The origin of the genetic code

amino acids as cofactors in an RNA world

The genetic code, understood as the specific assignment of amino acids to nucleotide triplets, might have preceded the existence of translation. Amino acids became utilized as cofactors by ribozymes in a metabolically complex RNA world. Specific charging ribozymes linked amino acids to corresponding RNA handles, which could basepair with different ribozymes, via an anticodon hairpin, and so deliver the cofactor to the ribozyme. Growing of the 'handle' into a presumptive tRNA was possible while function was retained and modified throughout. A stereochemical relation between some amino acids and cognate anticodons/codons is likely to have been important in the earliest assignments. Recent experimental findings, including selection for ribozymes catalyzing peptide-bond formation and those utilizing an amino acid cofactor, hold promise that scenarios of this major transition can be tested.

As Crick *et al.* aptly remarked in 1976, 'the origin of protein synthesis is a notoriously difficult problem'[1]. Essentially, two approaches can be taken to the problem of the origins of the code[2]. The top-down approach analyzes patterns in the contemporary genetic code and tries to identify chemical and selective forces that are likely to have contributed to the shaping of it (see Fig. 1 for a summary of observed patterns in the genetic code). The bottom-up approach is rooted in biochemistry and molecular biology, and aims at constructing consistent and plausible scenarios for the origin of coding and translation. Ideally, the two approaches should meet somewhere in the middle, like stalagmites and stalactites in a cave forming columns connecting its floor and roof.

It has usually been understood that the origin of the genetic code and that of translation must have gone hand in hand. In my opinion, this assertion has hindered breakthroughs in the field. In recent years, new ideas have been put forward that regard the linking of amino acids to nucleic acids as playing a role, in some form, in the metabolism of organisms in the RNA world, prior to the utilization of these links in protein synthesis (translation)[11–13]. According to this view, evolution of this central molecular mechanism has to be broken down into consecutive steps, where an adaptation at one step could serve as a preadaptation in the next one. The origin of the genetic code and translation has been one of the major transitions in evolution[2,14]: it meant a radically new way of using genetic information[15], and it allowed the division of labour between nucleic acids and proteins serving primarily as genes and enzymes, respectively[2].

The RNA world

I consider those approaches that accept an RNA world[16] with a considerable metabolic complexity. The idea of such a world can be traced back to Woese[17], Crick[18] and Orgel[19]. It was later strengthened considerably by the observation that many coenzymes have nucleotide (-like) parts, and could descend from an earlier metabolic stage where enzymatic RNAs (now called ribozymes) served as macromolecular biocatalysts[20,21].

A growing number of experiments provide support for the RNA world idea. Theoretical interest prompted the suggestion to select *in vitro* for RNA molecules with binding capacity to smaller molecules (including amino acids) or showing enzymatic function[22,23]. Since then, artificial selection for ribozymes has become commonplace. The capabilities of ribozymes is reassuring: a number of ribozymes catalyzing reactions at phosphodiester and carbon bonds have been selected: the rate enhancement, relative to the non-catalyzed reaction, ranges between 10^3–10^{10} (reviewed in Refs 24, 25). Emphatically, there is evidence for ribozymes with 3′ and 5′ aminoacyl transferase activity. Recently, a selected ribozyme was shown to catalyze the formation of an amide bond[26]. Even today, rRNA is active in peptidyl transfer in translation[27]. Thus, elementary reactions important in protein synthesis could have been carried out by ribozymes. The key question is this: why would an evolved ribo-organism bother to venture into the protein world? The generally accepted answer is that proteins provided a greater catalytic versatility than nucleic acids (20 versus 4 building blocks). But evolution is myopic: an event happening now would not be selected for, just because it will turn out advantageous thousands or millions of years later.

Some researchers (e.g. Ref. 28) suggested that the full-blown RNA world in fact never existed: instead, they envisage a very early coevolution of RNA and peptides in a 'ribonucleopeptide' world. Unfortunately, such scenarios fail to provide a convincing explanation of how this could

Eörs Szathmáry
szathmary@colbud.hu

Department of Plant Taxonomy and Ecology, Eötvös University, Budapest and Collegium Budapest, Szentháromság u. 2, H-1014 Budapest, Hungary.

 PII: S0168-9525(99)01730-8

TABLE 1. Some examples of binding and usage of cofactor by selected RNA or DNA aptamers

Cofactor	Nature of interaction	Ref.
ATP	Selective binding	33
FAD, NAD	Selective binding	34
Cyanocobalamin	Selective binding	35
NAD^+ and dephosphorylated coenzyme A	Self-incorporation by group I ribozyme	36
Biotin	Self-alkylation (3×10^6-fold enhancement)	37
ATP	Phosphate transfer (polynucleotide kinase, $\geq 10^5$-fold enhancement)	38
UUU (itself catalytic template)	RNA cleavage	39
pCGG-dansyl (trideoxynucleotide)	DNA self-cleavage ($\geq 10^3$-fold enhancement)	40
Histidine	RNA cleavage by DNA ($\approx 10^6$-fold enhancement)	41

have been coupled to the simultaneous solution of the coding problem (including aminoacyl-tRNA synthetases made of protein). Moreover, they miss a strong argument: namely, that short peptides (with less than 30 amino acids) do not readily form defined structures (especially without disulphide bonds), by sharp contrast with polynucleotides, which can be stable and enzymatic with such a limited number of monomers[29]. Also, the lesson from *in vitro* ribozyme selection experiments is that the evolutionary appearance of catalytic RNA is fast and straightforward[24,25,30]. Because evolution is myopic and opportunistic, it is hard to see how an RNA world could have been skipped, given that every other solution seems to be considerably more complicated and time-consuming.

