

An alternative to the RNA world hypothesis

Brian R. Francis

Department of Molecular Biology,
University of Wyoming, Laramie, WY,
USA

Abstract

The RNA world hypothesis for the origin of life is widely accepted in spite of the complexity of RNA synthesis. An alternative hypothesis is advanced for the origin and evolution of protein and nucleic acid synthesis. From an early stage synthesis of the evolutionary predecessors of nucleic acids and polypeptides was coupled. This evolved into an RNA replication process that used 3'-aminoacyl-NMP monomers (N= pyrimidine or purine) with a singlet coding system in which the four bases coded for glycine, alanine, valine and aspartic acid. Ribosomal protein synthesis (RPS) evolved from coupled replication by using tRNAs and separating fidelity checking and peptide bond formation functions into small and large ribosomal subunits. Continuity was maintained during the transition to the triplet process by using the same catalysts that aminoacylated NMPs to aminoacylate four different bases at the 3' ends of the original tRNAs. Four GNC codons used the central base to designate the same four amino acids that were coded in singlet replication. Triplet-coded protein synthesis had the advantage of producing multiple copies of protein from a single copy of RNA, and eventually replaced protein synthesis via singlet-coded replication. Evolution of the simple triplet-coded process into RPS is described.

Introduction

The RNA world hypothesis (RWH) for the origin of life has become thoroughly ingrained in the scientific discourse. It is an integral part of the genetics first approach to the problem of the origin of life that compares with the metabolism first approach that advocates a period during which metabolic reactions evolved prior to the appearance of informational polymers.^{1,2} At the center of the RWH is the observation that RNA, an informational polymer, is also in certain instances a biological catalyst. These two properties allow RNA to provide a solution to the *chicken or egg* problem of which came first, proteins or DNA; RNA catalysis by ribozymes occurred prior to the emergence of protein enzymes.³ In the RNA

world there is no need for the complex translational apparatus required for the production of coded proteins. Hence, the central assumption of the RWH is that RNA synthesis preceded encoded protein synthesis.^{4,5}

The distance between prebiotic chemistry and biology, and the problem presented by a claim for RNA as a replicating polymer early in the origin of life have been described by Whitesides: Most chemists believe, as do I, that life emerged spontaneously from mixtures of molecules in the prebiotic Earth. How? I have no idea. Perhaps it was by the spontaneous emergence of *simple* autocatalytic cycles and then by their combination. On the basis of all the chemistry that I know, it seems to me astonishingly improbable. The idea of an RNA world is a good hint, but it is so far removed in its complexity from dilute solutions of mixtures of simple molecules in a hot, reducing ocean under a high pressure of CO₂ that I don't know how to connect the two.⁶ In this paper I extend a previous hypothesis that attempted to bridge the gap between prebiotic chemistry and biology by proposing that RPS evolved from an earlier process in which synthesis of proteins and nucleic acids was coupled.⁷ The coupled process used a singlet coding system, and the main focus of this paper is to describe the transition of the singlet process into the triplet ribosomal process.

RNA has an evolutionary history

The principal difficulty for the RWH is that RNA is an extraordinarily complicated polymer. Biological synthesis of RNA starting from simple molecules such as carbon dioxide, ammonium, and phosphate ions requires a very large number of enzymes. Critics of the RWH point to this complexity and the difficulties for non-enzymatic RNA synthesis from simple starting reagents.^{8,9} A brilliant new pyrimidine nucleotide synthesis needs special reagents and conditions.¹⁰ Proponents of the RNA world recognize the difficulty in first assembling RNA.³ The problem of a simple mechanism for RNA synthesis remains unanswered. In response some markedly different catalytic systems have been proposed for the origin of life including stacking of heterocyclic bases in the absence of a backbone,^{11,12} compositional catalysis by assemblies of lipid molecules,^{13,14} and catalysis by mixtures of random oligomers formed from prebiotically available molecules.¹⁵⁻¹⁷ Proposals that do not include linear replicating polymers have to explain the eventual synthesis of RNA together with the metabolism needed to support it. A number of somewhat simpler RNA-type polymers have been synthesized.^{18,19} In all cases the purines and pyrimidines are retained for base pairing purposes and the polymer backbone is altered. Implicit in these syntheses is the notion that RNA was produced from simpler polymers as

Correspondence: Brian R. Francis
Department of Molecular Biology, University
of Wyoming, Laramie, WY 82071, USA.
Tel: +1.307.460.0408. E-mail: brianrf@uwyo.edu.

Key words: RNA world ribosome evolution.

Acknowledgment: I would like to thank Professor Jordanka Zlatanova for her critical reading of the manuscript and helpful suggestions.

Received for publication: 23 November 2011.

Revision received: 22 February 2011.

Accepted for publication: 23 February 2011.

This work is licensed under a Creative Commons Attribution 3.0 License (by-nc 3.0).

©Copyright B.R. Francis, 2011

Licensee PAGEPress, Italy

Trends in Evolutionary Biology 2011; 3:e2

doi:10.4081/eb.2011.e2

the result of an evolutionary process. Such a concept opens what Orgel calls a Pandora's box, because there is no way of knowing the structures of the evolutionary predecessors, and therefore the origin of life from an RNA world point of view becomes opaque.¹⁸ In addition, the catalytic activity of evolutionary predecessors of RNA would not have been the same as that of RNA (which is highly dependent on 2'-OH groups), but it was the catalytic properties of the predecessors that would have been required for the original synthesis of RNA.²⁰

Evidence for the RNA world hypothesis

Proponents of the RWH look for support to the biochemical activities of ribozymes, which are presumed to be holdovers from the previous RNA era, and an impressive number of non-biological RNA-catalyzed activities generated by *in vitro* selection show the range of reactions that can be catalyzed by RNA.²¹ In spite of their potential as catalysts, only a few biological reactions are catalyzed by ribozymes.²² Most of them involve hydrolysis of RNA or self-excision of introns from RNA transcripts.²³⁻²⁶ In many cases ribozymes are components of ribonucleoprotein (RNP) complexes in which protein components are required for *in vivo* activity.⁴ Surprisingly from an RWH perspective, no RNA component is involved in RNA synthesis, even though a central role is proposed for an RNA polymerase ribozyme in the RNA world,²⁷ and the core of such a polymerase has a similar catalytic site to that of the RNA polymerase enzyme.²⁸ Two particular RNA-catalyzed reactions are considered to be living molecular fossils of the RNA world.²⁹ One is catalysis by ribonuclease P (RNase P), which is best known for cleavage of 5'-leader

sequences from tRNA transcripts, and the other is catalysis of peptide bond formation by the large subunit of the ribosome. A core catalytic RNA sequence in RNase P is conserved, and under high salt conditions the RNA subunit by itself catalyzes the reaction; however, it is inactive *in vivo* in the absence of the protein subunits, which stabilize the superstructure of the RNA. RNase P has one RNA subunit and either one (bacterial), four (archaeal) or multiple (eukaryotic) protein subunits. Protein subunits appear to have been added to the RNA component to broaden its functions.³⁰⁻³⁵

The ribosome is an RNP

The most important biochemical evidence for the antiquity of ribozymes comes from crystal structure determinations and other studies of the large subunit of the ribosome; the evidence confirms that the peptidyl transferase active site is formed from domain V of the 23S rRNA and does not involve ribosomal proteins.³⁶⁻³⁹ Therefore, it is reasoned that, *Because the catalytic element that synthesizes proteins must have preceded proteins, the ribosome and, thus, ribozymes must have preceded protein synthesis.*⁴⁰ The decoding center where mRNA interacts with complementary tRNA anticodons is located on the small subunit and is primarily formed from 16S rRNA.^{39,41,42} The ribosome was regarded as such solid evidence for the antiquity of the RNA world that those of us who were unconvinced were invited to *wake up and smell the coffee*,⁴³ and implored to accept the new RNA world-centered concept of translation.⁴⁴ Is the evidence really that compelling? What about other aspects of ribosomal protein synthesis (RPS)?

Ribosomal proteins are not passive contributors to ribosome function.^{45,46} Assembly of the large and small subunits is dependent upon ribosomal proteins. In the absence of ribosomal proteins 23S rRNA is unable to serve as a peptidyl transferase.⁴⁷ Only domain V of the 23S rRNA is a large independently folding RNA segment,⁴⁸ but RNAs based on domain V have not (yet) been found to support peptidyl transferase activity in the absence of proteins.⁴⁹ Several ribosomal proteins assist in assembly of the large subunit by providing unstructured, highly positively charged polypeptide sequences that bind RNA segments together and extend into the center of the subunit.³⁶ These extensions fold cooperatively with the RNA.⁴⁵ 16S rRNA also folds cooperatively with ribosomal proteins to produce the small subunit.⁵⁰ Mutations of ribosomal proteins markedly affect translation or growth (for details, see the Appendix). Several protein factors are involved in initiation, elongation and termination of translation. Soluble GTPase factors such as the aminoacyl-tRNA loading factor, EFTu, and the translocation factor, EFG, are not required for protein synthesis. The energy

of the aminoacyl-ester bond of aminoacyl-tRNA is sufficient to drive translocation.^{51,52} Nonetheless, GTP hydrolysis by EFG is regarded as a critical step in the selection process for a cognate tRNA in the A site of the ribosome.⁵³ GTPases improve the rate and accuracy of protein synthesis.⁵⁴ The majority of the bridges between the small and large ribosomal subunits involve RNA-RNA interactions but several proteins are also involved.^{55,56} At minimum, proteins are required for correct assembly of the ribosome and maintenance of the rate and accuracy of protein synthesis.⁵⁴ Numerous metal ions, particularly Mg²⁺, fulfill structural roles in the RNA components of the ribosome, including the immediate vicinity of the peptidyl transferase center but not the active site itself.⁵⁷⁻⁵⁹ Overall, the ribosome, like RNase P, is best described as an RNP catalyst that also requires Mg²⁺ for its function.

