6 The RNA World: Hypotheses, Facts and Experimental Results

Marie-Christine Maurel, Anne-Lise Haenni

A biochemical world that would have existed before the contemporary DNA-RNA-protein world, and named in 1986 “The RNA World” by Walter Gilbert (Gilbert, 1986), such a world had already been proposed during the preceding decades by Carl Woese, Francis Crick and Leslie Orgel (Woese, 1965; Crick, 1968; Orgel, 1968).

By demonstrating the remarkable diversity of the RNA molecule, molecular biology proved these predictions. RNA, present in all living cells, performs structural and metabolic functions many of which were unsuspected only a few years ago. A truly modern “RNA world” exists in each cell; it contains RNAs in various forms, short and long fragments, single and double-stranded, endowed with multiple roles (informational, catalytic, that can serve as templates, guides, defense, etc.), certain molecules even being capable of carrying out several of these functions.

Are the sources of this RNA world to be found in the bygone living world?

6.1 The Modern RNA World

6.1.1 Where in the Living Cell is RNA Found?

Synthesized (transcribed) in the nucleus, mature messenger RNAs (mRNAs), transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) are exported as single strands to the cytoplasm of the cell after various maturation steps. A ribonucleic acid (RNA) is formed by linking nucleotides\(^1\), themselves composed of heterocyclic bases associated with a sugar, β-D-ribofuranose, and a phosphate molecule (phosphoric acid). The four main nucleotides contain the heterocyclic purine (adenine and guanine) or pyrimidine (cytosine and uracil) bases\(^2\). However, RNAs, in particular rRNAs and tRNAs contain a very large diversity of modified nucleotides, since more that a hundred modified nucleotides\(^3\) have now been identified in these two classes of molecules (Grosjean and Benne, 1998).

---

\(^{1}\) To yield a polyribonucleotide

\(^{2}\) Adenine, A; guanine, G; cytosine, C; uracil, U

\(^{3}\) Post-transcriptional modifications
RNAs are usually single stranded\(^4\). Nevertheless, these strands can base pair locally or over long stretches (intramolecular pairing). Finally, from a structural point of view, they contain a reactive hydroxyl group in the 2' position of ribose (a group that is absent in DNA). The stacking forces and pairing of bases produce “stems and helices”; defined structures bring together the helices and the regions separating them, into “motifs”.

RNA helices: Through the action of the stacking forces, the skeleton of the single strand by itself tends to take the shape of a simple, right-handed and irregular helix. However, the important conformation is the double helix composed of two strands of RNA or of RNA/DNA (hydrbids formed transiently during transcription) or that occurs when two distantly located complementary segments of the same RNA base pair.

The motifs identified are bulges, elbows, or loops.

Hairpins are other important structural motifs related to certain functions of RNAs. They can lead to interactions with special sequences, such as the GNRA loops\(^5\), seven-base-long loops, etc. Large RNAs possess independent domains formed by the arrangement of a certain number of motifs. An RNA molecule can adopt several reversible conformations, depending on the presence of ions, specific surfaces or bound ligands. RNAs possess a repertory of structures reminiscent of proteins (motifs or domains) allowing them to express certain functions such as catalysis. Finally, non-Watson–Crick base pairs\(^6\) are frequently encountered in RNAs (G-U pairs are common) and modified bases are involved, and by their strong steric hindrance with the bases, the 2' OH groups of the ribose moieties tend to prevent folding in the B helical conformation\(^7\).

6.1.1.1 The Three Large Classes of RNA

- Messenger RNAs (mRNAs of 400 to 6000 nucleotides) are the copy of DNA genes\(^8\). The RNA transcripts are considerably modified in the nucleus during maturation, and during transcription of DNA into RNA, short hybrids of the A conformation appear. Their life is short in prokaryotes (a few minutes to a few dozen minutes) and can be of several hours in higher eukaryotes; mRNAs correspond to only a few per cent of the total cellular RNAs. The step-by-step decoding of the mRNA by the ribosome known as translation is regulated by specific proteins, and in some cases also by hairpin motifs and/or by pseudoknots.

---

\(^4\) Paired two-stranded RNAs are exceptions found in a few rare viruses

\(^5\) N is any nucleotide, R is a purine nucleotide

\(^6\) See glossary. Watson–Crick pairings are the standard pairs (A-U and G-C)

\(^7\) The bends they impose to the plane of the bases – of about 20° – on the axis results in a structure resembling the A conformation (also designated RNA 11 to stress the 11 base pairs per turn). The A form of RNA double helices is characterized by 11 base pairs per helical turn (instead of 10 for the B form), and by bending of the base pairs by 16°/helical axis (instead of 20° for DNA A)

\(^8\) A gene is a fragment of DNA whose information is expressed via the genetic code
Pseudoknots result from base pairing between nucleotides within a loop and complementary nucleotides outside of the loop.