Another important point concerns the coenzymes used in the RNA world. In contemporary organisms a given coenzyme is used by many different enzymes. Evolutionary conservation of the chemical structure of the former must, therefore, be the rule[20,21], because otherwise simultaneous changes in the coenzyme binding sites of all those enzymes would be required. Many contemporary coenzymes possess nucleotide-like moieties, so they might have descended from the RNA world. This point is supported by the observation that, in some cases, the nucleotide part is not involved in the catalytic process[31]. In ribo-organisms, nucleotide-like parts of coenzymes could have been utilized as 'handles' (perhaps carrying other small molecules), whereby ribozymes could grab and manipulate them easily[32].

Experiments demonstrate that RNA can bind coenzymes selectively, and ribozymes can utilize cofactors in their activity (Table 1). Other ways of increasing the catalytic versatility of nucleic acids, including the extension of the genetic alphabet and post-synthetic covalent base modifications, cannot solve the problem in general[42]. I conclude that a ribozyme-run early metabolism, complemented by cofactors (many of them with nucleotide-like parts), is a plausible context in which to seek the origin of coding.

Amino acids as cofactors in the RNA world

With their delivery of novel functional groups, amino acids must have been among these attached molecules. One idea, the hypothesis of coding coenzyme handles (CCH)[2,12,14,42–44], asserts that linking of amino acids to specific oligonucleotide sequences allowed ribozymes to associate with specific amino acid cofactors by binding reversibly through conventional basepairing to the oligonucleotide handle. The specificity of unambiguous linking of amino acids to handles constituted the first useful assignment of an amino acid to a specific sequence; furthermore, the ribozymes catalyzing this charging reaction were the first assignment catalysts. Later in evolution, handles turned into adaptors (tRNA), assignment ribozymes were replaced by protein aminoacyl-tRNA synthetases, and many ribozymes became mRNA molecules while losing their original enzymatic activity. Thus, assignment is a preadaptation for coding and preceded translation (encoded protein synthesis).

Why should the use of amino acids as cofactors have led to their specific attachment to oligonucleotide handles? It is more straightforward for an RNA molecule to bind and recognize another molecule by conventional basepairing than by any other means, as in all components of the splicing apparatus. If amino acids, for their coenzymatic role, were coupled to oligonucleotide handles by specific charging ribozymes, then ribozymes busy with intermediate metabolism would have been able to grab them by the handles. Thus, the amino acid part was taken care of by a few highly specific ribozymes acting as handle-charging enzymes. All other ribozymes did not have to deal with this recognition problem at all, provided the handle-amino acid assignment followed the 'degeneracy combined with unambiguity' principle of the present genetic code[12,44]. This means that one type of amino acid could be linked to several different handles, but not the other way round. In the opposite case a given ribozyme would become associated with a variety of amino acid cofactors and catalytic activity would have been compromised. The idea that in an RNA-based metabolism coenzymes were conveniently equipped with

BOX 1. The testimony of tRNA modifications

Di Giulio has called attention to evidence in favour of the idea that anticodon-bearing ancestors of present-day tRNA molecules did take part in metabolic reactions[49] (as would be expected on the basis of evolution being opportunistic). In the tRNA for Ile the first position of the anticodon is modified by a lysine. This is essential for recognition of the codon AUA. Without this modification, the tRNA is charged with Met. It is noteworthy that all three amino acids mentioned belong to the Asp biosynthetic family. Another modification, that of position 37 (next to the anticodon), of tRNAs recognizing codons that start with A, and code for Ile, Met, Thr, Asn, Lys, Ser and Arg, involves Thr. Gly and a non-proteinogenic amino acid are also involved in some tRNA modifications. According to Di Giulio, these examples hint at 'a possible coevolution of anticodons and amino acids' (Ref. 49, p. 194). Wong suggested that originally amino acids were N-linked to nucleotides because the contemporary O-ester link is too labile[13]. A corollary to the above thesis is that some tRNA modifications might be faint 'palimpsests' of this ancient N-charging of handles by amino acids. For example, sulphur in the modified nucleoside (position 37) originates from cysteine in the codon group containing the Cys codon. In the same group (also containing serine), *cis*-hydroxylation of the isopentenyl-group also occurs (also at position 37), and the arising $-CH_2OH$ group is formally identical to the specific sidechain of serine. It has been shown that the relevant modifying enzymes are crucially sensitive to the identity of base 36 (third anticodon base) in tRNA. These modifications seldom determine tRNA identity in aminoacylation (see Ref. 50 for data), which rules out such a contemporary functional role, and renders a 'molecular fossil' interpretation more likely (compare with Ref. 65).

One can accept Di Giulio's view: 'the hypothesis … here suggests that there was a global coevolution which involved (i) RNA hairpin structures, which were the precursors of tRNAs …;' (ii) the biosynthetic pathways of amino acids occurring on these hairpin structures; (iii) a selection by the evolving anticodons on the emerging product amino acids… (Ref. 49, p. 194).

nucleotide-like handles has, of course, been around for some time[20,32]. The recognition that this idea can be linked to the origin of coding for amino acids is the subject matter of the present paper. An evolutionary bifurcation is envisaged: some coenzymes (like NAD, NADP, FMN, FAD, etc.) were equipped with handles of a non-standard nature (containing, for example, dinucleotides with 5′–5′ links), whereas the amino acids happened to be linked to standard type oligonucleotides[12]. The reason for this was, presumably, that in the latter case a regular family of compounds (amino acids) became linked to a regular family of oligonucleotide sequences (handles), aided by a coevolution of amino acids and their cognate handles. (This implies canonical machinery of amino acid synthesis and charging.) In contrast, the non-nucleotide parts of the other coenzymes are idiosyncratic and do not belong to a regular class of biochemical compounds.

A recent experiment by Roth and Breaker is of outstanding importance for the 'amino acids as coenzymes' idea: they have selected a DNA enzyme (deoxyribozyme) that uses a histidine cofactor for an RNA cleavage reaction[41]. What is now left to do is the repetition of this experiment with a handle-linked histidine. The clear prediction is that such an experiment must be successful, based on the evidence presented in Table 1.