Evolution of the ribosome

Determination of the initial structure of the large subunit was accompanied by the conclusion that such a complex structure must have evolved from a small RNA domain.³⁷ A dimeric proto-ribosome based on the segment of domain V that binds to the 3'-termini of tRNAs in the A and P sites, has been proposed as a simple precursor of 23S RNA that produced non-coded polypeptides.⁶⁰ Enhancement of the activities of ribozymes has been proposed as the function of the early non-coded polypeptides.⁶¹ Interstrand RNA binding often involves A-minor interactions between adenines of single-stranded RNA and the minor groove of double-stranded RNA helices, particularly those containing G-C base pairs.^{62,63} Addition of RNA modules that bind to domain V via A-minor interactions has been proposed as the means by which the final 23S rRNA was produced.⁶⁴ In this model, ribosomal proteins were added to the structure at the end, allowing the hypothesis to be in accord with the RWH. An alternative explanation is that expansion from domain V into the 23S rRNA was accompanied by changes in ribosomal proteins and that they retained importance in various aspects of ribosomal function. The conclusion that seems to be most in accord with the biochemical evidence is not that RNA synthesis preceded RPS, but that RNPs preceded RPS.

Evolution of the genetic code

Before discussing how ribosomal and other proteins could have been synthesized in the absence of RPS, I will briefly address another related issue, the evolution of the genetic code. Ever since the genetic code was deciphered the questions of the origin and evolution of the code have loomed large.⁶⁵ These problems have been addressed by numerous hypotheses,^{66,67} some of which have accepted the notion that the genetic code originated in an

RNA world, perhaps to assist ribozyme activity.^{68,69} The idea that the full genetic code evolved from a subset of codons coding for a subset of amino acids began with the proposal that an RRY mRNA (R=purine, Y=pyrimidine) could only be read in one frame.⁷⁰ Eigen and Winkler-Oswatitsch proposed that the universal code evolved from one in which glycine, alanine, valine, and aspartic acid were coded by GNY codons (N=any base), with the central base determining which one of the four amino acids was incorporated into a growing polypeptide chain.⁷¹ Proteins comprised of these four amino acids will be referred to as GAVD proteins. GNY codons were proposed as the codons used in the transition from a singlet code to a triplet code with a G at the first anticodon position of tRNAs (position 34) able to recognize a C or U at the third codon position.⁷ In subsequent proposals the GNY code has been refined to a GNC set of codons that would bind to complementary GNC anticodons, allowing two pairs of G-C triple hydrogen bonds to stabilize codon-anticodon binding.^{69,72-74} This set of codons would enable opposite strands of RNA to each code for proteins.⁷⁵⁻⁷⁷

The strengths of codon-anticodon interactions vary, with important contributions coming from G-C triple hydrogen bonds and stabilization of G-C base pairs by A-minor interactions. Among the most stable codon-anticodon interactions are those involving GGC, GCC, and GAC anticodons,⁷⁸ and the GAC anticodon binds more tightly to the GUC codon than the GAA anticodon to the UUC codon.⁷⁹ Therefore, interactions of GNC codons with their corresponding anticodons in the earliest ribosomal process would have been among the most stable.

tRNAs are generally extensively modified, especially at anticodon base 34 and base 37 immediately 3' of the anticodon. These modifications produce changes in the structure of the bases in the anticodon loop that enable the three bases of the anticodon to adopt a structure that improves codon-anticodon pairings in the decoding center of the ribosome. These structural changes also assist in reading frame fidelity.⁸⁰⁻⁸³ However, bases 34 and 37 are generally not modified when the anticodon is GNC,^{84,85} showing that an early set of GNC codons could have been read without these tRNA modifications. An exception is methylation of G37 in tRNA^{Asp}(GUC) that prevents its aminoacylation using arginyl-tRNA synthetase.⁸⁶ This modification would not be needed before arginine became a coded amino acid. Modification of archaeal tRNAs appears to be significantly limited compared to bacterial and eukaryotic tRNAs,⁸⁷ and in the ancient hyperthermophilic archaeum, *Methanopyrus kandleri*, the bases at position 34 are unusually simple, almost all G or modified U. Because of the wobble rules, these two bases at position 34 enable all of the codons in the genetic code to

be read except for a modified C of a CAU anticodon to read the Ile-AUA codon. Another archaeum, *Methanococcus jannaschii*, has almost the same base 34 pattern.⁸⁸ These species may have preserved the ancestral set of GNC anticodons together with anticodons to be used for coding as the genetic code expanded.

Coupled nucleic acid and protein synthesis

If the earliest ribosome had proteins as well as RNA, how were the RNA and proteins produced? By logic, the proteins must have been produced by a non-ribosomal process. My view is that the existence of an RNA world is unlikely and Pandora's box must be opened. Is the evolution of RNA as much of an open-ended question as Orgel suggests? The complexity of RNA demands that its synthesis evolved in several stages from a simple polymer, with the polymer at each stage able to infer an advantage over the previous stage in terms of hydrolytic and thermal stability, leading to RNA (and subsequently to DNA) with optimization of function. Each stage would have required an accompanying expansion in metabolism so that new monomers could be synthesized and activated, and the new monomers would need to use an earlier polymer as a template for synthesis in the same way that an RNA template is used for synthesis of DNA from dNTPs catalyzed by reverse transcriptase. Two possible RNA predecessors, one with a backbone of glycerol-phosphate and another with threose-phosphate, have been shown not to be sequential in evolution because they cannot form a mixed double-stranded helix.⁸⁹

In a previous paper I introduced the hypothesis that the earliest predecessor of nucleic acids was a β -linked linear polyester synthesized from malic acid, which is a simple molecule at the center of modern metabolism. Such a polymer has carboxylic acid side chains capable of interstrand hydrogen bonding. By itself, poly(β -malic acid) is not a genetic polymer because that would require at least two different side chains, but it could be converted into a genetic polymer by modification of the carboxylic acid side chains to produce two different side chains. Modifications that produced two different side chains with carboxamide functional groups would allow interstrand amide pairing. From this beginning a series of evolutionary steps converted the polyester backbone into the 3'-5' ribose phosphodiester backbone of RNA, and the amide-containing side chains into the four bases.⁷

RNA polymerization from NTP monomers using RNA polymerase is driven in the direction of product by the energy generated by formation of a phosphodiester bond at the expense of a phosphoanhydride bond, together with a shift in the equilibrium towards polymerization resulting from hydrolysis of the

other product, pyrophosphate, by pyrophosphatase. The same shift in equilibrium driving linear polyester synthesis could have arisen if the monomers were esterified at the free hydroxyl group with amino acids, and during polymerization the amino acid was not released but incorporated into a growing α -polypeptide on the end of the polyester chain. In this way polymerization of β -polyester and α -polypeptide chains could have been coupled: one copy of polypeptide produced per strand of polyester. The simplest coding system would have been one in which two monomers with different side chains were esterified with glycine and alanine, the two simplest amino acids. Coupled polyester/polypeptide synthesis evolved into coupled synthesis of RNA and proteins from 3'-aminoacylated-NMPs, and the early coding system evolved into one in which the 3'-hydroxyl groups of AMP, GMP, CMP and UMP, were specifically aminoacylated with aspartic acid, glycine, alanine, and valine, respectively, allowing coupled synthesis of RNA and proteins as shown in Figure 1A.⁷ Peptide bond synthesis in this process would resemble that directed by polyU from 2'(3')-glycyl esters of adenosine derivatives.⁹⁰ Judging by the existing editing requirements for mischarged tRNAs (reviewed in reference 91), aminoacylation of NMPs may only have required proofreading to remove Gly from mischarged CMP, which could have been accomplished by an independent editing catalyst.^{92,93}

Four different bases coding for four different amino acids would have been a singlet process. The four amino acids coded by the singlet process were the same as those resulting from translation of GNC codons in the triplet system. The types of protein that would have been possible from the singlet process can only be inferred from the properties that