- Transfer RNAs (tRNAs) are small molecules whose maximum length is about 100 nucleotides. They are strongly conserved and are involved in the central metabolism of all types of cells. Their main function is to ensure the interaction between the codon presented by the mRNA and the specific amino acid (corresponding to this codon) and contained in the anticodon of the aminoacyl-tRNA. tRNAs possess two extremely specific sites: the first is the sequence CCA located at the 3' OH of the molecule; the second site is located in a loop that contains the anticodon. The cloverleaf-shaped secondary structure (Fig. 6.1) possesses several motifs. tRNAs also serve as primers during replication of certain viruses and are involved in the activity of telomerases. Synthesized as pre-tRNAs they undergo a maturation step during which RNase P cleaves off a short fragment from the 5' end of the RNA (Guerrier-Takada et al., 1983). As already mentioned, tRNAs contain a large number of modified bases that are probably the most visible “relics” of an ancient RNA world (Cermakian and Cedergren, 1998).

![Fig. 6.1. Secondary cloverleaf structure of a tRNA. Arrows indicate number of nucleotides in the loop, stem and bulge](Image)
The size of the ribosomal RNAs (rRNAs) is variable, from 120 to 4718 nucleotides. rRNAs are located in the ribosome, the site of protein synthesis. In addition to about fifty proteins, the prokaryotic ribosome contains three rRNAs and the eukaryotic ribosome four rRNAs. The rRNAs are methylated (sometimes in the 2’OH position of the ribose, protecting the polymer from hydrolysis). Their typical secondary structure is remarkably well conserved (Fig. 6.2).

They possess complex global tertiary conformations that compact the molecule into different domains, and it has now been clearly demonstrated that the rRNA catalyzes the formation of the peptide bond during protein biosynthesis (Ban et al., 2000; Nissen et al., 2000; Ysupov et al., 2001).

![Fig. 6.2. Typical secondary structure 1) 16S rRNA of the bacterium Escherichia coli, 2) 18S rRNA of the yeast Saccharomyces cerevisiae](image)

### 6.1.1.2 Noncoding RNAs (ncRNAs)

In addition to rRNAs, tRNAs and mRNAs a variety of RNA molecules have been discovered that possess very diverse functions in the living cell (Maurel, 1992; Meli et al., 2001; Zamore, 2002; Grosshans and Slack, 2002; Westhof, 2002). Before involvement of the ribosome, the RNA transcripts must undergo maturation steps. In eukaryotes, these post-transcriptional modification steps require the participation of small RNAs, the snoRNAs (small nucleolar RNAs) that together with proteins, form the snoRNP (small nucleolar ribonucleoprotein particles). Over 150 snoRNPs have been described in eukaryotes (in different lineages). They form a snoRNP complex, the snorposome, that participates in RNA maturation. The origin of the modification systems is still unknown. One of the various hypotheses put forward suggests that the snoRNAs of the RNA world would have been involved in the assembly of the protoribosomes, and more generally in the scaffolding of ribozymes (Terns and Terns, 2002).
Moreover, large snRNPs (small nuclear ribonucleoprotein particles) responsible for intron excision from pre-mRNAs have been identified. Each snRNP is composed of snRNA and about a dozen snRNP proteins. Two classes of such spliceosomes cleave different introns, whereas excision and ligation of the exons is achieved by the same biochemical mechanism (Tarn and Steitz, 1997). Spliceosomes are restricted to eukaryotes, even though containing introns bacteria have been reported.

The telomerase is an enzyme that uses a small RNA as primer during replication to elongate the linear DNA located at the end of eukaryotic chromosomes (Maizels et al., 1999).

Vault RNAs are ribonucleoprotein particles located in the cytoplasm of eukaryotes (Kong et al., 2000). They are associated with the nuclear “pore complex”; their function has not been clearly defined, but their structure suggests that they may be involved in cell transport or in the assembly of macromolecules. The history of the evolution of vault RNAs remains unknown, but these RNAs could have participated in primitive compartmentation.

Finally, an RNA-protein complex, the SrpRNA (signal recognition particle RNA) is highly conserved in the three kingdoms (Wild et al., 2002). It is involved in translation, and during secretion of proteins from the plasma membrane or from the endoplasmic reticulum.

About 15 years ago, the existence of a correcting mechanism, “editing”, was demonstrated (Lamond, 1988). This co- or post-transcriptional mechanism modifies the sequence of the mRNA by the insertion or deletion of nucleotides, or by the modification of bases. Up to 55% modifications can take place with respect to the gene (in this case it is designated “cryptogene”). The sites where editing takes place are determined by the structure of the RNA, or by guide-RNAs (Stuart and Panigrahi, 2002). In kinetoplastid protozoa, guide RNAs are required to edit mitochondrial pre-mRNAs by inserting or deleting uridy late residues in precise sites (Kable et al., 1997).

