CMLS, Cell. Mol. Life Sci. 56 (1999) 85–93
1420-682X/99/020085-09 $ 1.50 + 0.20/0

CMLS Cellular and Molecular Life Sciences

Review

Evolution of the genetic code, protein synthesis and nucleic acid replication

S. Alberti

Unit of Experimental Oncology, Department of Cell Biology and Oncology, Istituto di Ricerche Farmacologiche Mario Negri-Consorzio Mario Negri Sud, I-66030 Santa Maria Imbaro, Chieti (Italy), Fax + 39 0872 578240, e-mail: alberti@cmns.mnegri.it

Received 16 December 1998; received after revision 15 February 1999; accepted 1 April 1999

Abstract. A better definition of the structural and thermodynamic determinants of the interaction of nucleic acids with proteins is shedding light on the origin of the genetic code, protein synthesis, and nucleic acid replication. This is also allowing to show a consistent biochemical framework for the appearance of these fundamental synthetic mechanisms. This article reviews recent significant developments in the field, and discusses an integrated model for a biochemically plausible evolution of these fundamental mechanisms of synthesis. This model is based on sequence-specific interactions between abiotically synthesized polynucleotides and polypeptides, and can account for a coordinate evolution of the genetic code, protein synthesis, and nucleic acid replication in living cells.

Key words. Genetic code; protein synthesis; nucleic acid replication; evolution.

Introduction

Several articles have recently reassessed the origin of the genetic code and protein synthesis [1–8]. This renewed interest stems from recent biochemical and structural advances, a major one being the discovery that RNA can have specific enzymatic properties [9–13]. This finding has profoundly changed current theories on the origin of life [14], and has offered a unifying view of the evolution of information-encoding molecules. A much better understanding of the rules of interaction between nucleotides and polypeptides [15–23], and improved knowledge of the regulation of nucleic acid and protein synthesis [24] are also contributing to this progress. Finally, technological advances now allow adequate testing of the basic tenets of evolutionary theories.

This article reviews recent developments in this field. A model of molecular evolution that may allow their integration in a biochemically plausible framework will be discussed.

Stereochemical affinity or 'frozen accident'?

The stereochemical [25, 26] and 'frozen accident' theories [27, 28] are two traditionally opposing views on the origin of the genetic code. The stereochemical theory predicts significant affinity between specific codon and amino acid pairs [25]. As a consequence, the code would derive from selective chemical interactions between nucleotide bases and amino acids. The existence of a strong driving force for the evolution of the code is appealing, as it may explain why the genetic code appeared so early, being almost as old as our planet [29]. This would also imply that the type of code that we know is essentially a necessity. However, this theory is difficult to accept in its original formulation, since there is little evidence for selective binding of amino acids to isolated codons or anticodons [5, 7, 16, 30]. This is hardly surprising, considering merely the sheer difference in size between amino acids and triplets of nucleic acid bases [31].

The apparently opposing theory of the 'frozen accident' is a stochastic model which predicts a random assignment of amino acids to codons. It also affirms that the genetic code is universal only because all living organisms derive from a single primordial cell. This theory highlights the vast selective advantage of encoded protein synthesis, whatever mechanism it originates from. Moreover, it accounts well for the invariance of the mature code. However, it says little about the molecular mechanisms that generated the genetic code, unless one postulates the probabilistic monster of the sudden appearance of a cell with mature genetic code and protein synthesis apparatus.

State of the art

A large variety of analyses have been recently provided with complementary contributions from different fields. An algebraic analysis of the evolution of the code interprets it as a symmetry-breaking of a 16-dimensional potential code [6]. This approach defines an evolutionary hierarchy among the different codons, i.e., it identifies groups of codons that are likely to derive from a common progenitor [28]. Using a different approach, an evolutionary hierarchy among different codons and amino acids has also been defined by extracting the simplest biochemical and combinatorial features from the 64 codons and 20 amino acids [32]. The recognition of codon symmetries [16, 30, 33], i.e., 'the code within codons' [34], is an operationally similar approach. This has succeeded in constructing homogeneous groups of codon-amino acid pairs, based on chemical relationships between amino acids and sequence similarity among the corresponding codons. Indeed, 'chemically close' codons are contiguous, i.e., only one base changes between chemically consecutive groups. More general models, like 'RRY,' 'RNY,' or 'G-nonG-N' [27], lead to similar conclusions. Thus, plausible models for an evolution of the code from a primitive form to the mature one can be drawn.

In a few instances, amino acids, e.g., arginine and isoleucine, have been demonstrated to specifically bind ribozymes or in-vitro-selected RNA containing their corresponding codons [21–23]. This indicates that a stereochemically specific binding between amino acids and RNA codons can exist, and that evolutionary models for the code based on stereochemical interactions are possible [35, 36]. However, specific RNA-amino acid recognition has been observed only after stringent RNA selection [9, 11, 22, 23], and it is unclear if a correspondingly selective pressure could have existed in a primordial environment.