Gly, Ala, Val, and Asp confer on proteins produced by RPS. All of the proteins would have been acidic. Glycine, alanine, and valine are well-established components of the transmembrane domains of integral membrane proteins, as found for example in the glycoprotein A dimer.⁹⁴ These amino acids are also able to form the core of a globular protein, as exemplified by the conserved core of the C-Ala globular domain.⁹⁵ Main chain C=O bonding to metal ions and main chain N-H hydrogen bonding to anions commonly depend on the conformational flexibility provided by glycine.⁹⁶⁻⁹⁸ Aspartate is found in the active sites of numerous enzymes, including polymerases and nucleases where it acts together with Mg^{2+} in binding to phosphate.^{99,100} In biochemistry, binding of proteins to phosphate commonly involves lysine and arginine side chains but it can also occur primarily through main chain N-H hydrogen bonding.¹⁰¹ Ribosomal proteins are nearly all basic, but binding to RNA through main chain N-H and Asp/ Mg^{2+} may have existed prior to the inclusion of basic amino acids into the genetic code. Therefore, ribosomal proteins coded by the singlet process would have been able to bind to RNA. With a low ratio of protein to RNA in the singlet process, perhaps close to 1:1, it would be almost inevitable that proteins and RNA would frequently interact and function together. In other words, many of the catalytic entities may have been RNPs. RNA would not be expected to traverse primitive membranes, and this capability probably rested with proteins. Some of the catalysts may have been ribozymes acting without a protein component. The singlet process could have established a complex RNP world into which RPS was born. RNPs that included proteins generated by the singlet process will be referred to as singlet RNPs.

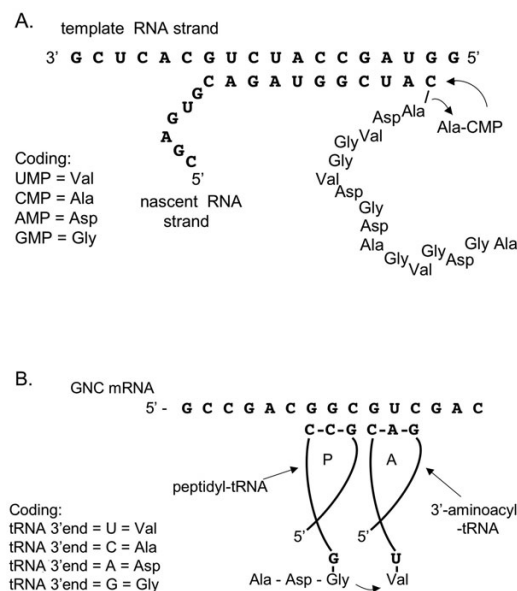


Figure 1. Evolution of ribosomal protein synthesis with GNC codons from coupled synthesis of RNA and GAVD proteins. (A) Coupled RNA and protein synthesis using 3'-aminoacylated NMPs. Phosphodiester bond formation occurs together with peptide bond formation and movement of the polypeptide chain onto the new 3'-terminal base of the nascent strand. (B) Ribosomal protein synthesis using GNC codons. Uncoupling of protein and nucleic acid syntheses used GNC codons for protein synthesis with four 3'-aminoacyl-tRNAs coding for Gly, Ala, Val, and Asp. A segment of mRNA is shown with P and A site tRNAs during elongation. The central base of GNC codons is the same as the 3'-nucleotide of the tRNA.

Although RNAs based on domain V of 23S rRNA have not been found to support peptidyl transferase activity,⁴⁹ other RNAs isolated by *in vitro* selection allow peptide bond formation in the absence of proteins including short RNA sequences.¹⁰²⁻¹⁰⁴ Similarly, RNAs have been isolated that bind to amino acids with some specificity, primarily to amino acids with complex side chains.¹⁰⁵ Binding of amino acids to RNA sequences is thought to resemble binding to codons or anticodons, allowing for a stereochemical explanation for the genetic code. In some instances RNAs act *in vivo* by binding small molecules upstream of the start sites of mRNAs, allowing regulation of protein synthesis.¹⁰⁶ These experimental observations have allowed development of versions of a direct RNA template hypothesis for the emergence of coded protein synthesis in the RNA world via intermediate formation of noncoded polypeptides.^{61,107,108} Binding of amino acids and activated amino acids to sites created by small folded RNA molecules is clearly established and close proximity of such sites is entirely reasonable, but when it comes to the transition to ribosomal protein synthesis these hypotheses become vague.^{107,108} They intersect with the coevolution hypothesis developed here at the point where singlet RNPs were used for activation of amino acids as 3'-aminoacyl-NMPs for singlet replication. RNA components of these RNPs could have had structures similar to those obtained by *in vitro* selection.

Evolution of tRNAs

The last step in this hypothesis is to describe the transition from a singlet to a triplet coding system for protein synthesis, but before doing so, it is necessary to address two other issues: the evolution of tRNAs and aaRSs. tRNA has a structure suggesting that it was originally formed from two half-tRNAs, the top half containing the acceptor stem and TYC stem/loop and the bottom half containing the D and anticodon stem/loops, and that they were subsequently linked together to form full-length tRNAs.^{109,110} Surprisingly, in the otherwise highly reduced genome of *Nanoarchaeum equitans* some tRNAs are indeed coded by two separate RNA segments.¹¹¹ One approach to understanding the evolution of tRNAs has been to suppose that their function in translation was preceded by a different function, and tRNAs and tRNA-like structures have several roles in addition to that in ribosomal protein synthesis.¹¹² In one of them tRNA-like segments are found at the 3' ends of some single-stranded RNA viral genomes with the same terminal 3'CCA sequences as those found in tRNAs.^{113,114} The existence of these and other tRNA-like structures prompted Weiner and Maizels (W-M) to hypothesize that during the RNA world they played a role in initiating ribozyme-catalyzed RNA replication, with the

3' CCA ends acting as templates and primitive telomeres.^{115,116} Consequently, tRNA-like structures first evolved their function in RNA replication, only later to be cleaved by RNase P for use as tRNAs in ribosomal protein synthesis. Although the coupling of protein and RNA synthesis proposed here seems to fit well with the W-M hypothesis, the use of 3' CCA ends is problematic. The implication of the W-M hypothesis is that the 3' CCA terminal residues were evolutionarily conserved from RNA-catalyzed replication to translation, but this is not the process of tRNA synthesis. tRNAs are transcribed as precursors and processed to remove sequences at the 5' end using RNase P and at the 3' end using the endonuclease, RNase Z,¹¹⁷ or endonuclease, RNase E, plus exonucleases.¹¹⁸ All mature tRNAs have CCA at the 3' end, but in archaea, eukaryotes, and many bacteria, tRNA transcripts end at base 73, the discriminator base, and CCA is added post-transcriptionally using CCA adding enzyme [ATP(CTP):tRNA nucleotidyl transferase]. In a hypothetical pathway by which tRNA-like structures evolved into transcribed tRNA sequences it is difficult to understand why CCA ends would be removed and then added back. Moreover, the CCA-adding enzymes come in two classes that code for structurally different proteins showing that the CCA-adding enzyme has evolved twice.¹¹⁹ This suggests that ribosomal protein synthesis has not always used tRNAs with CCA ends, and raises the possibility that the discriminator base, now mostly an A, was a universal A and the site of aminoacylation prior to addition of CCA.

Identity elements are those tRNA bases that are recognized by their cognate aaRSs and provide specificity for tRNA aminoacylation. Primary identity elements in the acceptor stem include the discriminator base and base pairs at positions 4-69, 3-70, 2-71, and 1-72.¹²⁰⁻¹²² The 3' terminal CCA sequence ending with A76 does not act as an identity element, nor in general do post-transcriptionally modified bases. Even since the split of the kingdoms, identity elements of tRNAs have evolved as illustrated particularly by tRNA^{Gly}.¹²³ tRNAs for GAVD amino acids illustrate many general features of the identity elements of tRNAs.¹²¹ Identity elements in tRNA^{Gly} include particularly G1-C72 and C2-G71 in the acceptor stem and anticodon positions 35 and 36, and to a lesser extent G3-C70. Thus, major identity elements are commonly found in the acceptor stem and anticodon. As its name suggests the discriminator base is often an identity element, but unusually the base involved in glycylation of eubacterial tRNA^{Gly} (generally U73) is different from that of archaeal or eukaryotic tRNA^{Gly} (A73).¹²³ G3-U70 is a completely conserved major element in tRNA^{Ala}. Less important are the acceptor stem pairs G1-C72 and G2-C71. Base 73 and the anticodon bases are not ele-

ments for tRNA^{Ala}.^{124,125} tRNA^{Val} has major identity elements at anticodon bases 35 and 36 and a minor element at base 73.¹²⁶ G73 and bases G34, U35, C36, and C38 in the anticodon loop are identity elements for tRNA^{Asp}.¹²⁷