Finally, the tmRNA (transfer-messenger RNA) is a stable cytoplasmic RNA found in eubacteria. tmRNAs contain a tRNA\textsuperscript{Ala}-like structure (with pairing between the 5' and 3' ends) and an internal reading frame that codes for a short peptide (peptide tag) (Fig. 6.3). It is thus at variance with the strict definition of snRNAs, since it encompasses a short reading frame. It performs a new type of recently discovered translation, known as trans-translation, during which a peptide is synthesized starting from two distinct mRNAs. tmRNA acts as tRNA and as mRNA to “help” ribosomes that are blocked on a truncated mRNA lacking a termination codon. tmRNA participates by adding alanine to the growing peptide chain. Thus, tmRNA plays a dual role: as tRNA\textsuperscript{Ala} it can be aminoacylated by the corresponding alanyl-tRNA synthetase, and as mRNA its open reading frame can be translated by the ribosome (Withey and Friedman, 2002; Valle et al., 2003). Could tmRNA be a bacterial adaptation, or could it have been lost by the archaeb and the eukaryae?
A eukaryotic system distantly related to tmRNA has recently been described (Barends et al., 2003) in the single-stranded Turnip yellow mosaic virus (TYMV) RNA. The 3’ end of the viral genome harbors a tRNA-like structure that is indispensable for the virus viability and can be valylated. During protein biosynthesis programmed with valylated TYMV RNA, the valine residue is N-terminally incorporated into the viral polyprotein, thereby introducing a novel mechanism of initiating protein synthesis (Fig. 6.4). Here again, the viral RNA would be bifunctional, serving both as tRNA and as mRNA.

It will be interesting to determine whether other viral RNAs whose 3’ end bears an aminoacylatable tRNA-like structure (Fechter et al., 2001) can also donate their amino acid for mRNA translation.

Viroids are subviral plant pathogens responsible for economically important diseases. They are small (246–401 nucleotides), single-stranded closed circular RNA molecules characterized by a highly compact secondary structure. They are devoid of coding capacity and replicate autonomously in the plant host. Two families of viroids have been characterized, the Pospiviroidae (type-member:
Fig. 6.4. Model of the tRNA-like structure-mediated internal initiation mechanism of TYMV RNA for polyprotein translation. I: Coat protein gene; II: Polyprotein gene; III: Movement protein gene; Adapted from Barends et al., 2003

Potato spindle tuber viroid, PSTVd) that replicates in the nucleus, and the Avsunviroidae (type-member: Avocado sun blotch viroid, ASBVd) that replicates in chloroplasts and possesses conserved hammerhead structures in the viroid and in the complementary RNA orientation. It has been suggested that the presence of hammerhead structures could reflect the early appearance of viroids in the course of evolution; they could correspond to “living fossils” of the primitive RNA world (Diener, 2001).

The few ncRNAs described here are probably only the tip of a huge iceberg (Bachellerie et al., 2002) since most of the transcriptional output of superior eukaryotes is nonprotein coding (97% for human). These ncRNAs could constitute a real RNA world in complex organisms (Eddy, 2001; Mattick, 2003). Their study may open new perspectives about the importance of RNA in primitive life. Certain RNAs that are presently being investigated are those involved in RNA interference (RNAi): the RNAs responsible for RNAi are the small interfering RNAs that target and cleave mRNAs (Nykanen et al., 2001). Micro-RNAs, another class of small RNAs, are involved in translation regulation (Grosshans and Slack, 2002). In eukaryotes, guide snoRNAs participate in selecting the sites on rRNAs that undergo modifications such as $\Psi$ formation or 2'-O-methylation (Lafontaine and Tollervey, 1998).

6.2 An RNA World at the Origin of Life?

The scenario of evolution postulates that an ancestral molecular world existed originally that was common to all the present forms of life; the functional properties of nucleic acids and proteins as we see them today would have been produced by molecules of ribonucleic acids (Joyce, 1989; Orgel, 1989; Benner et al., 1989, 1993; Joyce and Orgel, 1999; Gesteland et al., 1999; Bartel and Unrau, 1999; McGinness et al., 2002; Joyce, 2002).
### 6.2.1 Facts

As we have seen, RNAs occupy a pivotal role in the cell metabolism of all living organisms and several biochemical observations resulting from the study of contemporary metabolism should be stressed. For instance, throughout its life cycle, the cell produces deoxyribonucleotides required for the synthesis of DNA that derive from ribonucleotides of the RNA. Thymine, a DNA specific base is obtained by transformation (methylation) of uracil a RNA specific base, and RNAs serve as obligatory primers during DNA synthesis (Fig. 6.5). Finally, the demonstration that RNAs act as catalysts is an additional argument in favour of the presence of RNAs before DNA during evolution.