The difficult issue of the relationship between the genetic code and protein synthesis has also been tackled. Reductionistic approaches, e.g., that ordinate coding might have preceded translation, have been proposed [4, 5]. Crucial to these models are diverse enzymatic activities of the primordial RNA that are consistent with experimental evidence [9–12]. However, these models do not indicate how an isolated genetic code could have developed and gained an evolutionary advantage before the appearance of protein synthesis. A significant advance toward the solution of this problem has been the recognition that specific RNA secondary structures could have been utilized to recognize specific amino acids [1]. This model is based on the presence of aminoacyl-tRNA ribozymic synthetases and of minihelix tRNAs [3, 37], and predicts the appearance of a non-coded protein synthesis, which subsequently evolved into a coded form. Thus, it provides a critical link between the genetic code and protein synthesis. However, it requires ordinate recognition between different tRNAs and aminoacyl-tRNA synthetases and the existence of abundant structurally homogeneous RNA. A further difficulty is the potential necessity for up to 8000 different aminoacylated tRNA forms [3].

Open problems

Despite the significant advances outlined above, these theories have been subjected to strong criticism. First, it appears highly unlikely that complex molecules, like coenzyme handles [5], minihelix tRNAs [3, 37] or ribozyme-type RNA [9, 11, 21–23, 38, 39] appeared all of a sudden as functional entities. Critical to this point is the unlikelihood of selective catalysis, necessary for ordinate synthetic pathways, in a 'primordial soup.' More generally, the models discussed generate a non-solvable incongruity, often referred to as the 'chicken and egg' problem. Synthetically, if complex structures are necessary for the appearance of other complex structures, one should first explain how the first ones were generated. A solution to this paradox inevitably requires the initial steps to be simple from both a structural and a biochemical standpoint. Later steps can subsequently become more and more complex, by introducing random changes (primordial mutations) and fixing them by selection, if these are evolutionarily advantageous [24].

A further critical problem is represented by the evolution of the code over time. Indeed, the evolution of the code, e.g. from a simple single-base to a triplet code, is bound to lead to a devastating shift in the reading frame of the pre-existing mRNA [28, 40], which would annihilate all the advantages from the preceding stages of evolution. On the other hand, there are consistent indications that this evolution not only occurred, but also occurred rapidly, since the present format of the code is extremely ancient and is essentially invariant in

all living organisms [29]. None of the theories presented allows for a consistent explanation of this dicotomy.
Moreover, most theories on the appearance of the genetic code say little about the evolution of protein synthesis [3, 27]. This point is far from trivial, since it is highly unlikely that a genetic code gained an evolutionary advantage in the absence of protein synthesis, i.e., of a selectable product. It is also difficult to imagine that protein synthesis could have been 'added' to an already existing code.

Biochemical plausibility of evolutionary models

Biochemical plausibility of models of the evolution of the code and protein synthesis is required. Consistency with knowledge of peptide/nucleic acid interactions and with the thermodynamic characteristics of these processes is also necessary.
A second characteristic of a credible evolutionary theory is chemical and structural simplicity, to be compatible with evolution in a poorly organized primordial environment. It should also be possible to define coherent paths of development for the genetic code, protein synthesis, and nucleic acid replication, using a minimum number of ad hoc assumptions. It is, indeed, difficult to accept that three intertwined pathways of such complexity could have evolved separately. On the other hand, parallel evolution is able to provide a significant evolutionary advantage, e.g., by cross-feeding relevant substrates and regulatory processes from one pathway to the other.
Finally, the model presented should allow testable predictions to be drawn, and be experimentally testable, exiting from 'theory only' untestable models [32].

A coordinate evolution of the genetic code and protein synthesis

Evolutionary models based on the above premises can be developed [8]. A critical requirement for these models, i.e., that they be simple enough to be compatible with a poor primordial biochemistry, can be tackled by exploiting current knowledge on protein-nucleic acid interactions [9–13, 15–23]. Schematically, strong and selective binding between amino acids and nucleotide bases can be provided by repetitive interactions between specific amino acid side chains and specific bases, i.e., between polypeptides and polynucleotides. Starting from this interaction between polymers, an 'updated' stereochemical theory can be built that overcomes the main problem of the original theory, i.e. the insufficient strength and selectivity of binding of different amino acids to specific nucleic acid triplets. The only critical assumption of this model is that polynucleotides and polypeptides coexisted in sufficient amounts, although the underlying chemistry is still far from having been solved [41–49].
Sequence-specific interactions between polypeptides and polynucleotides would result in the accumulation of specific polypeptide-polyribonucleotide pairs. This created the correspondence between defined peptide and nucleotide sequences that is the core of a primordial genetic code.
Proximity between a peptide and an RNA molecule is likely to favor the formation of ester bonds between them (fig. 1A) [9, 16, 50, 51]. As a consequence of the covalent bonding, sequence-specific pairs of peptides and RNA (primordial-loaded tRNA) accumulated.
The appearance of primordial-loaded tRNA played a central role not only in the generation of the genetic code, but also in the evolution of protein synthesis. Single-stranded regions of loaded tRNA interacted with other RNA with complementary sequence (fig. 1A–D) (primordial mRNA). As a consequence, different charged tRNA molecules were brought together by primordial mRNA in a sequence-dependent way (fig. 1B). Proximity between two RNA-bound peptides increased chances of trans-esterification (fig. 1C) [49]. This resulted in one free tRNA and one loaded tRNA now bearing a longer peptide. Release of the free tRNA (fig. 1D) and the juxtaposition of further loaded tRNA molecules allowed a progressive increase in the polypeptide length. This process is a catalyzed synthesis of longer polypeptides, and is templated by specific polynucleotide sequences, i.e., it possesses the critical characteristics of encoded protein synthesis. Interestingly, this model predicts that the primordial synthetic apparatus was composed of nucleic acids only, which is consistent with other independent predictions [25, 27, 28].