My previous proposal was that the earliest tRNAs had at their 3' ends four different bases that were aminoacylated with four different amino acids in the same way as the NMPs were aminoacylated for the singlet replication process.⁷ The earliest site of aminoacylation may have been position 71, which is still C for Ala and G for Gly. The central base of the four GNC codons was the same as base 71 of the corresponding tRNA (Figure 1B). From the point at which tRNAs were formed from the top and bottom halves, a hypothetical path for tRNA evolution involves the following steps. Initially, tRNAs ended at position 71, and the base at position 71 was the only identity element. The 3-70 and 4-69 base pairs then developed as identity elements for aminoacylation. Bases 72 and 73 were added with base A73 becoming a universal site of aminoacylation. Identity elements for aminoacylation expanded further to include a 2-71 base pair. Lastly, CCA was added to the 3' end of tRNAs and A76 became the site of aminoacylation, allowing base 73 and base pair 1-72 to become identity elements. In order to expand the codons for tRNAs ending in CCA (A76 tRNAs), new identity requirements had to be met to maintain fidelity. This involved recognition of other bases, particularly anticodon bases, as identity elements. Each step in the evolution of tRNAs would have been accompanied by changes in the ribosome that improved translation. For example, the last step, adding the CCA ends, would have been accompanied by changes in the large subunit rRNA in which Gs in the A- and P-loops of domain V provided binding to cytosines in the CCA ends of tRNAs in the peptidyl transferase active site.¹²⁸

Evolution of aminoacyl-tRNA synthetases

aaRSs play critical roles in translation by ensuring that amino acids are bound to their cognate tRNAs. Two structural classes of aaRSs are defined by specific sequence motifs in their catalytic sites.¹²⁹ Lysine is exceptional in having aaRSs in both classes.¹³⁰ Each class has a modular structure, and evolution of aaRSs is explained by fusion of domains to a core catalytic domain for amino acid adenylation and transfer to the 3' terminal A76 of tRNA.¹³¹⁻¹³³ Fused domains recognize identity elements in the acceptor stem, the anticodon, and other tRNA segments, and in some cases provide an editing function for removal of mischarged amino acids. As an example, AlaRS consists of catalytic, tRNA recognition, editing, and oligomerization domains.¹³⁴⁻¹³⁶ The idea that separate domains became fused to form aaRSs

is supported by the existence of separate editing domains in some cases.^{95,137} Sometimes the catalytic domain itself makes significant contacts with the acceptor stem and other tRNA segments.¹³ Sequence and structure comparisons of aaRSs have allowed classes I and II to be subdivided and their evolution to be traced by comparison of aaRS catalytic domain sequences.¹³⁸⁻¹⁴⁰ Gene duplications first established the subclasses of aaRSs and continued until the present members in each class were produced. Anticodon binding domains (ABDs) were added at particular points in the process. After aaRSs gained the ability to recognize anticodon and other tRNA bases, identity elements in the acceptor stem were relaxed and acceptor stem sequences including the discriminator base diverged.¹²¹

Analysis of the sequences and structures of several protein components of the translation process including aaRSs, nucleic acid polymerases, and ribosomal proteins show that substantial diversification of proteins within each of these classes has already occurred before the emergence of universal and currently indispensable components of the translation system.¹⁴¹ For example, class I aaRSs and other enzymes are evolutionarily linked to an antecedent containing a Rossmann fold structure. Supporting the conclusion that RPS was well advanced before establishment of the aaRSs are phylogenetic and experimental evidence that at least one of two LysRSs arose after establishment of tRNA^{Lys},^{142,143} similar identity elements for tRNA^{Cys} recognized by two evolutionarily distinct aaRSs,¹⁴⁴ and lineage differences of Tyr-RS and Gly-RS.¹⁴⁵⁻¹⁴⁷ How did class I aaRSs that were indispensable for protein synthesis evolve with other proteins from the same basic structure? This conundrum has been explained by prior evolution of tRNAs in an RNA world.^{141,142} Hohn et al. conclude that tRNA identities and the universal genetic code were established in the RNA world before the aaRSs.¹⁴⁴

An alternative explanation for the *late* appearance of the aaRSs is that expansion of RPS to encompass the genetic code used tRNAs ending with a universal A73 (A73 tRNAs). At that time the A73 would have been bound into the peptidyl transferase active site of an earlier form of the large ribosomal subunit by A-minor interactions similar to those formed by A76 of A76 tRNAs in the A- and P-sites of the large ribosomal subunit.^{56,62} Structures of aaRSs suggest that binding of CCA ends to catalytic domains is an original property of aaRSs.¹³⁸ Their catalytic domains were not used for aminoacylation of base 73 and then expanded to aminoacylate base 76. Therefore, aaRSs have always used CCA ends, A73 tRNAs were not substrates for aaRSs, and aminoacylation of A73 tRNAs used different catalysts from the aaRSs. These catalysts

would have needed continuity with the earlier process of aminoacylation of base 71, which means that they recognized identity elements primarily in the acceptor stem of A73 tRNAs. For instance, recognition of the identity elements for A73 tRNA^{Ala} would include G2-C71 plus the G3-U70 that is the major identity element of tRNA^{Ala}. Coordination of A73 tRNA binding and aminoacylation would have been easier when the identity elements in the acceptor stem were closer to the active site than they are for A76 tRNAs, as shown by the two additional domains that are needed to recognize G3-U70 of tRNA^{Ala} by AlaRS.¹³⁴ Another possibility is that recognition of A73 tRNAs involved ABDs similar to those found in aaRSs, but as independent subunits that could be later fused to aaRS catalytic domains. The OB-fold used by AspRS for anticodon binding,¹⁴⁸ probably had an early role in translation,¹⁴⁹ but other ABDs have novel RNA binding properties,¹³⁸ suggesting that they evolved for use with aaRS catalytic domains and were not holdovers from aminoacylation of A73 tRNAs. Thus, identity elements for A73 tRNA aminoacylation, principally in the acceptor stem, may have been sufficient for expansion of the code.

Adaptation of aaRSs to A73 tRNAs with added 3'CCA ends presented a continuity problem; the aaRS catalytic domains did not necessarily have binding sites for the amino acids corresponding to those required by the anticodons of the new A76 tRNAs and determined previously by identity elements in the acceptor stems of A73 tRNAs. For continuity purposes, identity elements in the acceptor stem of A76 tRNAs would have had to select for those aaRS catalytic domains that bound the *correct* amino acid. To provide binding to the acceptor stem for recognition of identity elements, the first aaRS catalytic domains may have had small inserts or additions such as the pair of antiparallel α helices that bind to the minor groove of the acceptor stem of tRNA^{Ala}.¹⁵⁰ Having selected for a particular tRNA via identity elements, the aaRS catalytic domains would need to select for the cognate amino acid. Cells in which A76 tRNAs were charged with the wrong amino acids would incorporate wrong amino acids into many proteins, which would be seriously detrimental to growth. In class II aaRSs the amino acid binding site is determined by amino acid side chains,¹³⁸ and the selection process for the correct aminoacylation reaction would involve variation in these side chains. (In passing it should be noted that introduction of the aaRS catalytic domains into an existing RNA world of aminoacylation reactions would present a similar, perhaps even worse, continuity problem.)

Replacement of pre-aaRS aminoacylation catalysts occurred sequentially via gene duplication and mutation of aaRS catalytic domains, and their fusion to other domains for recogni-

tion of new identity elements and editing of mischarged amino acids. Complete changeover to aaRSs and A76 tRNAs may not have been accomplished prior to the separation of eubacteria and eukaryotes.¹⁵¹ During the changeover the peptidyl transferase site of the ribosome would have had to accommodate a mixture of tRNAs ending in either A73 or A76. Addition of Gs in the A- and P-loops of the large rRNA would also occur in a stepwise manner to improve binding to the CCA ends.

In cases where amino acid binding sites of aaRSs bind to amino acids with similar side chains, editing domains have been incorporated to selectively cleave misacylated amino acids from tRNAs.⁹¹ Proofreading to remove amino acids from noncognate tRNAs was likely to have been as necessary for A73 tRNAs as it is for A76 tRNAs. Given the abilities of separately expressed editing domains to deacylate A76 tRNAs,^{152,153} and the existence of separate proteins for trans-editing in some cases,^{92,93,137} similar editing proteins may have served the same purpose for A73 tRNAs. Hence, editing domains in aaRSs may have been carried over from the A73 tRNA period.

Installation of aaRSs and A76 tRNAs must have provided a significant advantage over the A73 *system* that it replaced. One advantage may have resided in the use of pyrophosphatase. Pyrophosphate may have had a prominent role as an energy currency in early cells.¹⁵⁴ When ATP became the principal energy currency the role of pyrophosphate was diminished and pyrophosphatase became a prominent enzyme for driving biosynthetic reactions involving cleavage between the α and β phosphates of NTPs. If this was the case, amino acid activation and transfer to A73 tRNAs may have been catalyzed by members of the ATP grasp superfamily of enzymes that ligate carboxylates with a range of substrates via intermediate formation of acyl phosphates and release of ADP.¹⁵⁵

In summary, evolution of tRNAs in three steps was accompanied by three stages in evolution of tRNA aminacylation: first, singlet RNPs aminoacylated tRNAs at base 71 allowing translation of GNC codons, then triplet-coded proteins, perhaps related to the ATP grasp superfamily, catalyzed aminoacylation of tRNAs at A73 allowing expansion of the genetic code into the universal code, and finally the catalytic domains of aaRSs aminoacylated A76, and aaRSs and A76 tRNAs evolved to replace the A73 tRNA system.

