposed by Wong (1975, 1976, 1990).

It is possible that the correlation of physical properties of amino acids and dimeric ribotides contributed to the non-randomic nature of the origin of the code at a very deep level, by favouring the cosolubilities of certain peptides and ribotide fragments. Concentration correlations then may have started through a mechanism like the one proposed in this paper. It could perhaps be said that we have followed Orgel's prescription (Orgel, 1989) that protein synthesis "was probably preceded by simpler processes by means of which interaction with amino acids conferred selective advantage on replicating RNA molecules".

REFERENCES

ANDERSON, P. W. (1983). *Proc. natn. Acad. Sci. U.S.A.* **80,** 3386–3390.
ANDERSON, P. W. & STEIN, D. (1985). *Self-organizing Systems: The Emergence of Order* (Yates, F. E., ed.). New York: Plenum Press.
BENNER, S. A. (1989). *Science* **236,** 789–806.
BRIGGS, G. E. & HALDANE, J. B. S. (1925). *Biochem. J.* **29,** 338–340.
CECH, T. R. (1986*a*). *Proc. Nat. Acad. Sci. U.S.A.* **83,** 4360–4365.
CECH, T. R. (1986*b*). *Sci. Am.* **255,** 44–48.

CECH, T. R. (1987). *Science* **236,** 1532–1539.
CRICK, F. H. C. (1968). *J. molec. Biol.* **18,** 367–379.
DE MEIS, L. (1989). *Biochem. Biophys. Acta.* **973,** 333–349.
DOUDNA, J. A. & SZOSTAK, J. W. (1989). *Nature* **339,** 519–522.
DUNNILL, P. (1966). *Nature, Lond.* **210,** 1267–1268.
DYSON, F. J. (1982). *J. molec. Evol.* **18,** 344–350.
EIGEN, M. (1971). *Naturwiss* **58,** 465–522.
EIGEN, M. & SCHUSTER, P. (1978). *Naturwiss* **65,** 341–369.
FELSENFELD, G. & MILES, H. T. (1967). *Ann. Rev. Biochem.* **36,** 407–448.
FERREIRA, R. & TSALLIS, C. (1985). *J. theor. Biol.* **117,** 303–313.
FERREIRA, R. (1987). *J. theor. Biol.* **128,** 289–295.
FERREIRA, R. (1988). *Leite Lopes Festschrift.* pp. 241–252. Singapore: World Scientific Co.
FOX, S. W. & DOSE, K. (1977). *Molecular Evolution and the Origin of Life.* New York: M. Dekker.
GAMOW, G. (1954). *Nature, Lond.* **173,** 318–320.
GILBERT, W. (1986). *Nature, Lond.* **319,** 618.
GUERRIER-TAKADA, C., GARDINER, K., MARSH, T., PACE, N. & ALTMAN, S. (1983). *Cell* **35,** 849–855.
GUIMARAES, R. C. & ERDMANN, V. C. (1989). *Origin Life, Int. Conference of Prague* **9,** 123–124.
GRANTHAM, R. (1974). *Science* **185,** 862–864.
HARTMAN, H. (1984). *Origin Life* **14,** 643–648.
INOUE, T. & ORGEL, L. E. (1982). *J. molec. Evol.* **16,** 201–217.
INOUE, T. & ORGEL, L. E. (1983). *Science* **219,** 859–862.
INOUE, T., JOYCE, G. F., GRZESKOWIAK, K., ORGEL, L. E., BROWN, J. M. & REESE, C. B. (1984). *J. molec. Biol.* **178,** 669–676.
JOYCE, G. F. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52,** 41–51.
JOYCE, G. F. (1989). *Nature, Lond.* **338,** 217–224.
JOYCE, G. F., INOUE, T. & ORGEL, L. E. (1984). *J. molec. Biol.* **176,** 279–306.
JOYCE, G. F. & ORGEL, L. E. (1986). *J. molec. Biol.* **181,** 433–438.
JOYCE, G. F., SCHWARTZ, A. W., MILLER, S. L. & ORGEL, L. E. (1987). *Proc. natn. Acad. Sci. U.S.A.* **84,** 4398–4402.
JUNGCK, J. R. (1978). *J. molec. Evol.* **11,** 211–224.
KAUFFMAN, S. A. (1986). *J. theor. Biol.* **119,** 124–135.
KIEDROWSKI, G. (1986). *Ang. Chem. Int. Ed.* **25,** 930–935.
KRUGER, K., GRABOWSKI, P. J., ZAUG, A. J., SANDS, J., GOTTSCHLING, D. E. & CECH, T. R. (1982). *Cell* **31,** 147–158.
LACEY, J. C. & WEBER, A. L. (1976). *Protein Structure and Evolution* (FOX, L., *et al.*, eds) pp. 213–222. New York: M. Dekker.
LACEY, J. C. & MULLINS, D. W. (1983). *Origin Life* **13,** 3–42.
LAZCANO, A., FASTAG, J., GARIGLIO, P., RAMIREZ, C. & ORÓ, J. (1988). *J. molec. Evol.* **27,** 365–376.
LOHRMANN, R. & ORGEL, L. E. (1977). *J. molec. Biol.* **113,** 193–205.
LOHRMANN, R. & ORGEL, L. E. (1978). *Tetrahedron* **34,** 853–857.
MACINO, G., CORUZZI, G., NOBREGA, F. G., LI, M. & TZAGOLOFF, A. (1979). *Proc. natn. Acad. Sci. U.S.A.* **76,** 3784–3785.
MANDEL, M. (1970). *Handbook of Biochemistry,* 2nd Edn. (Sober, H. A., ed.) pp. H75–H78. Cleveland, OH: The Chemical Rubber Co.
MULLINS, D. W., LACEY, J. C. & HEARN, R. A. (1973). *Nature, Lond. New Biol.* **242,** 80–82.
OPARIN, A. I. (1924). *Proiskhozhdanie Zhizny.* Rabochi, Moscow: Izd. Mosk.
ORGEL, L. E. (1968). *J. molec. Biol.* **38,** 381–388.
ORGEL, L. E. (1986). *J. theor. Biol.* **123,** 127–149.
ORGEL, L. E. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52,** 9–16.
ORGEL, L. E. (1989). *J. molec. Evol.* **29,** 465–474.
PORSCHKE, D. (1978). *Biopolymers* **17,** 315–323.
ROOT-BERNSTEIN, R. S. (1982). *J. theor. Biol.* **94,** 885–894.
ROOT-BERNSTEIN, R. S. (1983). *J. theor. Biol.* **100,** 99–106.
SALEMME, F. R., MILLER, M. D. & JORDAN, S. R. (1977). *Proc. natn. Acad. Sci. U.S.A.* **74,** 2820–2824.
SONNEBORN, T. M. (1965). *Evolving Genes and Proteins* (Bryson, V. & Vogel, H. J., eds) pp. 322–397. New York: Academic Press.
STAVES, M. P., BLOCH, D. P. & LACEY, J. C. (1987). *Z. Naturforsch.* **42c,** 129–133.
TSALLIS, C. & FERREIRA, R. (1986). *Physica* **140A,** 348–360.
TS'O, P. O. P. (1974). *Basic Principles in Nucleic Acid Chemistry,* Vol. 1. New York: Academic Press.
WALD, G. (1957). *Ann. N.Y. Acad. Sci.* **69,** 352–359.