Evolution of the genetic code

Significant interactions between nucleic acids and peptides are likely to have occurred only over relatively long regions. The main reason for this assumption is the need for sufficiently strong binding [17–20] at the higher temperatures characteristic of the 'primordial soup' [48]. As a result, the early genetic code was composed by elements much longer than the base-triplet single amino acids seen nowadays. How could a 'longer' code evolve to its present form? The evolution of cofactors that stabilized RNA-RNA binding (primordial ribosomes) led to a progressive parallel reduction in the length of the interacting polynucleotides and polypeptides (see below). This permitted the evolution of a translation machinery without causing frameshifts, i.e., it allowed conservation of the sequence and viabil-

ity of earlier encoded proteins. Maintaining the reading frames of pre-existing mRNA is a fundamental advantage over evolutionary models that postulate evolution from a one-base to a two/three-base codon [28]. The mature code then evolved to the absolute minimum of a single amino acid per RNA coding unit. Sterical constraints [28], the need for sufficient thermal binding stability, and the requirement for a sufficient number of codons is likely to have imposed the *lower* limit of a three-base codon. In summary, the triplet code is not the longest evolved code, but the shortest.

Evolution of protein synthesis

The appearance of accessory molecules allowed further evolution of protein synthesis. Likely early cofactors were polypeptides with either catalytic properties for tRNA loading (primordial aminoacyl-tRNA synthetases) [52], or that stabilized the interaction between tRNA and mRNA (primordial ribosomes) [53]. Primordial aminoacyl-tRNA synthetases are likely to have added finer discrimination among the different amino acids, whereas primordial ribosomes increased both the fidelity and the efficiency of the primordial translation process. They also improved the utilization of the available biologically significant material. Indeed, while the coding capability, i.e., the sequence-specific recognition between polynucleotides and polypeptides, of long versus short polymers is equivalent, long polymers were probably rare [40, 41, 44, 46–48, 54, 55], whereas short ones were more abundant, but could bind less strongly to each other.

In summary, both primordial aminoacyl-tRNA synthetases and primordial ribosomes were favorably selected, allowing a critical step up in the efficiency of protein synthesis.

Further evolution of the primordial code and of protein synthesis

The core of the presented model is stereochemically deterministic. However, it allows for the evolution of