Transition from singlet- to triplet-coded protein synthesis and expansion of the genetic code

The triplet process uncoupled protein and RNA synthesis so that one copy of RNA could code for many copies of protein. The new process would have been so advantageous that

the singlet process for protein synthesis was eventually replaced, and RNA synthesis evolved to use nucleoside triphosphates and pyrophosphatase. Crick did not think that the triplet coding system could have been preceded by a singlet or doublet coding system: *It might be argued that the primitive code was not a triplet but that originally the bases were read one at a time (giving 4 codons), then two at a time (giving 16 codons) and only later evolved to the present triplet code. This seems highly unlikely, since it violates the Principle of Continuity. A change in codon size necessarily makes nonsense of all previous messages and would almost certainly be lethal.*⁶⁵ If all of the RNA sequence was converted from a singlet process to a triplet process this would be correct, but if the triplet process was introduced for segments of RNA while the singlet process was still operating for RNA replication, it would not. A triplet process for protein coding using RNA segments with GNC codons could coexist with a singlet process without violating continuity if the singlet RNP catalysts that were used for aminoacylation of NMPs for replication were also used for aminoacylation of the four bases at position 71 of the earliest tRNAs.

No independent function for the bottom half of the tRNA molecule has been found, leaving the origin of the codon-anticodon interaction as a major problem in biology. What is the evolutionary predecessor of two adjacent seven-membered loops each forming a minihelix with another strand of RNA? My previous hypothesis was that they functioned in opening a bubble in dsRNA and translocation of the bubble along dsRNA.⁷ Crystal structures and other studies of the small ribosomal subunit suggest another possibility, namely that the minihelices evolved from a singlet RNP replication process in which RNA provided fidelity checking for W-C pairing. In the A site of the decoding center of the ribosome, base pairs involving codon base 1 and anticodon base 36, and codon base 2 and anticodon base 35, are monitored by A-minor interactions with 16S rRNA bases for correct W-C base pairing. This is one of the principal ways by which fidelity in translation is maintained. RNA based W-C fidelity monitoring is thought to be older than protein-based fidelity monitoring used by polymerases.⁴² The mechanism for checking correct W-C pairing in a singlet RNP-catalyzed replication process may also have involved an A-minor interaction for each base incorporated into the nascent strand with the monitoring system translocating one base at a time along the template strand. Another, perhaps higher fidelity, mechanism may have involved more than one A-minor interaction to allow some type of back-tracking for excision of misincorporated bases. Singlet RNP replication may even have added three bases at a time before the monitoring process translocated along the

template strand. Replication may have had two RNP components, one for fidelity checking and the other for catalysis of phosphodiester and peptide bond synthesis. Ribosomal translocation evolved from RNP replicase translocation by separating fidelity checking and peptide bond formation in the small and large ribosomal subunits, and losing phosphodiester bond formation. Codon-anticodon interactions in the A- and P-sites mimicked a segment of double stranded RNA binding between the nascent and template strands in the singlet RNP replicase. Addition of each new tRNA anticodon added to the *template* strand (Figure 2).

During the A73 tRNA era the genetic code developed from the four original amino acids into the universal code. To reach that point, gene duplication allowed evolution of new metabolic pathways for synthesis of new amino acids. Codons were not reallocated from one amino acid to another as new amino acids were incorporated into the code. Stop codons were used for amino acid expansion.¹⁵⁶ Use of stop codons for pyrrolysine and selenocysteine has extended the universal genetic code.¹⁵⁷ Duplication and mutation of tRNA genes and modification of tRNA bases generated tRNAs with new anticodons and identity elements that allowed stop codons to be used to introduce new amino acids. Mutations in mRNA genes that produced stop codons permitted new amino acids to be incorporated into existing sequences, allowing selection for improved function. Alterations to the ribosome included adding segments to rRNA genes and new ribosomal proteins. Additions to large subunit rRNA may have displaced ribosomal proteins away from the peptidyl transferase center leav-

ing unstructured peptide sequences for binding of the new rRNA segments. Refinement of the ribosome also permitted fidelity checking after peptide bond formation.¹⁵⁸

Intergenic RNA segments remained after the transition to NTP-based RNA synthesis and triplet-coded translation. The proposal by Eigen and Winkler-Oswatitsch for an early GNY coding system was based on analysis of tRNA sequences.⁷¹ It is possible that tRNA genes were coding sequences that lost their coding capabilities, but it is not necessary for either rRNAs or tRNAs to have coded for proteins in the triplet process. Some RNA sequences that did not code for proteins became introns allowing the length of triplet coded sequences to increase and linking up sequences of rRNAs or tRNAs.^{159,160} Introns are rarely found in archaeal mRNAs,¹⁶¹ but are found in tRNAs and rRNAs.^{159,162-164} The origin of introns has been the subject of much discussion.^{165,166} The introns-first hypothesis is that exons emerged from noncoding regions interspersed between RNA genes (genes coding for ribozymes and other RNAs) in RNA sequences.¹⁶⁶ As developed in this hypothesis, introns were derived from non-coding sequences remaining after the transition to the triplet coding system. The large number of non-coding RNA sequences performing regulatory and guide roles¹⁶⁷⁻¹⁷² would also have evolved from sequences that were not involved in triplet coding.

Concluding remarks

The essence of this hypothesis is that a singlet coding system for coupled protein and nucleic acid synthesis existed prior to the triplet process for ribosomal protein synthesis.

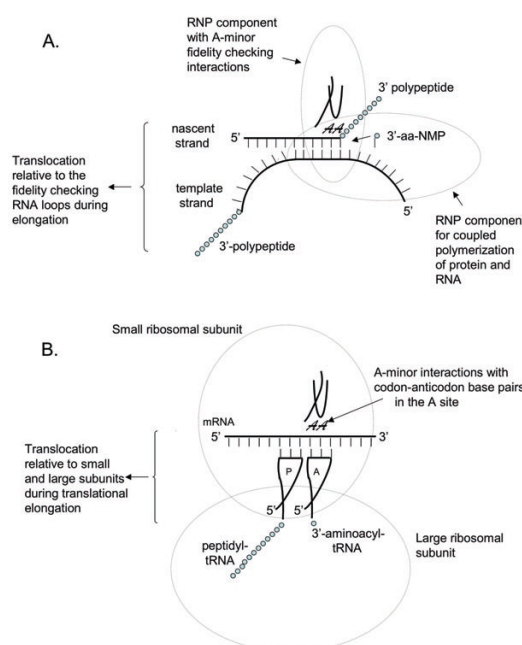


Figure 2. Proposed evolution of codon-anticodon interactions from RNP catalyzed coupled RNA-polypeptide synthesis. (A) RNA replication using 3'-aminoacyl-NMPs and an RNP polymerase that included A-minor interactions (A) with W-C base pairs in the active site for fidelity monitoring. A template-nascent strand duplex with six base pairs is shown. RNP components are involved in fidelity checking and the polymerase activity that catalyzed formation of phosphodiester and peptide bonds. **(B)** Preservation of the fidelity checking process in ribosomal protein synthesis at the A site of the decoding center of the small ribosomal subunit. The RNP polymerase component evolved to use tRNAs for peptide bond formation in the peptidyl transferase active site of the large ribosomal subunit.

During the transition from the singlet to the triplet process continuity of coding was maintained. Explaining the emergence of RPS from the RNA world is major problem for the RWH.^{54,107,108,173} Evolution of RPS from a singlet coding system facilitates an explanation for the origin of RPS. Ostensibly, coevolution of two macromolecules, RNA and proteins, is not as parsimonious an evolutionary pathway as that of RNA alone. RNA has the ability to perform some of the functions of proteins, as shown by the ability of ribozymes to catalyze reactions similar to that of enzymes, but in other crucial respects RNA is unable to fulfill the function of proteins. In particular, RNA cannot act as an integral membrane protein, and therefore cannot be involved in nutrient and ion uptake, discharge of waste products, response to changes in the extracellular environment, and energy transduction. The coevolution approach allows proteins, from those initially containing only Gly and Ala to those containing Gly/Ala/Val/Asp to those containing the full range of amino acids, to be integral membrane proteins, providing a better evolutionary model than that provided by the apparently more parsimonious RWH.