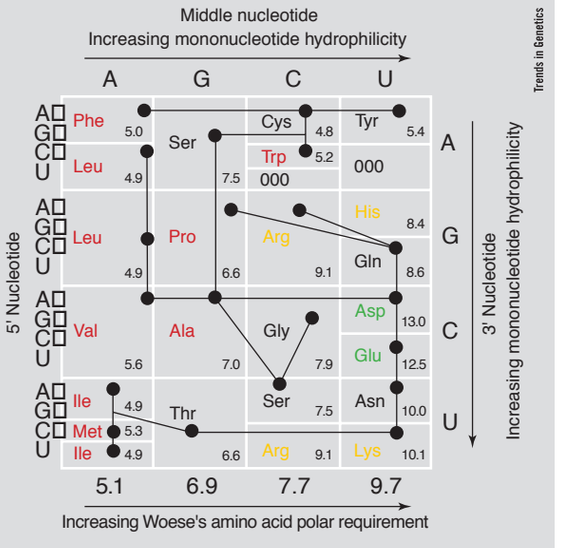
The nature of handles (adaptors) at the coenzyme stage

Having accepted that amino acids might have been equipped with oligonucleotide handles, it is time to enquire about the structure of the latter in detail. Contrary to my first suggestion[12], it seems that handles could not have been shorter than trinucleotides, simply because there would have been no measurable binding between ribozymes and mono- or dinucleotides. Later, I asserted that handles were composed of anticodonic trinucleotides[43], which is more plausible. Kazakov and Altman have shown that a trinucleotide can catalyze a metal-ion dependent cleavage of a specific RNA sequence motif, GAAA (Ref. 39). On the contrary, when 3′-end single strands of tRNA acceptor arms are provided in a solution with pieces of the complementary 5′-end strand of varying lengths, it is found that at least tetranucleotides are required to achieve successful aminoacylation, catalyzed by these oligonucleotide cofactors and the cognate protein synthetase[45]. In effect, the two complementary strands form a so-called microhelix (i.e. just the tRNA acceptor arm), known to be sufficient for aminoacylation in several cases (cf. Ref. 58). Thus, it is equivocal that mere trinucleotides would have generally been sufficient.

Data on oligonucleotide interactions[46] suggest that a more satisfactory solution can be found. Complementary free trinucleotides in solution do not associate measurably. Interaction between a free codon and its complementary anticodonic stem–loop is a thousand times stronger. tRNAs with complementary anticodons interact with each other still a thousand times more strongly. This is mostly attributable to base-stacking interactions, because binding between two anticodon-containing, linear fragments that lack a self-paired stem but have flanking nucleotides, is only ten times weaker[46].

What is the comparable situation in the case of handles bound by ribozymes? The handle-binding site (having a complementary sequence to the 'anticodon' site) is expected to be similar to a tRNA loop: it has a fixed three-dimensional structure, presumably allowing for base stacking. A simple trinucleotide would be expected, on average, to bind to such a site with a constant of equilibrium of about 10^{-3} M, which is at the lower end of the range of binding strengths between protein enzyme active sites and their substrates, the upper end of which lies at 10^{-8} M (Ref. 47); this is approached by tRNAs with complementary anticodons (10^{-6}–10^{-7} M). The lesson is that weaker binding can always be arranged for between two loops, but binding strength between a loop and a free trinucleotide has a fixed, and low, maximum value. The hairpin structures of handles and the corresponding binding sites on the ribozymes would also have solved the problem of the avoidance of undesirable cofactors: because handle-binding sites would have been adaptations themselves, mere trinucleotides just anywhere in a ribozyme would not have interfered.