WALD, G. (1964). *Proc. natn. Acad. Sci. U.S.A.* **52,** 595–601.
WEBER, A. L. & LACEY, J. C. (1978). *J. molec. Evol.* **11,** 199–210.
WEINER, A. M. & MAIZELLS, N. (1987). *Proc. natn. Acad. Sci. U.S.A.* **84,** 7383–7387.
WESTHEIMER, F. H. (1986). *Nature, Lond.* **319,** 534–535.
WOESE, C. R., DUGRE, D. H., DUGRE, S. A., KONDO, M. & SAXINGER, W. C. (1966). *Cold Spring Harbor Symp. Quant. Biol.* **31,** 723–736.
WOESE, C. R. (1967). *The Genetic Code*. p. 186. New York: Harper & Row.
WONG, J. T. (1975). *Proc. natn. Acad. Sci. U.S.A.* **72,** 1909–1912.
WONG, J. T. (1976). *Proc. natn. Acad. Sci. U.S.A.* **73,** 2336–2340.
WONG, J. T. (1980). *Proc. natn. Acad. Sci. U.S.A.* **77,** 1083–1086.
WONG, J. T. (1981). *Trends Biochem. Sci.* **6,** 33–35.
WONG, J. T. (1988). *Microb. Sci.* **5,** 174–181.
WONG, J. T. (1991). *Orig. Life Evol. Biosphere.* **21,** 165–176.
ZAUG, A. J. & CECH, T. R. (1986). *Science* **231,** 470–473.
ZIELINSKI, W. S. & ORGEL, L. E. (1987). *Nature, Lond.* **327,** 346–347.
ZIELINSKI, W. S. & ORGEL, L. E. (1989). *J. molec. Evol.* **29,** 281–283.