References

- Orgel LE. Geothermal synthesis and metabolism. *Astrobiology* 2006;6:297-98.
- Orgel LE. The implausibility of metabolic cycles on the prebiotic Earth. *PloS Biol* 2008;6:e18.
- Joyce GF. The antiquity of RNA-based evolution. *Nature* 2002;418:214-21.
- Cech TR. Crawling out of the RNA World. *Cell* 2009;136:599-602.
- Lincoln TA, Joyce GF. Self-sustained replication of an RNA enzyme. *Science* 2009;323:1229-32.
- Whitesides GM. Revolutions in chemistry. *Chem Eng News* 2007;85:12-17.
- Francis BR. A hypothesis that ribosomal protein synthesis evolved from coupled protein and nucleic acid synthesis. *Chemtracts Biochem Mol Biol* 2000;13:153-191.
- Shapiro R. A replicator was not involved in the origin of life. *IUBMB Life* 2000;49:173-6.
- Shapiro R. Small molecule interactions were central to the origin of life. *Q Rev Biol* 2006;82:105-25.
- Powner MW, Gerland B, Sutherland JD. Synthesis of activated pyrimidine ribonucleotides in prebiotic plausible conditions. *Nature* 2009;459:239-42.
- Sowerby SJ, Petersen GB. Life before RNA. *Astrobiology* 2002;2:231-39.
- Hud NV, Jain SS, Li X, Lynn DG. Addressing the problems of base pairing and strand cyclization in template-directed synthesis: a case for the utility and necessity of 'molecular midwives' and reversible backbone linkages for the origin of proto-RNA. *Chem Divers* 2007;4:768-83.
- Segre D, Shenhev B, Kafri R, Lancet D. The molecular roots of compositional inheritance. *J Theor Biol* 2001;213:481-91.
- Wu M, Higgs PG. Compositional inheritance comparison of self-assembly and catalysis. *Orig Life Evol Biosph* 2008;38:389-418.
- de Duve C. Chemistry and selection. *Chem Biodivers* 2007;4:574-83.
- Kauffman SK. Origin of life and the living state. *Orig Life Evol Biosph* 2007;37:315-22.
- Rode BM, Fitz D, Jakschitz T. The first chemical evolution towards the origin of life. *Chem Biodivers* 2007;4:2674-702.
- Orgel LE. Prebiotic chemistry and the origin of the RNA World. *Crit Rev Biochem Mol Biol* 2004;39:99-123.
- Anastasi C, Buchet FF, Crowe MA, et al. RNA: prebiotic product or biotic invention. *Chem Biodivers* 2007;4:721-39.
- Orgel LE. Some consequences of the RNA world hypothesis. *Orig Life Evol Biosph* 2003;33:211-18.
- Chen X, Li N, Ellington AD. Ribozyme catalysis of metabolism in the RNA world. *Chem Biodivers* 2007;4:633-55.
- Doudna JA, Cech TR. The chemical repertoire of natural ribozymes. *Nature* 2002;418:222-8.
- Haugen P, Simon DM, Bhattacharya D. The natural history of group I introns. *Trends Genet* 2005;1:111-9.
- Lehmann K, Schmidt U. Group II introns: structure and versatility of large natural ribozymes. *Crit. Rev. Biochem Mol Biol* 2003;38:249-303.
- Dai L, Chai D, Gu SQ, et al. A three-dimensional model of a group II intron RNA and its interaction with intron-encoded reverse transcriptase. *Mol Cell* 2008;30:472-85.
- Rupert PB, Massey AP, Sigurdsson ST, Ferré-D'Amaré AR. Transition state stabilization by a catalytic RNA. *Science* 2002;298:1421-4.
- Iyer L, Koonin EV, Aravind L. Evolutionary connection between the catalytic subunits of DNA-dependent RNA polymerases and eukaryotic RNA-dependent RNA polymerases and the origin of RNA polymerases. *BMC Struct Biol* 2003;3:1-23.
- Shechner DM, Grant RA, Bagby SC, et al. Crystal structure of the catalytic core of an RNA-polymerase ribozyme. *Science* 2009;326:1271-5.
- Wilson TJ, Lilley DMJ. The evolution of ribozyme chemistry. *Science* 2009;323:1436-8.
- Guerrier-Takada C, Gardiner K, Marsh T, et al. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 1983;35:849-57.
- Beebe JA, Kurz JC, Fierke CA. Magnesium ions are required by *Bacillus subtilis* ribonuclease P RNA for both binding and cleaving precursor tRNAAsp. *Biochemistry* 1996;35:10493-505.
- Kazantsev AV, Pace NR. Bacterial RNase P: a new look at an ancient enzyme. *Nat Rev Microbiol* 2006;4:729-40.
- Ellis JC, Brown JW. The RNase P family. *RNA Biology* 2009;6:362-9.
- Marvin MC, Engelke DR. Broadening the mission of an RNA enzyme. *J Cell Biochemistry* 2009;108:1244-51.
- Lai LB, Vioque A, Kirsebom LA, Gopalan V. Unexpected diversity of RNase P, an ancient tRNA processing enzyme: challenges and prospects. *FEBS Lett* 2010;584:287-96.
- Ban N, Nissen P, Hansen J, Moore PB, Steitz TA. The complete atomic structure of the large ribosomal subunit at 2.4Å resolution. *Science* 2000;289:905-20.
- Nissen P, Hansen J, Ban N, et al. The structural basis of ribosome activity in peptide bond synthesis. *Science* 2000;289:920-30.
- Cech TR. Structural biology. The ribosome is a ribozyme. *Science* 2000;289:878-9.
- Yusupov MM, Yusupova GZ, Baucom A, et al. Crystal structure of the ribosome at 5.5Å resolution. *Science* 2001;292:883-96.
- Steitz TA, Moore PB. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends Biochem Sci* 2003;28:411-8.
- Carter AP, Clemons WM, Brodersen DE, et al. Structural insights from the structure of the 30S ribosomal subunit and its interaction with antibiotics. *Nature* 2000;407:340-8.
- Ogle JM, Ramakrishnan V. Structural insights into translational fidelity. *Ann Rev Biochem* 2005;74:129-77.
- Williamson JR. Small subunit, big science. *Nature* 2000;407:306-7.
- Woese CR. Translation: In retrospect and prospect. *RNA* 2001;7:1055-67.
- Klein DJ, Moore PB, Steitz TA. The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J Mol Biol* 2004b;340:141-77.
- Wilson DN, Nierhaus KH. Ribosomal proteins in the spotlight. *Crit Rev Biochem Mol Biol* 2005;40:243-67.
- Noller HF, Hoffarth V, Zimniak L. Unusual