FIGURE 1. Patterns in the genetic code

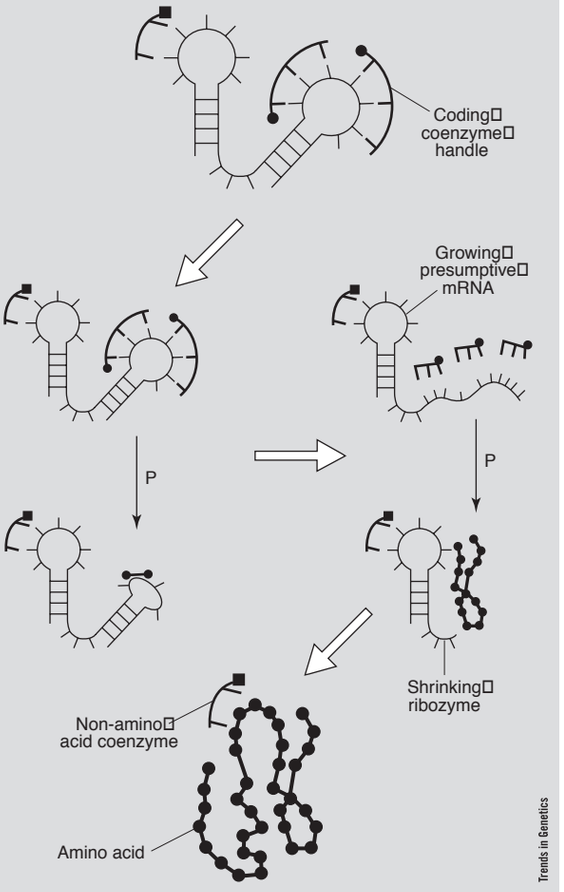


A comprehensive summary of several patterns discerned so far in the genetic code. The basic representation is due to Jungck, showing anticodon rather than codon nucleotides[3]. The numbers next to the amino acids indicate Woese's polar requirement (see Ref. 3 for a discussion of such values). The shading of amino acids corresponds to neutral and hydrophobic (red), basic (yellow), acidic (green), and neutral and polar (black), following Ref. 4. Amino acids adjacent in biosynthetic pathways[5–7] are connected by a bold line. Note that each line between two adjacent dots corresponds to a single nucleotide change in the code. It is apparent that neighbouring amino acids tend to be related by polarity value[8,9], biosynthetic kinship[6,7], or both. In some parts, these two effects do not correlate (on the pathway from Asp to Ile, the change in amino acid polarity is enormous). In other cases, however, there is a correlation (polarity in the Phe, Leu, Cys, Trp and Tyr shikimate family is consistently low, except for Ser at its root). 'The code within the codon' statistical rule[7] can also be observed as a tendency (3′- and middle anticodon nucleotides indicating biosynthetic kinship and amino acid polarity, respectively). It is hard to escape the conclusion that both effects have contributed to shaping the code, sometimes independently. This could explain that some amino acids are 'out of place' regarding one or the other effect. For example, Ser with the NGA anticodon is fine in its biosynthetic family, but its polarity is too high for that region. Or, Arg with anticodon YCU is at the 'right' position regarding its polarity, whereas Arg with anticodon NCG is consistent with biosynthetic contiguity. The tendency that polar amino acids have polar anticodons is also apparent. In fact, Jungck found the following significant (0.92) relationship between anticodon dinucleotides (middle and 3′-base) and amino acid values:

Relative hydrophilicity of dinucleotide =
0.037 polarity requirement − 0.0074 bulkiness + 0.0320

(See Refs 3 and 10 for further correlations.) (Bulkiness is the ratio of side chain volume to length.) The fact that a physicochemical measure and a size measure both appear in the above equation is consistent with the idea that some kind of stereochemical effect was also influential during early evolution of the code.

FIGURE 2. Ribozyme replacement and coenzyme takeover



This drawing is a variation of a theme pursued before[2,21,28]. White and black arrows indicate evolutionary transitions and intracellular transformations, respectively. P, peptidyl transferase ribozyme. First, some catalytic dipeptides form by virtue of P, then the former might associate to mutant, somewhat smaller ribozymes. Evolution drives the system to increasingly divergent sub-populations of presumptive mRNAs and shrinking ribozyme cores. A memory of this process might be reflected by the existence of some basic metabolic protein enzymes that bind nucleotide-like cofactors as well as to their cognate mRNA molecules[64]. Note the different fate of amino acids and other type cofactors: while the former are incorporated into proteins, the latter are taken over from the original ribozymes.

It is safe to conclude that handles generally must have been made of stem–loop (hairpin) structures. Note that another argument in favour of such structures, as opposed to longer single-stranded oligonucleotides, is the greater resistance against spontaneous degradation of the former[48], which is important for a catalytic agent. The fact that specificity of the interaction between coenzyme handles and ribozymes was ensured by Watson–Crick pairing of three (rather than fewer or more) nucleotides must have followed from the physicochemistry of pairing, providing an optimal residence time. Thus, I envisage primordial handles consisting of an ancestral anticodon stem–loop (hairpin) with a linked amino acid.

Evolution of assignment: the genetic code as frozen stereochemistry

According to the present version of the CCH hypothesis, present-day adaptors evolved from RNA hairpins, quite closely matching the anticodon loop of the former. This idea echoes that of Crick *et al.*[1], and is consistent with Di Giulio's finding that amino acids are involved in tRNA modifications in anticodon loops only (Box 1). An important question is what role stereochemical interactions played in the assignment of amino acids to such anticodonic hairpins.

At least three parties must have been involved: hairpins, amino acids and synthetases. The 'weak' form of the stereochemical hypothesis holds that although charging must have involved specific binding of amino acids to RNA sites, this is no longer reflected in the structure of the contemporary genetic code. The 'strong' form in contrast says that traces of such an interaction should be found in the present genetic code (cf. Ref. 51).

Let us first examine the strong form. Although the question is not settled, there is various evidence to suggest that anticodons (or codons) could have played a role in interaction specificity with the amino acid. For example, there is a physicochemical correlation between amino acids and cognate anticodons (Fig. 1). Moreover, a complex of four nucleotides: the discriminator base (upstream of the CCA end of tRNA) and the anticodon ('C4N model') was shown by model-building to interact through a lock-and-key relation with the cognate amino acid[52]. Other stereochemical models suggest a fit between codons and amino acids[53].

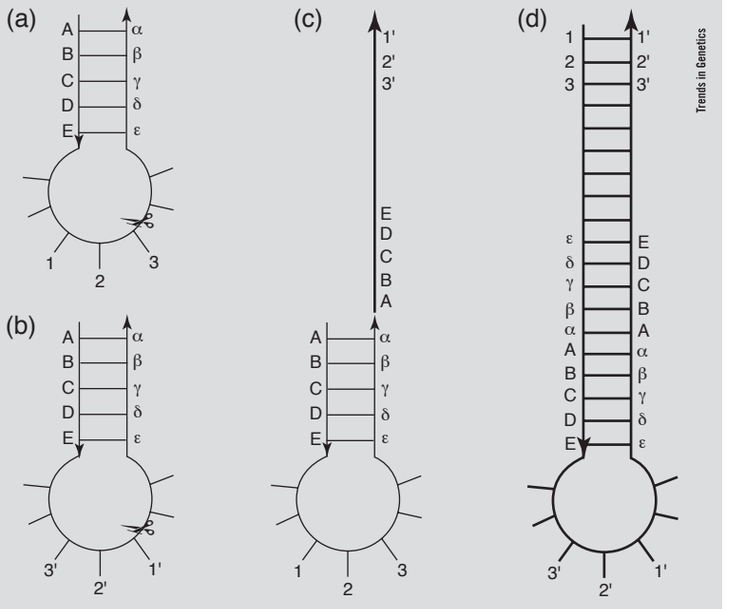
Direct, albeit limited, evidence for the C4N model has recently been obtained. RNAs, consisting of a single hairpin, with a 5′-flanking free anticodon and a conventional discriminator base-CCA sequence at the 3′-end, can be specifically, though modestly, charged with the cognate amino acids (glycine, alanine and valine) in the presence of a valyl-aspartic or alanyl-aspartic acid catalyst and the corresponding aminoacyl adenylates[54]. This experiment indicates that the suggested interactions might in fact occur in solution, but raises further questions about the degree of generality, specificity and ionic strength of the medium (people argued repeatedly that similar interactions were either too weak or occurred under artificial conditions only). We do not yet know whether this result could be generalized to other amino acids, or whether the dipeptide catalyst could be replaced by an oligonucleotide, as would be required in an RNA world.