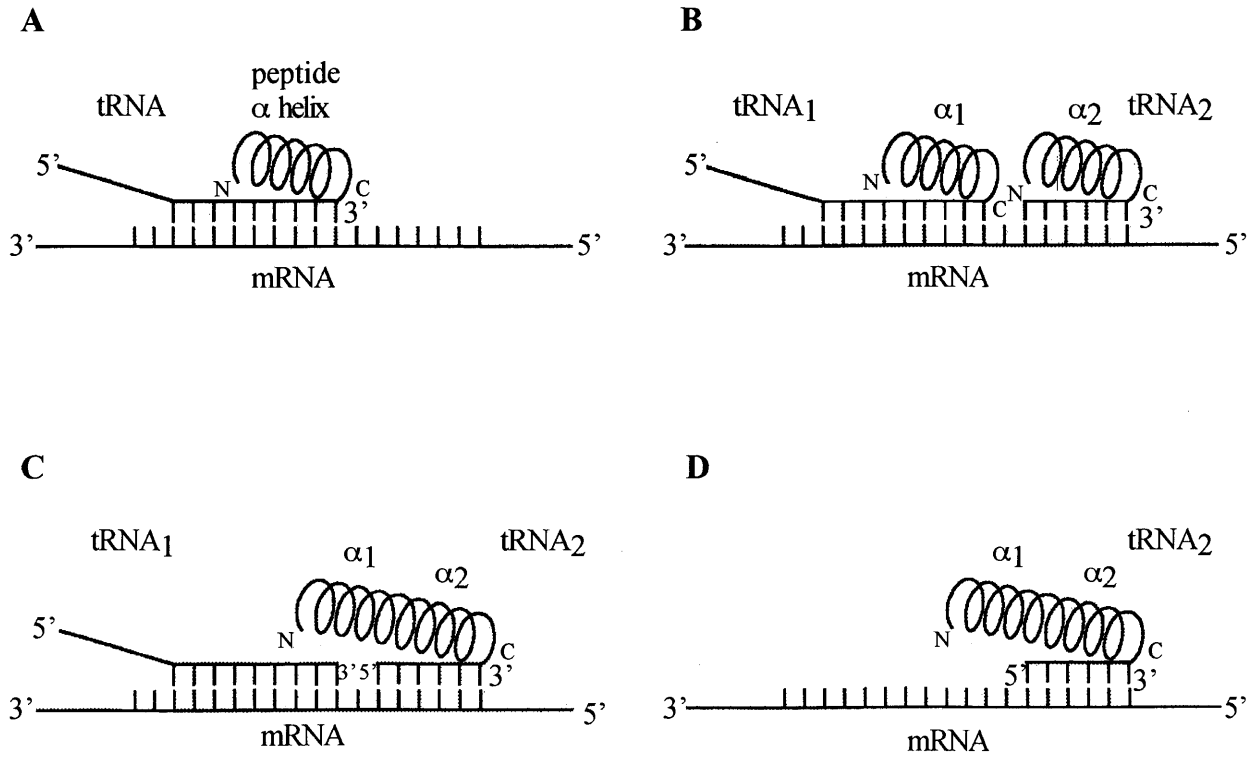

Figure 1. Evolution of the primordial genetic code and protein synthesis. (*A*) Generation of a primordial genetic code. (*B–D*) Generation of primordial protein synthesis. N: N terminal; C: C terminal; α_1: α helix polypeptide 1; α_2: α helix polypeptide 2. (*A*) Sequence-specific interactions between double-stranded RNA regions and oligopeptides in α helix conformation, and of complementary primordial tRNA and mRNA. (*B*) Sequence-specific juxtaposition of different loaded tRNA. (*C*) Proximity-enhanced trans-esterification reaction between tRNA-bound peptides. (*D*) Release of unloaded tRNA. (Reproduced from ref. 8, with kind permission of Springer-Verlag.)

the code and for the fixation of specific subsets of its potential variants. It is, indeed, difficult to imagine that primordial, i.e. unselected for, stereochemical interactions were sufficient to confer adequate selectivity among similar amino acids. On the other hand, cofactors (primordial aminoacyl-tRNA synthetases) are likely to have helped in refining tRNA-amino acid recognition. This stage of evolution of the genetic code was, thus, guided, which probably accounts for most of its orderly features, i.e., internal symmetries [6]. Good examples of these are the codons for acidic amino acids and the corresponding amides. GAA and GAG encode glutamic acid, while CAA and CAG encode glutamine. Correspondingly, GAC and GAU encode aspartic acid, while AAC and AAU encode asparagine. Thus, GAN encodes acidic residues, whereas the corresponding amide codons are encoded by AAN, i.e., they only differ in the first base. In summary, the appearance of aminoacyl-tRNA synthetases allowed a first expansion of the probably incomplete primordial alphabet [5, 6, 28], by the addition of chemically similar amino acids. Further expansion also occurred through the stabilization of aminoacid/RNA interactions that were too weak to be significant, or by developing some that did not previously exist.

Discrimination between similar amino acids is likely to have meant utilization of multiple tRNA recognition elements and specific three-dimensional structures. It also required guidance by more and more evolved aminoacyl-tRNA synthetases. This process of structural specialization led to a separation of functions, e.g., of codon versus acceptor stems in tRNAs.

Each step of the evolution of the code and protein synthesis is likely to have conferred a vast selective advantage to the primordial cell possessing it. Thus, it probably allowed a quick replacement of the preceding stages of cellular evolution. Further evolution then occurred among the descendants of each evolutionary bottleneck. This indicates that the core feature of the 'frozen accident' theory, i.e., that the code is unique because all living organisms derive from a single primordial cell [27], is consistent with a stereochemical theory. Coherent with this analysis, significant exceptions to the universality of the code appear in cases of radically different evolutionary history, e.g., the genomes of intracellular organelles [16].

Nucleic acid replication

A powerful model for the evolution of the genetic code and protein synthesis should be coherent with models for nucleic acid replication (see above). Template-directed RNA synthesis probably appeared in the form of RNA-guided RNA synthesis [13], and was the precursor of both mRNA transcription and nucleic acid replication. This process would have both increased the availability of useful primordial mRNA, and favored the genetic process of nucleic acid replication, as is the case for replicating RNA viruses [24]. The template-replicated RNA subsequently evolved into primordial DNA by retrotranscription. The appearance of the primordial DNA and genome-containing cells permitted the evolution of replicating organisms, with an ordinate capacity to fix, maintain and transmit the advantages of previous evolutionary steps.

Several models of primordial RNA transcription/replication can be envisaged. However, the most parsimonious one is probably an RNA-only nucleic acid replication mechanism [13]. 'Folded-back' RNA molecules were probably abundant in the primordial RNA pool (fig. 2A). These would have provided a simple start for nucleic acid synthesis, since they would have contained both a template and a primer for polynucleotide elongation. This mechanism would have directed the incorporation of single nucleotides or the condensation of longer complementary oligonucleotides in a strand-complementary fashion (fig. 2B, C). A consistent remnant of this primordial mechanism is the use of RNA primers for the initiation of the synthesis of replicating DNA [24, 32].