![Diagram](image)

**Fig. 6.5.** Facts in favour of an RNA world. (a) Synthesis of deoxyribonucleotide; (b) Structure of uracil and thymine; (c) DNA synthesis primed by RNA

### 6.2.2 Hypotheses

DNA replication triggered by ribonucleotide primers can be considered as a modified transcription process during which polymerisation of RNA is “replaced” by that of DNA. In addition, DNA a double-stranded molecule lacking a hydroxyl
group in 2’ of the desoxyfuranose, appears more stable than RNA. Therefore it seems highly likely that RNA arose before DNA during biochemical evolution, and for this reason DNA is sometimes considered as modified RNA better suited for the conservation of genetic information. This genetic privilege would constitute a logical step in an evolutionary process during which other molecules could have preceded RNA and transmitted genetic information.

The idea of an “RNA” world rests primarily on three fundamental hypotheses, developed by Joyce and Orgel (1999):

- during a certain period in evolution, genetic continuity was assured by RNA replication,
- replication was based on Watson–Crick type base pairing,
- genetically coded proteins were not involved in catalysis.

6.2.3 But What do We Know about Primitive Replication?

Synthesis of a strand complementary of the template was studied extremely thoroughly in vitro in the group of Orgel (Inoue and Orgel, 1983; Joyce and Orgel, 1986; Orgel, 1992). During this directed synthesis, the mononucleotides (activated under the form of 5′-phosphorimidazolides) are positioned according to the Watson and Crick pairing rules along a preformed polypyrimidine template. Since these monomers are activated, they can bind together to form the complementary strand (Fig. 6.6). Orgel and his coworkers showed that starting from activated monomers, it is possible in certain conditions to copy a large number of oligonucleotide sequences containing one or two different nucleotides in the absence of enzyme (Hill et al., 1993).

Ferris and his coworkers spent some 15 years studying the assembly of RNA oligomers on the surface of montmorillonite (clay of Montmorillon in the Vienne region in France) (Ferris, 1987; Ferris and Ertem, 1992). The monomers used, nucleoside 5′-phosphorimidazolides, were probably not prebiotic molecules. Nevertheless, experimental results demonstrated that minerals that serve as adsorbing surfaces and as catalysts (Paecht-Horowitz et al., 1970; Ferris et al., 1996), can lead to accumulation of long oligonucleotides, as soon as activated monomers are available. One can thus envisage that activated mononucleotides assembled into oligomers on the montmorillonite surface or on an equivalent mineral surface. The longest strands serving as templates, direct synthesis of a complementary strand starting from monomers or short oligomers, and double-stranded RNA molecules accumulate. Finally, a double RNA helix of which one strand is endowed with RNA polymerase activity, would dissociate to copy the complementary strand to produce a second polymerase that would copy the first to produce a second complementary strand, and so forth. The RNA world would thus have emerged from a mixture of activated nucleotides. However, a mixture of activated nucleotides would need to have been available! In addition, this nucleotide chemistry is restricted in another way, since a copy of the template can be started only if the nucleotides are homochirals (Joyce et al., 1987).
Finally, when either the first replicative molecule, the template or one of its elements (nucleotides) is to be synthesized from the original building blocks in particular the sugars that are constituents of nucleotides, a number of difficulties are encountered (Sutherland and Whitfield, 1997). Synthesis of the sugars from formaldehyde produces a complex mixture in which ribose is in low amounts. On the other hand, production of a nucleoside from a base and a sugar leads to numerous isomers, and no synthesis of pyrimidine nucleosides has so far been achieved in prebiotic conditions. Finally, phosphorylation of nucleosides also tends to produce complex mixtures (Ferris, 1987). Onset of nucleic acid replication is almost inconceivable if one does not envisage a simpler mechanism for the prebiotic synthesis of nucleotides. Eschenmoser succeeded in producing 2,4-diphosphate ribose during a potentially prebiotic reaction between glycol...
aldehyde\textsuperscript{9} monophosphate and formaldehyde (Eschenmoser, 1999). It is thus possible that direct prebiotic nucleotide synthesis can occur by an alternative chemical pathway. Nevertheless, it is more likely that a certain organized form of chemistry preceded the RNA world, hence the notion of “genetic take-over”. Since the ribose-phosphate skeleton is theoretically not indispensable for the transfer of genetic information, it is logical to propose that a simpler replication system would have appeared before the RNA molecule.

6.3 A Pre-RNA World

6.3.1 Evolutive Usurpation

During the evolutionary process, a first genetic inorganic material, would have been replaced by organic material. The hypothesis of a precursor of nucleic acid\textsuperscript{10} (Cairns-Smith, 1966, 1982) is a relatively ancient idea, but it is only within the last few years that research has been oriented towards the study of molecules simpler than present-day RNAs, yet capable of autoreplication. Models with predictably retroactive activities can thus be tested experimentally.