The facts discussed above indicate a possible stereochemical importance of the anticodon in primordial handle charging. An exception to the anticodonic recognition of amino acids is presented by the binding of arginine to essentially codonic triplets, as it occurs in variants of the *Tetrahymena* group I self-splicing intron[55]. Yarus gives an account of amino acid binding sites of the natural and artificially selected RNAs (Ref. 51). Binding sites for arginine and isoleucine contain cognate codons, whereas no codonic sequence is found in an RNA binding valine. Additionally, five of six of the modern arginine codons have been observed in these RNAs. On this ground, Yarus favours a codonic version of the strong form of direct templating. Yet closer inspection of selected RNA aptamers indicates that his conclusion could have been drawn too readily. In fact, in almost all cases (with one exception) when one or more codonic triplets are observed in a specific binding site, one or more anticodonic triplets are also present. The exception is compensated for by the observation that, whereas a selected binding site to valine contains no codonic triplet, it does harbour two anticodonic triplets!

All in all, the jury is still out on the question whether anticodons or codons played a widespread role in primordial amino acid charging of handles. The available evidence seems to point to a possible importance of both, in agreement with hypotheses resting on the building of stereochemical models[52,53]. However, there are some other points of importance. First, codons and anticodons are so short that they are easily formed by chance, even in relatively short sequences[51]. This is certainly true, but there are two kinds of statistical result that can, in principle, corroborate the stereochemical theory. If one repeatedly finds that codonic or anticodonic triplets are in excess in the binding sites of different, independently selected RNA aptamers for the same amino acid, then this strongly suggests a stereochemical importance for the given triplets. Another kind of favourable finding would be if cognate triplets tended to occur in selected binding sites in aptamers for different amino acids. Even if the *a priori* probability for a single favourable case is quite high, the joint probability becomes remorselessly smaller as the number of independent, positive results increases. The first kind of result has been obtained by Knight and Landweber: whereas the expected, and observed, frequency of arginine codons in nonbinding RNA sites is 30%, they make up 72% of the sequences in the specific binding regions[56]. The results surveyed in the previous paragraph point in the direction of the second kind of finding, which is not to deny that some specific binding sites contain neither codonic nor anticodonic triplets[51]. A possible resolution is offered by Yarus: 'RNA is unexpectedly versatile, such that the code may be a frozen stereochemical accident. In other words, the code may preserve a set of biochemical interactions, but the choice of particular interactions now appears so broad that many other codes could have resulted' (Ref. 51, p. 114). This frozen (bottom-up) stereochemistry would, in part, explain the frozen correlations (Fig. 1) of the top-down approach. Clearly, more systematic screening of possible RNA binding sites for all amino acids should be carried out, as I suggested in 1989 (Ref. 22). To this I might add that one should try to select for ribozymes that could charge amino acids to an anticodon loop. Group I introns were conjectured to be descendants of primordial charging enzymes[57,43], based on the facts that they frequently occur next to (downstream of) the anticodon in eubacterial tRNA genes[58] and that they can naturally bind arginine[55]. In the tRNA genes the anticodon in one exon forms a pair with a formal 'codon' in the intron[58]. Thus, selection for various synthetases, starting from a group I intron sequence, might be rewarding, especially because the system itself could choose whether to use codonic or anticodonic triplets for recognition[43].

Successful selection for synthetase ribozymes might resolve the apparent paradox of anticodonic correlations and stereochemical fits occurring in space-filling models and the weakness of such interactions measured in realistic systems. The anticodonic/codonic relations might be amplified sufficiently only in a macromolecular context, and not in free solution[43]. Although the CCH hypothesis is compatible with direct involvement of codons and/or anticodons in primordial amino acid recognition, it allows any mixture between complete involvement of these triplets and no such involvement at all[43]. My guess is that, for some amino acids, selection will result in anticodonic (codonic) binding sites, and for some others, not. Assignments using the former could be primitive, those using the latter could be evolved (cf. Ref. 49).

FIGURE 3. Evolution of tRNA



Evolution of tRNA from an anticodon hairpin to a longer structure with an operational code at the 'acceptor' end. (a) A hypothetical ancestral anticodon hairpin. (b) A complementary anticodon hairpin. The two hairpins are replicational copies (+ and – strands) of each other[61]. The scissors indicate the common location of Group I tRNA introns; anticodon nucleotides are labelled by numerals. Corresponding letters of the Roman and Greek alphabets, as well as simple and dashed numerals, indicate complementary bases. Arrowheads point in the 5′→3′ direction. (c) This molecule is a fortuitous ligated product of a hairpin (a) and the 5′-part of a hairpin (b). (d) A replicational copy of molecule (c). Newly synthesized parts are indicated by thick lines. At the end of the process one has an extended version of adaptor (b), with the anticodon repeated at the 3′-end of the 'acceptor' stem, which is similar to one of the patterns observed in tRNA sequences[60]. See Box 2 for further details.