The condensation of complementary oligonucleotides over a nucleic acid template is in principle equivalent to the primordial mRNA/tRNA recognition described above as the core of protein synthesis (fig. 1). The only significant difference is the nature of the tRNA, i.e., peptide loaded in the case of protein synthesis versus peptide free in the case of nucleic acid replication. The end point of the two types of RNA-RNA interactions is also different, i.e., release of peptide-free tRNA in the case of peptide synthesis (fig. 1D) versus condensation into longer nucleotide polymers in the case of nucleic acid synthesis (fig. 2C). Assuming the naïve model of two tRNA pools, i.e., peptide loaded versus peptide free, in competition with each other, nucleic acid replication would be alternative to protein synthesis, a feature largely maintained by mature DNA synthesis versus mRNA transcription [24]. An experimental finding coherent with the model proposed is the conservation of tRNA-like structures in the genome that are likely to have been involved in nucleic acid replication [56]. An intriguing similarity between the folded-back RNA precursor (fig. 2A) and telomeres [57] suggests that the former could have originated the process of chromosome end-capping. Consistent with the model presented is also the observation that the top half (acceptor stem and TpsiC) of the tRNA is more ancient than the anticodon-containing half [56]. Equally consistent with this model of RNA-catalyzed condensation of

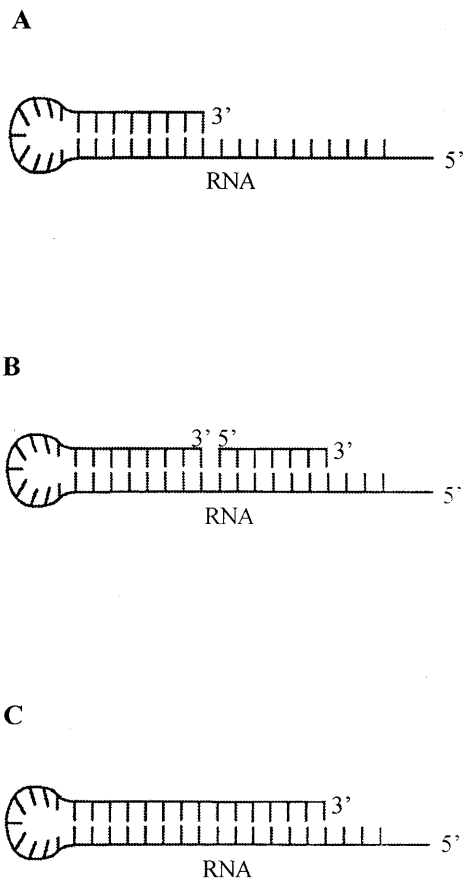

Figure 2. Evolution of primordial RNA replication. (*A*) Generation of double-stranded RNA regions by RNA back-folding. (*B*) Sequence-specific juxtaposition of RNA building blocks over complementary primordial RNA templates. (*C*) Condensation of the juxtaposed RNA into longer sequence-specific polymers.

short polymers into longer ones is the demonstration that RNA can possess an RNA ligase activity [58]. Primordial replication did not need to be limited to 'folded-back' RNA. The use of independent RNA molecules allowed the process of physical separation of the replicated molecules, permitting further rounds of synthesis and exponential increase in the copy number of sequence-identical/complementary RNA.

The proposed model indicates a strong potential link between nucleic acid replication and protein synthesis, consistent with suggestions from previous work. Orgel [59] proposed that amino acid charging at the 3′ terminus of tRNA would have favored RNA replication from the opposite end. Interestingly, this model implies that nucleic acid synthesis would follow the 'correct' 5′ to 3′ direction. Elongation of the 'folded-back' RNA in the model depicted would also impose a 5′ to 3′ direction of synthesis. Correspondingly, it would progressively shield the single-stranded RNA portions from hybridization with peptide-loaded tRNA. Thus, a 5′ to 3′ directed RNA elongation would impart an N- to C-terminal direction to peptide elongation.

Evolution of a primordial cell

The appearance and evolution of the first living cells is critically linked to the evolution of the biochemical pathways discussed in this article. It is, thus, proposed here that a primordial cell consisted in a confined environment (primordial cell membrane) that could favor the (re)production of its major constituents, e.g., polynucleotides and polypeptides. Even the limited biochemical interactions between the latter as depicted in the preceding paragraphs were probably sufficient to cross-feed evolutionary advantages accumulated by each synthetic pathway. Instrumental to this had to be the evolution of a primordial cell membrane. Amphipathic lipids, e.g., phospholipids, can self-assemble in an aqueous medium in liposome-like structures, upon minor energy inputs [60]. The diameter of such structures and the organization into lipid bilayers are remarkably close to those of cells and cell membranes, respectively, making this one of the most successful, and least recognized, experimental tests of the evolution of life. Sequestration of fluid within the primordial liposome was a critical step for the evolution of life, since it allowed the local accumulation of products and generated an immensely higher concentration of the components that were no longer free to diffuse away. This also allowed some degree of protection for biochemically labile compounds. An equally important, albeit more subtle, evolutionary advantage of the primordial cell was to constitute a 'selectable unit.' In other words, if an advantageous change occurred, e.g., in the synthesis of primordial RNA, this would have translated into an overall survival advantage for the whole cell. This allowed the accumulation of disparate evolutionarily favorable changes, and further supports the assumption that different biochemical pathways are highly likely to have evolved coherently with each other.

Experimental findings

Several experimental findings support the model proposed, and are listed below.

1) The existence of variants of the genetic code indicates that evolution of the code is possible [31, 61].