6.3.2 Alternative Genetic Systems

In the peptide nucleic acids (PNA) of Nielsen and coworkers, the ribofuranose-phosphate skeleton is replaced by a polyamidic skeleton on which purine and pyrimidine bases are grafted (Fig. 6.7). PNAs form very stable double helices with an RNA or a complementary DNA (Egholm et al., 1993) and can serve as template for the synthesis of RNA, or vice versa (Schmidt et al., 1997). PNA-DNA chimeras containing two types of monomers have been produced on DNA or PNA templates (Koppitz et al., 1998). The information can be transferred from PNAs (achiral monomers) to RNA during directed synthesis; the double-helical molecule with a single complementary RNA strand is stable. Transition from a “PNA world” to an “RNA world” is hence possible. Nevertheless, the formation of oligomers from PNA monomers seems particularly difficult in prebiotic conditions.

Eschenmoser (1994) explored the properties of nucleic acid analogues in which ribofuranose is replaced by one of its isomers, ribopyranose (Furanose, 5-membered ring; pyranose, 6-membered ring). p-RNAs (pyranosyl RNAs) (Fig. 6.7) form more stable double helices (with Watson–Crick pairings) than RNA with ribofuranose. In addition, the double helices of p-RNA wind and unwind more easily than those formed with standard nucleic acids, and this should facilitate their separation during replication. p-RNAs could therefore constitute good candidates as precursor genetic systems, but a p-RNA strand cannot pair

\textsuperscript{9} Recently shown to exist in interstellar clouds and comets (Cooper et al., 2001)

\textsuperscript{10} This is the idea of genetic take-over developed by Cairns-Smith in the 1960s
with an RNA of complementary sequence, and this makes it difficult to imagine a transition from p-RNA to RNA.

The group of Eschenmoser recently replaced the ribose moiety by a four-carbon sugar, threose, whose prebiotic synthesis seems easier. The resulting oligonucleotides designated TNAs, \((3' \rightarrow 2')-\alpha\text{-L-threose nucleic acid}\) (Fig. 6.8), can form a double helix with RNA (Schönig et al., 2000). TNA is capable of antiparallel, Watson–Crick pairing with complementary DNA, RNA and TNA oligonucleotides. Furthermore, Szostak and his collaborators have recently found that certain DNA polymerases can copy limited stretches of a TNA template, despite significant differences in the sugar-phosphate backbone, (Chaput and Szostak, 2003).

Finally, hexitol nucleic acids (HNA) (Fig. 6.7), whose skeleton is composed of 1, 5-anhydrohexitol (six-membered cyclic hexitol) and their isomers altritol nucleic acids (ANA), form stable duplexes with complementary oligonucleotides, and are very efficient templates since they favour assembly of a complementary strand during directed synthesis (Kozlov et al., 1999a, 1999b, 2000). The shape of the duplexes formed is reminiscent of that of DNA in the A form. Double-helical DNA is mainly in the B form\(^{11}\), whereas the double helices of RNA in the DNA-RNA hybrids adopt the A form\(^{12}\). Kozlov et al. (1999c) have demonstrated that

---

\(^{11}\) d-ribose in the 2′-endo form. The characteristics of these forms are indicated above in the text.

\(^{12}\) In this case, the sugar is in the 3′-endo conformation. In the A form of RNA double helices, there are 11 base pairs per helical turn (instead of 10 for the B form); the inclination of the base pairs is 16°/helical axis (20° for DNA A).
the more the template is in the A form, the better the efficiency of directed synthesis. Based on these studies one can imagine an entire series of templates that would supply the “good” structural preorganization. Furthermore, these same authors have shown that RNA partially preorganized in the A form, is a more efficient matrix than single-stranded DNA. Finally, whatever the precursor skeleton adapted to the formation of stable duplexes may have been, the bond at the mineral surface could have imposed the necessary geometrical constraints: yet this still remains to be experimentally demonstrated.

This leads us to two major conclusions, namely that on the one hand a transition may have occurred between two different systems without loss of information, and that on the other hand the HNA and ANA nucleic acids are very efficient templates. Even if it is difficult to imagine prebiotic synthesis of these molecules, they are good model systems that show the importance of a necessary structural preorganization for directed synthesis by a template.

From the point of view of evolution, the studies described previously demonstrate that other molecules capable of transmitting hereditary information may have preceded our present day nucleic acids. This is what Cairns-Smith coined the “take-over” (Cairns-Smith, 1982), the evolutionary encroachment or genetic take-over, or to some extent what François Jacob (1970) calls genetic tinkering, in other words, making new material from the old. This also sheds light on the precision with which the various elements or processes progressively adjusted themselves, thanks to successive trials and errors.