- resistance of peptidyl transferase to protein extraction procedures. *Science* 1992;256:1416-19.
48. Smith TF, Lee JC, Gutell RR, Hartman H. The origin and evolution of the ribosome. *Biol Direct* 2008;3:16.
 49. Anderson RM, Kwon M, Strobel SA. Toward ribosomal RNA catalytic activity in the absence of protein. *J Mol Evol* 2007;64:472-83.
 50. Sykes MT, Williamson JR. A complex assembly landscape for the 30S ribosomal subunit. *Annu Rev Biophys* 2009;38:197-215.
 51. GavriloVA LP, KostiasHKina OE, KoteliAnsky VE, et al. Factor-free (non-enzymic) and factor dependent systems of translation of polyuridylic acid by *Escherichia coli* ribosomes. *J Mol Biol* 1976;101:537-52.
 52. Southworth DR, Brunelle JL, Green R. EFG-independent translocation of the mRNA:tRNA complex is promoted by modification of the ribosome with thiol-specific reagents. *J Mol Biol* 2002;324:611-23.
 53. Cochella L, Brunelle JI, Green R. Mutational analysis reveals two independent molecular requirements during transfer RNA selection on the ribosome. *Nat Struct Mol Biol* 2007;14:30-6.
 54. Noller HF. Evolution of protein synthesis from an RNA world. *Cold Spring Harb Perspect Biol* doi:10.1101/cshperspect.a003681.
 55. Liiv A, O'Conner M. Mutations in the intersubunit bridge regions in 23S rRNA. *J Biol Chem* 2006;281:29850-62.
 56. Selmer M, Dunham CM, Murphy FV, et al. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 2006;313:1935-42.
 57. Klein DJ, Moore PB, Steitz TA. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA* 2004;10:1366-79.
 58. Rodnina MV, Beringer M, Wintermeyer W. How ribosomes make peptide bonds. *Trends Biochem Sci* 2007;32:20-6.
 59. Hsiao C, Williams LD. A recurrent magnesium-binding motif provides a framework for the ribosomal peptidyl transferase center. *Nucleic Acids Res* 2009;37:3134-42.
 60. Agmon I. The dimeric proto-ribosome: structural details and possible implications on the origin of life. *Int J Mol Sci* 2009;10:2921-34.
 61. Noller HF. The driving force for molecular evolution of translation. *RNA* 2004;10:1833-7.
 62. Nissen P, Ippolito JA, Ban N, et al. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci USA* 2001;98:4899-903.
 63. Battle DJ, Doudna JA. Specificity of RNA-RNA helix recognition. *Proc Natl Acad Sci USA* 2002;99:11676-81.
 64. Bokov K, Steinberg SV. A hierarchical model for evolution of 23S ribosomal RNA. *Nature* 2009;457:977-80.
 65. Crick FHC. The origin of the genetic code. *J Mol Biol* 1968;38:367-79.
 66. Knight RD, Freeland SJ, Landweber LF. Selection, history and chemistry: the three faces of the genetic code. *Trends Biochem Sci* 1999;24:241-7.
 67. Koonin EV, Novozhilov AS. Origin and evolution of the genetic code: the universal enigma. *IUBMB Life* 2009;61:99-111.
 68. Wolf YI, Koonin EV. On the origin of the genetic code in the RNA world by means of natural selection, exaptation, and sub-functionalization. *Biol Direct* 2007;2:14.
 69. Rodin SN, Rodin AS. On the origin of the genetic code: signatures of its primordial complementarity in tRNAs and aminoacyl-tRNA synthetases. *Heredity* 2008;100:341-55.
 70. Crick FHC, Brenner S, Klug A, Piecznik G. A speculation on the origin of protein synthesis. *Origins Life* 1976;7:389-97.
 71. Eigen M, Winkler-Oswatitsch R. Transfer RNA, an early gene? *Naturwissenschaften* 1981;68:282-92.
 72. Trifonov EN. Consensus temporal order of amino acids and evolution of the triplet code. *Gene* 2000;261:139-51.
 73. Ikehara K. Origins of gene, genetic code, protein and life: comprehensive view of life systems from a GNC-SNS primitive genetic code hypothesis. *J Biosci* 2002;27:165-86.
 74. Lehmann J, Riedo B, Dietler G. Folding of small RNAs displaying the GNC base pattern: implications for the self-organization of the genetic system. *J Theor Biol* 2004;227:381-95.
 75. Rodin S, Ohno S. Two types of aminoacyl-tRNA synthetases could be originally encoded by complementary strands of nucleic acids. *Orig Life Evol Biosph* 1995;25:565-89.
 76. Carter CW, Duax WL. Did tRNA synthetase classes arise on opposite strands of the same gene? *Mol Cell* 2002;10:705-8.
 77. Rodin AS, Rodin SN, Carter CW Jr. On primordial sense-antisense coding. *J Mol Evol* 2009;69:555-67.
 78. Olejniczak M, Dale T, Fahlman RP, Uhlenbeck OC. Idiosyncratic tuning of tRNAs to achieve uniform ribosome binding. *Nat Struct Mol Biol* 2005;12:788-93.
 79. Dale T, Fahlman RP, Olejniczak M, Uhlenbeck OC. Specificity of the ribosomal A site for aminoacyl-tRNAs. *Nucleic Acids Res* 2009;37:1202-10.
 80. Urbonavicius J, Qian Q, Durand JM, et al. Improvement of reading frame maintenance is a common function for several tRNA modifications. *EMBO J* 2001;20:4863-73.
 81. Agris PF. Decoding the genome: a modified view. *Nucleic Acids Res* 2004;32:223-38.
 82. Konevega AL, Soboleva NG, Makhno VI, et al. Purine bases at position 37 of tRNA stabilize codon-anticodon interaction in the ribosomal A site by stacking and Mg²⁺-dependent interactions. *RNA* 2004;10:90-101.
 83. Agris PF. Bringing order to translation: the contributions of transfer RNA anticodon-domain modifications. *EMBO Rep* 2008;9:629-35.
 84. Auffinger P, Westhof E. Location and distribution of modified nucleotides in tRNA. In: Grosjean H, Benne R, editors. *Modification and Editing of RNA*. Washington, DC: ASM Press;1998.
 85. Bjork GR. Appendix 6. Modified nucleosides at positions 34 and 37 of tRNAs and their predicted coding capacities. In: Grosjean H, Benne R, editors. *Modification and Editing of RNA*. Washington, DC: ASM Press;1998.
 86. Pütz J, Florentz C, Benseler F, Giegé R. A single methyl group prevents the mischarge of a tRNA. *Nat Struct Biol* 1994;1:580-2.
 87. Agris PF, Vendeix FA, Graham WD. tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol* 2007;366:1-13.
 88. Tong K-L, Wong JT. Anticodon and wobble evolution. *Gene* 2004;333:169-77.
 89. Yang Y-W, Zhang S, McCullum EO, Chaput JC. Experimental evidence that GNA and TNA were not sequential polymers in the prebiotic evolution of RNA. *J Mol Evol* 2007;65:289-95.
 90. Weber AL, Orgel LE. Poly(U)-directed peptide bond formation from the 2'(3')-glycyl esters of adenosine derivatives. *J Mol Evol* 1980;16:1-10.
 91. Ling J, Reynolds N, Ibba M. Aminoacyl-tRNA synthesis and translational quality control. *Annu Rev Microbiol* 2009;63:61-78.
 92. Ahel I, Korencic D, Ibba M, Söll D. Trans-editing of mischarged tRNAs. *Proc Natl Acad Sci USA* 2003;100:15422-7.
 93. Sokabe M, Okada A, Yao M, et al. Molecular basis of alanine discrimination in editing site. *Proc Natl Acad Sci USA* 2005;102:11669-74.
 94. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. *Science* 1997;276:131-3.
 95. Guo M, Chong YE, Beebe K, et al. The C-Ala domain brings together editing and aminoacylation functions on one tRNA. *Science* 2009;325:744-7.

96. Doyle DA, Morais Cabral J, Pfuetzner RA, et al. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science* 1998;280:69-77.
97. Boudker O, Ryan RM, Yernool D, et al. Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature* 2007;445:387-93.
98. Hirsch AKA, Fischer FR, Diederich F. Phosphate recognition in structural biology. *Angew Chem Int Ed* 2007;46:338-52.
99. Ng KK, Arnold JJ, Cameron CE. Structure-function relationships among RNA-dependent RNA polymerases. *Curr Top Microbiol Immunol* 2008;320:137-56.
100. Tadokoro T, Kanaya S. Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J* 2009;276:1482-93.
101. Luecke H, Quijcho FA. High specificity of a phosphate transport protein determined by hydrogen bonds. *Nature* 1990;347:402-6.
102. Zhang B, Cech TR. Peptidyl-transferase ribozymes: trans reactions, structural characterization and ribosomal RNA-like features. *Chem Biol* 1998;5:539-53.
103. Sun L, Cui Z, Gottlieb RL, Zhang B. A selected ribozyme catalyzing diverse dipeptide synthesis. *Chem Biol* 2002;9:619-28.
104. Chumachenko NV, Novikov Y, Yarus M. Rapid and simple ribozymic aminoacylation using 3 conserved nucleotides. *J Am Chem Soc* 2009;131:5257-63.
105. Yarus M, Widmann JJ, Knight R. RNA-amino acid binding: a stereochemical era for the genetic code. *J Mol Evol* 2009;69:406-29.
106. Roth A, Breaker RR. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* 2009;78:305-34.
107. Yarus M. Amino acids as RNA ligands: a direct RNA-template theory for the code's origin. *J Mol Evol* 1998;47:109-17.
108. Ma W. The scenario on the origin of translation in the RNA world: in principle of replication parsimony. *Biology Direct* 2010;5:65.
109. Di Giulio M. The non-monophyletic origin of the tRNA molecule. *J Theor Biol* 1999;197:403-14.
110. Tanaka T, Kikuchi Y. Origin of the cloverleaf shape of transfer RNA – the double-hairpin model: implications for the role of tRNA intron and the long extra loop. *Viva Origino* 2001;29:134-42.
111. Randau L, Münch R, Hohn MJ, et al. Nanoarchaeum equitans creates functional tRNAs from separate genes for their 5'- and 3'-halves. *Nature* 2005;433:537-41.
112. Söll D. Transfer RNA: an RNA for all seasons. In: Gesteland RF, Atkins JF, editors. *The RNA World*. New York: Cold Spring Harbor Press;1993.
113. Blumenthal T, Carmichael GG. RNA replication: function and structure of Q β -replicase. *Annu Rev Biochem* 1979;48:525-48.
114. Dreher TW. Role of tRNA-like structures in controlling plant virus replication. *Virus Res* 2009;139:217-29.
115. Weiner AM, Maizels N. 3'terminal tRNA-like structures tag genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc Natl Acad Sci USA* 1987;84:7383-7.
116. Weiner AM, Maizels N. The genomic tag hypothesis: modern viruses as molecular fossils of ancient strategies for genomic replication, and clues regarding the origin of protein synthesis. *Biol Bull* 1999;196:327-30.
117. Vogel A, Schilling O, Spath B, Marchfelder A. The tRNase Z family of proteins: Physiological functions, substrate specificity, and structural properties. *Biol Chem* 2005;386:1253-64.
118. Ow MC, Kushner SR. Initiation of tRNA maturation by RNase E is essential for cell viability in *E.coli*. *Genes Dev* 2002;16:1102-15.
119. Vörtler S, Mörl M. tRNA-nucleotidyltransferases: highly unusual RNA polymerases with vital functions. *FEBS Lett* 2009;584:297-302.
120. McClain WH. Rules that govern tRNA identity in protein synthesis. *J Mol Biol* 1993;234:257-80.
121. Saks ME, Sampson JR, Abelson JN. The transfer RNA identity problem; a search for rules. *Science* 1994;263:191-7.
122. Giegé R, Sissler M, Florentz C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res* 1998;26:5017-35.
123. Mazauric MH, Roy H, Kern D. tRNA glycylation system from *Thermus thermophilus*. tRNAGly identity and functional interrelation with glycylation systems from other phylae. *Biochemistry* 1999;38:13094-105.
124. Hou YM, Francklyn C, Schimmel P. Molecular dissection of a transfer RNA and the basis for its identity. *Trends Biochem Sci* 1989;14:233-7.
125. McClain WH, Foss K, Jenkins RA, Schneider J. Four sites in the acceptor helix and one site in the variable pocket of tRNA^{Ala} determine the molecule's acceptor identity. *Proc Natl Acad Sci USA* 1991;88:9272-6.
126. Fukai S, Nureki O, Sekine SI, et al. Mechanism of molecular interactions for tRNA^{Val} recognition by valyl-tRNA synthetase. *RNA* 2003;9:100-11.
127. Giegé R, Florentz C, Kern D, et al. Aspartate identity of transfer RNAs. *Biochimie* 1996;78:605-23.
128. Voorhees RM, Weixlbaumer A, Loakes D, et al. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* 2009;16:528-33.
129. Eriani G, Delarue M, Poch O, et al. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* 1990;347:203-6.
130. Ibba M, Bono JL, Rosa PA, Söll D. Archaeal-type lysyl-tRNA synthetase in the spirochete *Borrelia burgdorferi*. *Proc Natl Acad Sci USA* 1997;94:14383-8.
131. Schimmel P, Giegé R, Moras D, Yokoyama S. An operational code for amino acids and possible relationship to genetic code. *Proc Natl Acad Sci USA* 1993;90:8763-8.
132. Ibba M, Söll D. Aminoacyl-tRNA synthesis. *Annu Rev Biochem* 2000;69:617-50.
133. Ribas de Pouplana L, Schimmel P. Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. *Cell* 2001;104:191-3.
134. Swairjo MA, Otero FJ, Yang XL, et al. Alanyl-tRNA synthetase crystal structure and design for acceptor-stem recognition. *Mol Cell* 2004;13:829-41.
135. Naganuma M, Sekine S, Fukunaga R, Yokoyama S. Unique protein architecture of alanyl-tRNA synthetase for aminoacylation, editing and dimerization. *Proc Natl Acad Sci USA* 2009;106:8489-94.
136. Sokabe M, Ose T, Nakamura A, et al. The structure of alanyl-tRNA synthetase with editing domain. *Proc Natl Acad Sci USA* 2009;106:11028-33.
137. Shimizu S, Juan EC, Sato Y, et al. Two complementary enzymes for threonylation of tRNA in crenarchaeota: crystal structure of *Aeropyrum pernix* threonyl-tRNA synthetase lacking a cis-editing domain. *J Mol Biol* 2009;394:286-96.
138. O'Donoghue P, Luthy-Schulten Z. On the evolution of structure in aminoacyl-tRNA synthetases. *Microbiol Mol Biol Rev* 2003;67:550-73.
139. Wolf YI, Aravind L, Grishin NV, Koonin EV. Evolution of aminoacyl-tRNA synthetases-analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res* 1999;9:639-710.
140. Woese CR, Olsen GJ, Ibba M, Söll D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev* 2000;64:202-36.
141. Aravind L, Anantharaman V, Koonin EV.