Towards translation

Translation is easier to evolve, logically as well as chemically, if there is already a triplet–amino acid assignment present. I imagine a process in which ribozymes began to bind handle-mounted amino acids to their adjacent triplets (Fig. 2). Thus, an evolutionary bifurcation took

BOX 2. Evolutionary origin of the operational code

The operational code is the set of rules through which present-day synthetases recognize the cognate tRNA molecule[59], mostly by the end of the tRNA acceptor stem. Because anticodon recognition by synthetases is occasional, idiosyncratic and is performed by a highly variable domain, it is usually assumed that the acceptor stem of tRNAs evolutionarily precedes the anticodon hairpin[59,60]. I suggest the opposite – that the anticodon hairpin is the primordial adaptor (Fig. 3). Whatever direction one chooses, the problem remains that the genetic code and the operational code must be evolutionarily linked. Data indicate that for tRNA consensus sequences having complementary anticodon hairpins, the bases at position 2 (and slightly less at position 3) in the acceptor stem are also complementary[61]. From this, Rodin and Ohno were able to deduce an evolutionary relationship between the anticodon and the triplet 1–3 (sometimes 70–72) in the acceptor stem[60,61]. Figure 3 shows how an ancestral anticodon hairpin can be extended into a longer one carrying a repeat of the anticodon at the 3′-end of the molecule. (This was possible only with the emergence of adaptors that bound the amino acid through more reactive ester links.) A simple exercise to the reader is to show how one can get an anticodon repeat at the 5′-end of an extended anticodon hairpin. Ultimately, both patterns observed by Rodin and Ohno are deducible from an 'anticodon first' scenario.

place: some copies of the same RNA served as shrinking RNA cores of evolving ribonucleoprotein enzymes, whereas other copies evolved into messenger RNAs that ultimately gave rise to protein genes[43]. This has called for a relocation of the charged amino from the anticodon loop to the 3′-end, concomitant with the accidental (but easy) emergence of adaptors that were longer and had a formal anticodon–codon pair at the end of the acceptor stem (Fig. 3, Box 2). Note that the latter structure echoes the same complex arising transiently in the ribozyme-catalyzed charging of anticodonic adaptors (outlined above). The ribozyme that catalyzed the peptidyl transfer between these novel adaptor molecules (similar to a ribozyme selected artificially[62]) became the large-subunit rRNA in the ribosome[43]. By the regular detachment of amino acids from their adaptors, the latter became reusable in protein synthesis[43]. Selection for peptidyl transfer took off when the advantageous juxtaposition of amino acid-bearing handles became frequent enough. Emerging peptides were used either in solution or in association with shrinking ribozymic cores (Fig. 2). Ultimately, protein enzymes also replaced the charging ribozymes.

Support for the primitive ancestry of tRNAs and the secondary nature of the cognate synthetases was presented very recently. In some archebacteria and eubacteria, the conventional lysyl-tRNA synthetase, belonging to class II synthetases, is replaced by a class I synthetase enzyme. Sequence analyses indicate that at least one of the two types of lysyl-tRNA synthetases appeared after the establishment of $tRNA^{Lys}$, because sequences of the latter belong to a single tree. This suggests that evolutionary tinkering with protein synthetases happened around a tRNA population that was already present (see Ref. 63 for a recent review).

Acknowledgements

I thank Viktor Müller, Csaba Pál and one anonymous referee for helpful suggestions in preparing the manuscript. This work has been supported by grants from the Hungarian Scientific Research Fund (OTKA) and the Ministry of Education.

STOP PRESS • STOP PRESS • STOP PRESS

Statistical evidence in favour of the various, not mutually exclusive, hypotheses for the origin of the genetic code will surely accumulate further. Two recent examples[1,2], both based on tRNA sequences, give support to the co-evolution theory and strengthen the view that, based on their primitive ancestry, tRNAs are a goldmine for researchers interested in the origin of the genetic code. In further agreement with Wong's theory, more and more cases of tRNA-linked amino acid transformations come to light. Two alternative pathways for the charging with asparagine and glutamine of the respective tRNAs are known: the direct and the indirect pathways. The schemes are illustrated with the example of asparagine below.

Direct

$$\text{Asparagine} + tRNA^{Asn} \xrightarrow{\text{Asn-tRNA-synthetase}} \text{Asn-}tRNA^{Asn}$$

Indirect

$$\text{Aspartic acid} + tRNA^{Asn} \xrightarrow{\text{Asp-tRNA-synthetase}} \text{Asp-}tRNA^{Asn}$$

$$\text{Asp-}tRNA^{Asn} \xrightarrow{\text{Transamidase}} \text{Asn-}tRNA^{Asn}$$

According to Wong's co-evolution theory, the indirect pathway historically precedes the direct one, because in the indirect pathway a precursor amino acid is first mis-charged to the tRNA of the product amino acid by the synthetase of the precursor amino acid. It is hard to see how the indirect mechanism could evolve into the direct one without an intermediate stage where both mechanisms co-occur. Sequence data indicate that in *Deinococcus* both mechanisms are present for asparagine as well as glutamine[3]. In a related *Thermus* species, the same holds for asparagine[4]. (The presence and identity of a tRNA-independent asparagine synthetase in these organisms needs further clarification.) The synthetases of the direct route are thought to be eukaryotic inventions, acquired through horizontal gene transfer by certain bacteria. The presence of both pathways in *Thermus* and *Deinococcus* might then be analogous, although not homologous, to a similar character state in a common ancestor or present-day eukaryotes.