6.4 Optimizing the Functional Capacities of Ribonucleic Acids

6.4.1 Coenzymes and Modified Nucleosides

The nucleotides that by post-transcriptional modification can today acquire the majority of functional groups present in amino acids, possess a great potential
diversity that is expressed at the level of ribonucleotide coenzymes (several coenzymes derive from AMP), and of the modified bases of tRNAs (Fig. 6.9). The role of cofactors at all steps of the metabolism and their distribution within the three kingdoms suggest that a great variety of nucleotides was present in the ancestor common to all forms of life.

Several authors have underscored the possible presence of coenzymes before the appearance of the translation machinery (White, 1976). Proteins would have appeared only at a later stage, coenzymes and ribozymes being fossil traces of past catalysts. Indeed, in the living cell, only a minority of enzymes function without coenzyme; they are mostly hydrolases, and apart from this group, 70% of the enzymes require a coenzyme. If metal coenzymes involved in catalysis are considered, the number of enzymes that depend on coenzymes increases further. Present-day coenzymes, indispensable cofactors for many proteins, would be living fossils of primitive metabolism catalysts.

Most coenzymes are nucleotides (NAD, NADP, FAD, coenzyme A, ATP, etc.) or contain heterocyclic nitrogen bases that can originate from nucleotides (thiamine pyrophosphate, tetrahydrofolate, pyridoxal phosphate, etc.).

Coenzymes would be vestiges of catalytic nucleic enzymes that preceded ribosomal protein synthesis, and tRNAs can be viewed as large coenzymes participating in the transfer of amino acids. It is even possible to consider that catalytic groups that were part of nucleic enzymes were incorporated in specific amino acids rather than being “retained” as coenzymes. This could be the case of imidazole, the functional group of histidine, whose present synthesis in the cell is triggered by a nucleotide.

<table>
<thead>
<tr>
<th>Coenzyme</th>
<th>R</th>
<th>R'</th>
<th>R''</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activated methionine</td>
<td>methionine</td>
<td>H</td>
<td>H</td>
<td>0</td>
</tr>
<tr>
<td>Amino acid adenylate</td>
<td>amino acid</td>
<td>H</td>
<td>H</td>
<td>1</td>
</tr>
<tr>
<td>Activated sulfate</td>
<td>SO$_3^{2-}$</td>
<td>H</td>
<td>PO$_3^{2-}$</td>
<td>1</td>
</tr>
<tr>
<td>Cyclic 3'-5' AMP</td>
<td>H</td>
<td>H</td>
<td>PO$_3^{2-}$</td>
<td>1</td>
</tr>
<tr>
<td>NAD</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>2</td>
</tr>
<tr>
<td>NADP</td>
<td>PO$_3^{2-}$</td>
<td>H</td>
<td>H</td>
<td>2</td>
</tr>
<tr>
<td>FAD</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>2</td>
</tr>
<tr>
<td>CoA-SH</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 6.9. List of coenzymes derived from AMP
The modified nucleosides present today in RNAs result from post-transcriptional modifications. Nevertheless, modified nucleosides could have been present in the primitive world and their distribution would have become established in the RNAs of the three living kingdoms (Cermakian and Cedergren, 1998).

Our working hypothesis is based on the demonstration of esterase activity in a nucleoside analogue N⁶-ribosyladenine (Fuller et al., 1972; Maurel and Ninio, 1987). This activity, which is due to the presence of an imidazole group that is free and available for catalysis, is comparable to that of histidine placed in the same conditions (Fig. 6.10). We have studied the kinetic behaviour of this type of catalyst (Ricard et al., 1996) and have shown that the catalytic effect increases greatly when the catalytic element, pseudohistidine, is placed in a favourable environment within a macromolecule (Décout et al., 1995). Moreover, primitive nucleotides were not necessarily restricted to the standard nucleotides encountered today, and because of their replicative and catalytic properties, the N⁶ and N³ substituted derivatives of purines could have constituted essential links between the nucleic acid world and the protein world.

6.4.2 The Case of Adenine

Purine nucleotides, and in particular those containing adenine, participate in a large variety of cellular biochemical processes (Maurel and Décout, 1999).
Their best-known function is that of monomeric precursors of RNAs and DNAs. Nevertheless, derivatives of adenine are universal cofactors. They serve in biological systems as a source of energy (ATP), allosteric regulators of enzymatic activity and regulation signals (cyclic AMP). They are also found as acceptors during oxidative phosphorylation (ADP), as components of coenzymes (such as in FAD, NAD, NADP, coenzyme A), as transfer agents of methyl groups and of S-adenosylmethionine, as possible precursors of polyprenoids in C5 (adenosylhopane) (Neunlist et al., 1987), and – last but not least – adenine 2451 conserved within the large rRNA in the three kingdoms, would be involved in acid-base catalysis during the formation of the peptide bond (Muth et al., 2000). However, this role of adenine has been refuted based on mutagenesis studies and phylogenetic comparisons (Muth et al. 2001; Green and Lorsch, 2002).