2) Chemically similar amino acids are often encoded by similar codons [5, 28, 34]. In the model proposed, similar RNA sequences are more likely to bind polypeptides with similar chemical properties.

3) Relatively short peptides (down at least to 17mers) recognize short specific sequences of double-stranded

RNA or DNA [17–20]. Double-stranded RNA can efficiently bind α-helical polypeptides through both the major [15, 19, 62] and the minor [63, 64] grooves. Finally, short repetitive motifs like the RGG box can recognize RNA with sequence specificity [65].

4) Relatively short RNA can specifically recognize amino acids in a genetic-code-consistent manner [21–23]. Analogously, short specific RNAs can be selected from large numbers of random sequences by selective binding to proteins on a solid substrate [38, 39].

5) Binding to double-stranded nucleic acids can induce α-helical conformation in peptides devoid of secondary structure [18, 19, 66–68]. Consistent with this, RNA can have a chaperone activity [69], i.e., facilitates the folding of polypeptide chains.

6) Small RNAs and oligopeptides can possess the enzymatic properties required by the model presented. Indeed, RNA can catalyze aminoacyl-tRNA synthesis [52], amino acid transfer reactions [50], and esterifications [9]. Ribozymes can have esterase activity for aminoacyl-tRNA bonds [11] and ribosomal RNA shows peptidyl-transferase activity [70]. Conversely, oligopeptides can affect the stability of ester bonds in nucleic acids [16].

7) Covalently bonded nucleic acid/polypeptide complexes have been identified in several different organisms [51].

8) tRNA-like structures probably involved in replication are conserved in the genome [56]. These could be remnants of primitive telomeres and/or replication primers (fig. 2).

9) RNA ligases can originate from random RNA sequences [58]. RNA can also acquire RNA-polymerase-like activity [13].

Experimental testing of the model

A large body of experimental evidence supports the model presented, and the model itself presents experimentally testable predictions.

SELEX-type binding assays can directly test the predicted preferential binding between specific RNA sequences and polypeptides of defined composition [21, 23, 38, 39]. Solid-phase-coupled homo- and heteropolymeric peptides could be used to select for RNAs that bind to specific amino acid sequences. The sequence of the latter will allow the determination of potential correspondences between polypeptides and polynucleotides [21, 22, 35]. This would also allow determination of the subset of codon-amino acids which shows strong stereochemical affinity. The latter may correspond to the 'early set' of codons that were progenitors of later additions (see above).

Recently, transcription factors and DNA have been analyzed at the level of fine structural details, e.g., by cocrystallization [64, 67, 71]. Extensive analysis of the structural determinants of these interactions may further define 'general' rules that would apply to the selective recognition between polymers postulated in the model analyzed. Important characteristics of these reactions that will need to be clarified are the minimum length of the interactive polymers and the discrimination rate between chemically similar structures, e.g., between L and D isomers. This might also shed light on the old controversy of 'genetic takeovers' [72, 73] which, however unlikely they may appear, might correspond to a progressive restriction of molecular possibilities among similar classes of compounds, e.g., choice of one RNA backbone versus another.

The chemistry of spontaneous synthesis of amino acid and base polymers needs to be elucidated. Recent advances indicate the feasibility of this approach [45, 47, 72, 74–79]. Experimental tests should probably include an analysis of condensation reactions within liposome cavities. This may allow determination of the effects of a closed environment on polymerizing reactions, and help to define its consequences on the development of primordial cells.

Acknowledgements. This work was supported by the Italian Association for Cancer Research and the Italian National Research Council, Convenzione CNR-Consorzio Mario Negri Sud.

1 Schimmel P., Giege R., Moras D. and Yokoyama S. (1993) An operational RNA code for amino acids and possible relationship to genetic code. Proc. Natl. Acad. Sci. USA **90:** 8763–8768
2 Maddox J. (1994) The genetic code by numbers. Nature **367:** 111
3 Schimmel P. and Henderson B. (1994) Possible role of aminoacyl-RNA complexes in noncoded peptide synthesis and origin of coded synthesis. Proc. Natl. Acad. Sci. USA **91:** 11283–11286
4 Szathmary E. and Smith J. M. (1995) The major evolutionary transitions. Nature **374:** 227–232
5 Szathmary E. (1993) Coding coenzyme handles: a hypothesis for the origin of the genetic code. Proc. Natl. Acad. Sci. USA **90:** 9916–9920
6 Hornos J. E. and Hornos Y. M. M. (1993) Algebraic model for the evolution of the genetic code. Phys. Rev. Lett. **71:** 4401–4406
7 Cedergren R. and Miramontes P. (1996) The puzzling origin of the genetic code. Trends Biochem. Sci. **21:** 199–200
8 Alberti S. (1997) The origin of the genetic code and protein synthesis. J. Mol. Evol. **45:** 352–358
9 Cech T. R. and Bass B. L. (1986) Biological catalysis by RNA. Annu. Rev. Biochem. **55:** 599–629
10 Cech T. R. (1986) A model for the RNA-catalyzed replication of RNA. Proc. Natl. Acad. Sci. USA **83:** 4360–4363
11 Piccirilli J. A., McConnell T. S., Zaug A. J., Noller H. F. and Cech T. R. (1992) Aminoacyl esterase activity of the *Tetrahymena* ribozyme. Science **256:** 1420–1424
12 Zhang B. and Cech T. R. (1997) Peptide bond formation by in vitro selected ribozymes. Nature **390:** 96–100