- Monophyly of class I aminoacyl tRNA synthetase, USPA, ETPF, photolyase, and PP-ATPase nucleotide-binding domains: implications for protein evolution in the RNA world. *Proteins* 2002;48:1-14.
142. Ribas de Pouplana L, Turner RJ, Steer BA, Schimmel P. Genetic code origins: tRNAs older than their synthetases? *Proc Natl Acad Sci USA* 1998;95:11295-300.
 143. Ibbá M, Losey HC, Kawarabayasi Y, et al. Substrate recognition by class I lysyl-tRNA synthetases: a molecular basis for gene displacement. *Proc Natl Acad Sci USA* 1999;96:418-23.
 144. Hohn MJ, Park HS, O'Donoghue P, et al. Emergence of the universal genetic code imprinted in an RNA record. *Proc Natl Acad Sci USA* 2006;103:18095-100.
 145. Nagel GM, Cumberledge S, Johnson MS, et al. The β subunit of *E. coli* glycyl-tRNA synthetase plays a major role in tRNA recognition. *Nucleic Acids Res* 1984;12:4377-84.
 146. Toth MJ, Schimmel P. A mutation in the small (α) subunit of glycyl-tRNA synthetase affects amino acid activation and subunit association parameters. *J Biol Chem* 1990;265:1005-9.
 147. Bonnefond L, Giegé R, Rudinger-Thirion J. Evolution of the tRNA^{Tyr}/TyrRS aminoacylation systems. *Biochimie* 2005;87:873-3.
 148. Charron C, Roy H, Blaise M, et al. Non-discriminating and discriminating aspartyl-tRNA synthetases differ in the anticodon-binding domain. *EMBO J* 2003;22:1632-43.
 149. Hartman H, Smith TF. GTPases and the origin of the ribosome. *Biol Direct* 2010; 5:36.
 150. Silvan LF, Wang J, Steitz TA. Insights into editing from an Ile-tRNA synthetase structure with tRNA^{Ile} and mupirocin. *Science* 1999;285:1074-7.
 151. Ribas de Pouplana L, Frugier M, Quinn CL, Schimmel P. Evidence that two present-day components needed for the genetic code appeared after nucleated cells separated from eubacteria. *Proc Natl Acad Sci USA* 1996;93:166-70.
 152. Lin L, Hale SP, Schimmel P. Aminoacylation error correction. *Nature* 1996;384:33-4.
 153. Zhao MW, Zhu B, Hao R, et al. Leucyl-tRNA synthetase from the ancestral bacterium *Aquifex aeolicus* contains relics of synthetase evolution. *EMBO J* 2005; 24:1430-9.
 154. Serrano A, Pérez-Castiñeira JR, Baltscheffsky M, Baltscheffsky H. H⁺-PPases: yesterday, today and tomorrow. *IUBM Life* 2007;59:76-83.
 155. Li H, Fast W, Benkovic SJ. Structural and functional modularity of proteins in the de novo purine biosynthetic pathway. *Protein Sci* 2009;18:881-92.
 156. Lehman N, Jukes TH. Genetic code by stop codon takeover. *J Theor Biol* 1988;135:203-14.
 157. Yuan J, O'Donoghue P, Ambrogelly A, et al. Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. *FEBS Lett* 2010;584:342-9.
 158. Zaher HS, Green R. Quality control by the ribosome following peptide bond formation. *Nature* 2009;457:161-6.
 159. Nomura N, Sako Y, Uchida A. Molecular characterization and postsplicing fate of three introns within the single rRNA operon of the hyperthermophilic archaeon *Aeropyrum pernix* K1. *J Bacteriol* 1998;180:3635-43.
 160. Itoh T, Nomura N, Sako Y. Distribution of 16S rRNA introns among the family Thermoproteaceae and their evolutionary implications. *Extremophiles* 2003;7:229-33.
 161. Yoshinari S, Itoh T, Hallam SJ, et al. Archaeal pre-mRNA splicing: a connection to hetero-oligomeric splicing endonuclease. *Biochem Biophys Res Commun* 2006;346:1024-32.
 162. Toro N, Jimenez-Zurdo JI, Garcia-Rodriguez FM. Bacterial group II introns: not just splicing. *FEMS Microbiol Rev* 2007;31:342-58.
 163. Brügger K, Chen L, Stark M, et al. The genome of *Hyperthermus butylicus*: a sulfur-reducing, peptide fermenting, neutrophilic Crenarchaeote growing up to 108°C. *Archaea* 2007; 2:127-35
 164. Omer AD, Ziesche S, Decatur WA, et al. RNA-modifying machines in archaea. *Mol Microbiol* 2003;48:617-29.
 165. Rodriguez-Trelles F, Tarrío R, Ayala FJ. Origins and evolution of spliceosomal introns. *Annu Rev Genet* 2006;40:47-76.
 166. Penny D, Hoepfner MP, Poole AM, Jeffares DC. An overview of the introns-first theory. *J Mol Evol* 2009;69:527-40.
 167. Dennis PP, Omer A. Small non-coding RNAs in archaea. *Curr Opin Microbiol* 2005;8:685-94.
 168. Storz G, Altuvia S, Wassarman KM. An abundance of RNA regulators. *Annu Rev Biochem* 2005;74:199-217.
 169. Terns M, Terns R. Noncoding RNAs of the H/ACA family. *Cold Spring Harb Symp Quant Biol* 2006;71:395-405.
 170. Hannon GJ, Rivas FV, Murchison EP, Steitz JA. The expanding universe of non-coding RNAs. *Cold Spring Harb Symp Quant Biol* 2006;71:551-64.
 171. Repoila F, Darfeuille F. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biol Cell* 2009;101:117-31.
 172. Sharp PA. The centrality of RNA. *Cell* 2009;136:577-80.
 173. Cech TR. Crawling out of the RNA world.