On the other hand, biosynthesis of an amino acid, histidine, that would have appeared late in evolution, begins with 5-phosphoribosyl-1-phosphate that forms N’-(5-phosphoribosyl)-ATP by condensation with ATP. This reaction is akin to the initial reaction of purine biosynthesis. Finally, the ease with which purine bases are formed in prebiotic conditions suggests that these bases were probably essential components of an early genetic system. The first genetic system was probably capable of forming base pairs of the Watson–Crick type, Hoogsteen and other atypical associations, by hydrogen bonds as they still appear today in RNA. It probably contained a different skeleton from that of RNA, and no doubt also modified bases, thereby adding chemical functions, but also hydrophobic groups, and functions such as amine, thiol, imidazole, etc. Wächterhäuser (1988) also suggested novel pairings of the purine–purine type.

Originally, the principle probably rested on forced cooperation of genetic and functional components, rather than on selection by individual competition. It may have first entailed testing and improvements (learning by trial and error) of the informational content of the genes, i.e. linking the genotype (sequence) to the phenotype (shape). One can consider that in such a system the unforeseen was faced, so that the living organism would need to adapt favourably and rapidly.

### 6.4.3 Mimicking Darwinian Evolution

Most of the “rational” biochemical approaches consist of deducing the active sequence of a nucleic acid or protein from a primary sequence, or in synthesizing a defined compound by modelling and structural analysis. However, “real life”, that of our ancestors as also that of our cells, does not proceed in this manner. The hunter-gatherers of prehistory survived only thanks to their extraordinary capacity to recognize objects. In addition, survival of a population in a new environment is often linked to the appearance of a few variants to which random mutations conferred the power to adapt and exploit the new situation to their advantage. Combinatorial methods, by modelling these observations, have now

---

13 Purines have also been found in the Murchison meteorite
14 And by giving access to many related molecules that can be sorted
become the alternative to the rational concept. Selection in vitro requires no information concerning the sequence of the molecules, and replaces the pre-established adjustments between the molecule and its target. What is needed, is to mimic the processes of evolution at the molecular level.

Indeed, it has been known since the experiments of Spiegelman (1971) and his colleagues (Kramer et al., 1974) that populations of different molecules capable of reproducing themselves in a hereditary manner, can evolve and adapt to an appropriate environment. Spiegelman, the inventor of non-natural selection indeed demonstrated in the 1960s, that RNA populations can evolve when they replicate with the help of an enzyme, the replicase of the bacteriophage Qβ. A population of macromolecules can thus comply with the prerequisites of Darwinian theory, and must find a form adapted to recognition of the target in a sufficiently rich population. Coexistence in the same entity of shape and sequence, can favour the emergence of favourable candidates by means of a selection step (linked to the shape) and an amplification step (linked to the sequence) at the end of this molecular evolutionary process. A selection of this type could have occurred during early molecular evolution, some 3.5 billion years ago.

The original polymers more or less related to RNA and formed in the primitive world must have randomly contained the A, U, G and C bases. There are over one million possible sequences for a decanucleotide composed of 10 monomers A, U, G, C, and over $10^{12}$ sequences for a polynucleotide of 20 monomers\textsubscript{15}. Nature does not appear to have exploited all the possible combinations before having reached the remarkable functional unity of the living world, and given the immense number of possibilities it is also useless to try to explore experimentally, one by one, all the potentially functional sequences.

The SELEX method (systematic evolution of ligands by exponential enrichment) (Tuerk and Gold, 1990) is an efficient, quasi automatic method based on repeated cycles of reproductive selection of those individuals that are best adapted to a given function. Established in the 1990s, this method makes it possible to obtain new structures, aptamers, selected through their aptitude to recognize other molecules (Ellington and Szostak, 1990). Aptamers are capable to recognize targets as small as metal ions, or as large as cells. They can interact with a great variety of molecules that are important for primitive metabolism, like amino acids, porphyrines, nucleotide factors, coenzymes, small peptides and short oligonucleotides (Illangasekare and Yarus, 1997; Jadhav and Yarus, 2002; Joyce, 2002; McGinness et al., 2002; Reader and Joyce, 2002).

At the molecular level, the Darwinian behaviour requires that a method of selection (RNA-aptamers), of amplification of selected species, and of mutations (introduction of variants in the population by means of mutations) be established. Through several cycles of selection, amplification and mutations,

\textsuperscript{15} For a nucleic acid of 200 nucleotides, $10^{120}$ different sequences are theoretically possible, and for a small protein containing 200 amino acids, $10^{280}$ arrangements are possible! Which also applies to the protein world (phage display and combinatorial synthesis of peptides)
populations of molecules are “pushed” to evolve towards novel properties. The molecules presenting the best “aptitudes” are selected and a new generation will thereby come out. Evolutionary processes performed experimentally thus make it possible for molecules to emerge that have not yet been produced by Nature, or allow the re-emergence of precursor molecules that have strongly diverged or naturally disappeared.