13 Been M. D. and Cech T. R. (1988) RNA as an RNA polymerase: net elongation of an RNA primer catalyzed by the *Tetrahymena* ribozyme. Science **239:** 1412–1416
14 Lewin R. (1986) RNA catalysis gives fresh perspective on the origin of life. Science **231:** 545–546
15 Weeks K. M. and Crothers D. M. (1993) Major groove accessibility of RNA. Science **261:** 1574–1577
16 Dillon L. S. (1978) The Genetic Mechanism and the Origin of Life, pp. 211–233, Plenum, New York
17 Poncet D., Laurent S. and Cohen J. (1994) Four nucleotides are the minimal requirement for RNA recognition by rotavirus non-structural protein NSP3. EMBO J. **13:** 4165–4173
18 Talanian R. V., McKnight C. J. and Kim P. S. (1990) Sequence-specific DNA binding by a short peptide dimer. Science **249:** 769–771
19 Tan R., Chen L., Buettner J. A., Hudson D. and Frankel A. D. (1993) RNA recognition by an isolated alpha helix. Cell **73:** 1031–1040
20 Anthony-Cahill S. J., Benfield P. A., Fairman R., Wasserman Z. R., Brenner S. L., Stafford W. F. et al. (1992) Molecular characterization of helix-loop-helix peptides. Science **255:** 979–983
21 Yarus M. (1988) A specific amino acid binding site composed of RNA. Science **240:** 1751–1758
22 Majerfeld I. and Yarus M. (1994) An RNA pocket for an aliphatic hydrophobe. Struct. Biol. **1:** 287–292
23 Famulok M. and Szostak J. W. (1992) Stereospecific recognition of tryptophan agarose by in vitro selected RNA. J. Am. Chem. Soc. **114:** 3990–3991
24 Lewin B. (1994) Genes V, 5th edn, Wiley, New York
25 Woese C. R. (1967) The Genetic Code, Harper & Row, New York
26 Porschke K. M. (1985) Differential effect of amino acid residues on the stability of double helices formed from polyribonucleotides and its possible relation to the evolution of the genetic code. J. Mol. Evol. **21:** 192–198
27 Crick F. H. C., Brenner S., Klug A. and Pieczenik G. (1976) A Speculation on the Origin of Protein Synthesis, pp. 389–397, Reidel, Dordrecht
28 Crick F. H. C. (1968) The origin of the genetic code. J. Mol. Biol. **38:** 367–379
29 Eigen M., Lindemann B. F., Tietze M., Winkler Oswatitsch R., Dress A. and Haeseler A. von (1989) How old is the genetic code? Statistical geometry of tRNA provides an answer. Science **244:** 673–679
30 Schuster L. S. (1981) Prebiotic evolution. In: Biochemical Evolution, pp. 15–87, Gutfreund H. (ed.), Cambridge University Press, Cambridge
31 Stryer L. (1988) Biochemistry, 3rd edn, pp. 652–658, Freeman, New York
32 Jimenez-Sanchez A. (1995) On the origin and evolution of the genetic code. J. Mol. Evol. **41:** 712–716
33 Jungck J. R. (1978) The genetic code as a periodic table. J. Mol. Evol. **11:** 211–224
34 Taylor F. J. and Coates D. (1989) The code within the codons. Biosystems **22:** 177–187
35 Yarus M. (1998) Amino acids as RNA ligands: a direct-RNA-template theory for the code's origin. J. Mol. Evol. **47:** 109–117
36 Yarus M. (1991) An RNA-amino acid complex and the origin of the genetic code. New Biol. **3:** 183–189
37 Schimmel P. and Ribas de Pouplana L. (1995) Transfer RNA: from minihelix to genetic code. Cell **81:** 983–986
38 Tuerk C. and Gold L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science **249:** 505–510
39 Ellington A. D. and Szostak J. W. (1990) In vitro selection of RNA molecules that bind specific ligands. Nature **346:** 818–822
40 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. **30:** 7–15
41 Pace N. R. (1991) Origin of life – facing up to the physical setting. Cell **65:** 531–533
42 Tyagi S. and Ponnamperuma C. (1990) Nonrandomness in prebiotic peptide synthesis. J. Mol. Evol. **30:** 391–399
43 Robertson M. P. and Miller S. L. (1995) Prebiotic synthesis of 5-substituted uracils: a bridge between the RNA world and the DNA-protein world. Science **268:** 702–705
44 Wachtershauser G. (1994) Life in a ligand sphere. Proc. Natl. Acad. Sci. USA **91:** 4283–4287
45 Graaf R. M. de, Visscher J. and Schwartz A. W. (1995) A plausibly prebiotic synthesis of phosphonic acids. Nature **378:** 474–477
46 Ertem G. and Ferris J. P. (1996) Synthesis of RNA oligomers on heterogeneous templates. Nature **379:** 238–240
47 Ferris J. P., Hill A. R. J., Liu R. and Orgel L. E. (1996) Synthesis of long prebiotic oligomers on mineral surfaces. Nature **381:** 59–61
48 Kiedrowski G. von (1996) Primordial soup or crepes? Nature **381:** 20–21
49 Keller M., Blochl E., Wachtershauser G. and Stetter K. O. (1994) Formation of amide bonds without a condensation agent and implications for origin of life. Nature **368:** 836–838
50 Lohse P. A. and Szostak J. A. (1996) Ribozyme-catalysed amino-acid transfer reactions. Nature **381:** 442–444
51 Drygin Y. F. (1998) Natural covalent complexes of nucleic acids and proteins: some comments on practice and theory on the path from well-known complexes to new ones. Nucleic Acids Res. **26:** 4791–4796
52 Illangasekare M., Sanchez G., Nickles T. and Yarus M. (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. Science **267:** 643–647
53 Purohit P. and Stern S. (1994) Interactions of a small RNA with antibiotic and RNA ligands of the 30S subunit. Nature **370:** 659–662
54 Hirao I. and Ellington A. D. (1996) Re-creating the RNA world. Curr. Biol. **5:** 1017–1022
55 Lazcano A. and Miller S. L. (1996) The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. Cell **85:** 793–798
56 Maizels N. and Weiner A. M. (1994) Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. Proc. Natl. Acad. Sci. USA **91:** 6729–6734
57 Blasco M. A., Lee H. W., Hande M. P., Samper E., Lansdorp P. M., DePinho R. A. et al. (1997) Telomere shortening and tumor formation by mouse cells lacking telomerase RNA. Cell **91:** 25–34
58 Ekland E. H., Szostak J. W. and Bartel D. P. (1995) Structurally complex and highly active RNA ligases derived from random RNA sequences. Science **269:** 364–370
59 Orgel L. E. (1989) The origin of polynucleotide-directed protein synthesis. J. Mol. Evol. **29:** 465–474
60 O'Brien D. F., Klingbiel R. T., Specht D. P. and Tyminski P. N. (1985) Preparation and characterization of polymerized liposomes. Ann. NY Acad. Sci. **446:** 282–295
61 Caron F. (1990) Eucaryotic codes. Experientia **46:** 1106–1117
62 Chen L. and Frankel A. D. (1995) A peptide interaction in the major groove of RNA resembles protein interactions in the minor groove of DNA. Proc. Natl. Acad. Sci. USA **92:** 5077–5081
63 Freire R., Salas M. and Hermoso J. M. (1994) A new protein domain for binding to DNA through the minor groove. EMBO J. **13:** 4353–4360
64 Schumacher M. A., Choi K. Y., Zalkin H. and Brennan R. G. (1994) Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science **266:** 763–770
65 Burd C. G. and Dreyfuss G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. Science **265:** 615–621
66 Spolar R. S. and Record M. T. J. (1994) Coupling of local folding to site-specific binding of proteins to DNA. Science **263:** 777–784