1 Chaley, M.B. *et al.* (1999) Relationships among isoacceptor tRNAs seem to support the coevolution theory of the origin of the genetic code. *J. Mol. Evol.* 48, 168–177
2 Bermúdez, C.I. *et al.* (1999) Characterization and comparison of *Escherichia coli* transfer RNAs by graph theory based on secondary structure. *J. Theor. Biol.* 197, 193–205
3 Curnow, A.W. *et al.* (1998) Glutamyl-$tRNA^{Gln}$ amidotransferase in *Deinococcus radiodurans* may be confined to asparagine biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12838–12843
4 Becker, H.D. and Kern, D. (1998) *Thermus thermophilus:* a link in evolution of the tRNA-dependent amino acid amidation pathways. *Proc. Natl. Acad. Sci. U. S. A.* 95, 12832–12837

References

1 Crick, F.H.C. *et al.* (1976) A speculation on the origin of protein synthesis. *Orig. Life* 7, 389–397
2 Maynard Smith, J. and Szathmáry, E. (1995) *The Major Transitions in Evolution,* Freeman
3 Jungck, J.R. (1978) The genetic code as a periodic table. *J. Mol. Evol.* 11, 211–224
4 Lewin, B. (1997) *Genes VI,* Oxford University Press
5 Dillon, L.S. (1973) The origins of the genetic code. *Bot. Rev.* 39, 301–345
6 Wong, J.T.F. (1975) A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. U. S. A.* 72, 1909–1912
7 Taylor, F.J.R. and Coates, D. (1989) The code within the codons. *BioSystems* 22, 177–187
8 Sonneborn, T.M. (1965) Degeneracy of the genetic code: extent, nature, and genetic implications, in *Evolving Genes and Proteins* (Bryson, V. and Vogel, H.J., eds), pp. 377–397, Academic Press
9 Woese, C.R. (1965) On the evolution of the genetic code. *Proc. Natl. Acad. Sci. U. S. A.* 54, 1546–1552
10 Weber, A.L. and Lacey, J.C., Jr (1978) Genetic code correlations: Amino acids and their anticodon nucleotides. *J. Mol. Evol.* 11, 199–210
11 Gibson T.J. and Lamond, A.I. (1990) Metabolic complexity in the RNA world and implications for the origin of protein synthesis. *J. Mol. Evol.* 31, 7–15
12 Szathmáry, E. (1990) Useful coding before translation: the coding coenzymes handle hypothesis for the origin of the genetic code, in *Evolution: from Cosmogenesis to Biogenesis* (Lukács, B. *et al.*, eds), pp. 77–83, KFKI-1990-50/C (preprint)
13 Wong, J.T.F. (1991) Origin of genetically encoded protein synthesis: a model based on selection for RNA peptidation. *Orig. Life Evol. Biosphere* 21, 165–176
14 Szathmáry, E. and Maynard Smith, J. (1995) The major evolutionary transitions. *Nature* 374, 227–232
15 Jablonka, E. and Szathmáry, E. (1995) The evolution of information storage and heredity. *Trends Ecol. Evol.* 10, 206–211
16 Gilbert, W. (1986) The RNA world. *Nature* 319, 818
17 Woese, C.R. (1967) *The Genetic Code,* Harper & Row
18 Crick, F.H.C. (1968) The origin of the genetic code. *J. Mol. Biol.* 38, 367–379
19 Orgel, L.E. (1968) Evolution of the genetic apparatus. *J. Mol. Biol.* 38, 381–393
20 White, H.B. (1976) Coenzymes as fossils of an earlier metabolic stage. *J. Mol. Evol.* 7, 101–104
21 White, H.B. (1982) Evolution of coenzymes and the origin of pyridine nucleotides, in *The Pyridine Nucleotide Coenzymes* (Everse, J., Anderson, B. and You, K.S., eds), pp. 1–17, Academic Press
22 Szathmáry, E. (1989) The emergence, maintenance, and transitions of the earliest evolutionary units. *Oxf. Surv. Evol. Biol.* 6, 169–205