In practice, how does one proceed? A “library” of oligonucleotides is a conformational population containing at least one particular conformation able to recognize the molecular target we are interested in (Fig. 6.11). The protocol is composed of five steps: the creation of double-stranded DNA carrying the random “box”16 flanked by regions required for amplification; transcription of this DNA into single-stranded RNA; selection; production of a DNA population by reverse transcription and PCR of the sequences retained during the selection step, then cloning and sequencing of the strands obtained after a certain number of selection and amplification cycles.

From a vast combination of nucleic acids, one can isolate aptamers that possess catalytic properties (RNA ligation, cleavage or synthesis of a peptide bond, transfer of an aminoacyl group, etc.). The first nucleic acids could possess independent domains, separated by flexible segments, creating reversible conformational motifs, dependent on ions and bound ligands. Thus, a peptide that is 10 amino-acids long can recognize fine structural differences within a micro-RNA helix (discrimination can be made between two closely placed microhelices). Just as protein and antibodies, RNA molecules can present hollows, cavities, or slits that make these specific molecular recognitions possible. RNAs must “behave as proteins”. Whatever the chronology and the order of appearance of the various

![Synthetic DNA Pool](image)

Fig. 6.11. The SELEX method (adapted from Wilson and Szostak, 1999)

16 Region of defined length, for instance, of some 50 randomly aligned nucleotides
classes of molecules, the importance lies in the shape, the scaffolding and the architecture that have allowed functional associations.

Starting from a heterogeneous population of RNAs with $10^{15}$ variants (a population of $10^{15}$ different molecules) we have selected 5 populations of RNAs capable of specifically recognizing adenine after about ten generations (Meli et al., 2002). When cloned, sequenced and modelled, the best one among the individuals of these populations, has a shape reminiscent of a claw capable of grasping adenine. Is it the exact copy of a primitive ribo-organism that feeds on prebiotic adenine in prebiotic conditions? Functional and structural studies presently under way will highlight other activities, other conformations, etc.

Following this line of investigation we have selected two adenine-dependent ribozymes capable of triggering reversible cleavage reactions (Fig. 6.12). One of them is also active with imidazole alone. This result leads to very important perspectives (Meli et al., 2003).

A considerable amount of research has been focused on the selection of ribozymes in vitro. Recently, it was demonstrated that a ribozyme is capable of continuous evolution, adding successively up to 3 nucleotides to the initial molecule (McGuinnes, 2002). It is also possible to construct a ribozyme with only two different nucleotides, 2,6-diaminopurine and uracil (Reader and Joyce, 2002). Finally, Bartel and coworkers have selected a ribozyme-polymerase, capable of self-amplification (Johnston et al., 2001).

![Fig. 6.12. Adenine-dependent hairpin ribozymes (ADHR). Arrowheads: cleavage sites; Grey dots: degenerated (mutated) sites; Vertical bars: separation between the primer binding region and the random sequence](image)

### 6.4.4 Other Perspectives

Very little is known to date about the behaviour of macromolecules in “extreme” environments. How do structures behave? What are the major modifications observed? What are the conditions of structural and functional stability? How are
the dynamics of the macromolecules and their interactions affected? What are the possibilities of conserving biological macromolecules in very ancient soils or in meteorites? Can we find traces of these macromolecules as molecular biosignatures, and if so in what form (Maurel and Zaccai, 2001; Tehei et al., 2002)?

The selection of thermohalophilic aptamers, RNAs resistant to high temperatures ($80^\circ C$) in the presence of salt (halites that are 30 million years old), undertaken in our laboratory, will perhaps allow us to answer some of these questions, that are fundamental for the search of past traces of life, and of life on other planets.

6.5 Conclusion

The RNA world thus contains innumerable perspectives. The combination of methods available today are the best adapted to explore the vast combinations of nucleic acids but also of peptides. Will they make it possible to reconstitute the first steps of the living world? Attractive simulations may emerge, opening new evolutionary paths that have not been envisaged or that Nature has not yet explored.

The RNA world, at whatever step we place it in the history of the living world, must be considered as a step in the history of life, an important step in the evolution of the contemporary cellular world. Because of its strong explanatory power, it also constitutes an important opening in the scientific study of the origin of life. Even if this concept does not explain how life appeared, it nevertheless promises a great number of experimental breakthroughs.

Acknowledgement

Figure 6.4 is reprinted from *Cell*, 2003,112, Barends S., Bink H.H.J., van den Worm S.H.E., Pleij C.W.A., Kraal B. Entrapping ribosomes for viral translation: tRNA mimicry as a molecular trojan horse. Copyright 2003, with permission from Elsevier. We also thank Dr. G.F. Joyce for his constructive comments on the manuscript.

References