67 Weiss M. A., Ellenberger T., Wobbe C. R., Lee J. P., Harrison S. C. and Struhl K. (1990) Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. Nature **347:** 575–578

68 Johnson N. P., Lindstrom J., Baase W. A. and Hippel P. H. von (1994) Double-stranded DNA templates can induce alpha-helical conformation in peptides containing lysine and alanine: functional implications for leucine zipper and helix-loop-helix transcription factors. Proc. Natl. Acad. Sci. USA **91:** 4840–4844

69 Huang Z. S. and Wu H. N. (1998) Identification and characterization of the RNA chaperone activity of hepatitis delta antigen peptides. J. Biol. Chem. **273:** 26455–26461

70 Noller H. F., Hoffarth V. and Zimniak L. (1992) Unusual resistance of peptidyl transferase to protein extraction procedures. Science **256:** 1416–1419

71 Somers W. S. and Phillips S. E. V. (1992) Crystal structure of the met repressor-operator complex at 2.8 A revolution reveals DNA recognition by β-strands. Nature **359:** 387–394

72 Eschenmoser A. (1994) Chemistry of potentially prebiological natural products. Origins Life Evol. Biosphere **24:** 389–423

73 Joyce G. F., Schwartz A. W., Miller S. L. and Orgel L. E. (1987) The case for an ancestral genetic system involving simple analogues of the nucleotides. Proc. Natl. Acad. Sci. USA **84:** 4398–4402

74 Joyce G. F. and Orgel L. E. (1988) Non-enzymatic template-directed synthesis on RNA random copolymers: poly(C,A) templates. J. Mol. Biol. **202:** 677–681

75 Service R. F. (1998) A biomolecule building block from vents. Science **281:** 1936–1937

76 Lewis R. (1997) Primordial soup researchers gather at watering hole. Science **277:** 1034–1035

77 Tarasow T. M., Tarasow S. L. and Eaton B. E. (1997) RNA-catalysed carbon-carbon bond formation. Nature **389:** 54–57

78 Unrau P. J. and Bartel D. P. (1998) RNA-catalysed nucleotide synthesis. Nature **395:** 260–263

79 Keefe A. D. and Miller L. (1995) Are polyphosphate or phosphate esters prebiotic reagents? J. Mol. Evol. **41:** 693–702