23 Szathmáry, E. (1990) Towards the evolution of ribozymes. *Nature* 344, 115
24 Lorsch, J.R. and Szostak, J.W. (1996) Chance and necessity in the selection of nucleic acid catalysts. *Acc. Chem. Res.* 29, 103–110
25 Pan, T. (1997) Novel variant ribozymes obtained through *in vitro* selection. *Curr. Opin. Chem. Biol.* 1, 17–25
26 Wiegand, T.W. *et al.* (1997) Selection of RNA amide synthases. *Chem. Biol.* 4, 675–683
27 Nitta, I. *et al.* (1998) Reconstitution of peptide bond formation with *Escherichia coli* 23*S* ribosomal RNA domains. *Science* 281, 666–669
28 Di Giulio, M. (1997) On the RNA world: Evidence in favor of an early ribonucleopeptide world. *J. Mol. Evol.* 45, 571–578
29 James, K.D. and Ellington, A.D. (1995) The search for missing links between self-replicating nucleic acids and the RNA world. *Orig. Life Evol. Biopshere* 25, 515–530
30 Hirao, I. and Ellington, A.D. (1995) Re-creating the RNA world. *Curr. Biol.* 5, 1017–1022
31 Benner, S.A. *et al.* (1987) Natural selection, protein engineering, and the last ribo-organism: rational model building in biochemistry. *Cold Spring Harbor Symp. Quant. Biol.* 52, 56–63
32 Orgel, L.E. (1989) The origin of polynucleotide-directed protein synthesis. *J. Mol. Evol.* 29, 465–474
33 Sassanfar, M. and Szostak, J.W. (1993) An RNA motif that binds ATP. *Nature* 364, 550–553
34 Burgstaller, P. and Famulok, M. (1994) Isolation of RNA aptamers for biological cofactors by *in vitro* selection. *Angew. Chem. Int. Ed. Engl.* 33, 1084–1087
35 Lorsch, J.R. and Szostak, J.W. (1994) *In vitro* selection of RNA aptamers specific for cyanocobalamin. *Biochemistry* 33, 973–982
36 Breaker, R.R. and Joyce, G.F. (1995) Self-incorporation of coenzymes by ribozymes. *J. Mol. Evol.* 40, 551–558
37 Wilson, C. and Szostak, J.W. (1995) *In vitro* evolution of a self-alkylating ribozyme. *Nature* 374, 777–782
38 Lorsch, J.R. and Szostak, J.W. (1994) *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* 371, 31–36
39 Kazakov, S. and Altman, S. (1992) A trinucleotide can promote metal ion-dependent specific cleavage of RNA. *Proc. Natl. Acad. Sci. U. S. A.* 89, 7939–7943
40 Burmeister, J. *et al.* (1997) Cofactor-assisted self-cleavage in DNA libraries with a 3′–5′-phosphoramidate bond. *Angew. Chem. Int. Ed. Engl.* 36, 1321–1324
41 Roth, A. and Breaker, R.R. (1998) An amino acid as a cofactor for a catalytic polynucleotide. *Proc. Natl. Acad. Sci. U. S. A.* 95, 6027–6031
42 Szathmáry, E. and Maynard Smith, J. (1997) From replicators to reproducers: the first major transitions leading to life. *J. Theor. Biol.* 187, 555–571
43 Szathmáry, E. (1993) Coding coenzyme handles: A hypothesis for the origin of the genetic code. *Proc. Natl. Acad. Sci. U. S. A.* 90, 9916–9920
44 Szathmáry, E. (1996) Coding coenzyme handles and the origin of the genetic code, in *From Simplicity to Complexity in Chemistry – and Beyond. Part I.* (Müller, A. *et al.*, eds), pp. 33–41, Vieweg
45 Musier-Forsyth, K. *et al.* (1991) Enzymatic aminoacylation of single-stranded RNA with and RNA cofactor. *Proc. Natl. Acad. Sci. U. S. A.* 88, 209–213
46 Grosjean H. *et al.* (1998). Modulatory role of modified nucleotides in RNA loop-loop interactions, in *Modification and Editing of RNA* (Grosjean, H. and Benne, R., eds), pp. 113–133, ASM Press
47 Stryer, L. (1995) *Biochemistry*, Freeman
48 Eigen, M. (1971) Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465–523
49 Di Giulio, M. (1998) Reflections on the origin of the genetic code: a hypothesis. *J. Theor. Biol.* 191, 191–196
50 Björk, G.R. (1995) Biosynthesis and function of modified nucleosides, in *tRNA: Structure, Biosynthesis, and Function* (Söll, D. and RajBhandary, U.L., eds), pp. 165–205, ASM Press
51 Yarus, M. (1998) Amino acids as RNA ligands: A direct-RNA-template theory for the code's origin. *J. Mol. Evol.* 47, 109–117
52 Shimizu, M. (1982) Molecular basis for the genetic code. *J. Mol. Evol.* 18, 297–303
53 Mellersh, A.R. (1993) A model for the prebiotic synthesis of peptides which throws light on the origin of the genetic code and the observed chirality of life. *Orig. Life Evol. Biosphere* 23, 261–274
54 Shimizu, M. (1995) Specific aminoacylation of C4N hairpin RNAs with the cognate aminoacyl-adenylates in the presence of a dipeptide: origin of the genetic code. *J. Biochem.* 117, 23–26
55 Yarus, M. (1988) A specific amino acid binding site composed of RNA. *Science* 240, 1751–1758
56 Knight, R.D. and Landweber, L.F. (1998) Rhyme or reason: RNA-arginine interactions and the genetic code. *Chem. Biol.* 5, 215–220
57 Ho, C.K. (1988) Primitive ancestry of transfer RNA. *Nature* 333, 24
58 Reinhold-Hurek, B. and Shub, D.A. (1992) Self-splicing introns in tRNA genes of widely divergent bacteria. *Nature* 357, 173–176
59 Schimmel, P. *et al.* (1993) An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U. S. A.* 90, 8763–8768
60 Rodin, S. and Ohno, S. (1997) Four primordial modes of tRNA-synthetase recognition, determined by the (G,C) operational code. *Proc. Natl. Acad. Sci. U. S. A.* 94, 5183–5188
61 Rodin, S. *et al.* (1996) The presence of codon-anticodon pairs in the acceptor stem of tRNAs. *Proc. Natl. Acad. Sci. U. S. A.* 93, 4537–4542
62 Zhang, B. and Cech, T.R. (1997) Peptide bond formation by *in vitro* selected ribozymes. *Nature* 390, 96–100
63 Schimmel, P. and Ribas de Pouplana, L. (1999) Genetic code origins: Experiments confirm phylogenetic predictions and may explain a puzzle. *Proc. Natl. Acad. Sci. U. S. A.* 96, 327–328
64 Kyrpides, N.C. and Ouzounis, C.A. (1995) Nucleic acid-binding metabolic enzymes: living fossils of stereochemical interactions? *J. Mol. Evol.* 40, 564–569
65 Cermakian, N. and Cedergren, R. (1998) Modified nucleotides always were: an evolutionary model, in *Modification and Editing of RNA* (Grosjean, H. and Benne, R., eds), pp. 535–541, ASM Press

NF-κB to the rescue

RELs, apoptosis and cellular transformation

The REL/NF-κB/IκB superfamily of signal transducers and transcription factors are paradigmatic of molecular mechanisms by which rapid responses in the immune system can be achieved. NF-κB proteins have been implicated in diverse processes such as the ontogeny of the immune system, immune responses to pathogens and, importantly, in contributions to the multistage processes of oncogenesis, as described in this review. NF-κB and its regulators, the IκBs, are linked to pro- and anti-apoptotic events as well as signaling systems contributing to cellular transformation. How are these disparate events controlled to effect normal and abnormal processes in cells? Here we explore a few of the many events in which NF-κB appears to participate and processes that integrate signals to control important stages of oncogenesis.

The ubiquity of the NF-κB/IκB signaling system is undisputed. A common feature of NF-κB signaling appears to be the use of NF-κB transducers in acute, inducible responses to a variety of extracellular signals. However, even within a single system, this family of proteins can be active in responses that lead to diametrically opposed outcomes for the cell. For instance, in neuronal systems, the β-amyloid protein associated with Alzheimer disease appears to mediate

0168-9525/99/$ – see front matter PII: S0168-9525(99)01719-9