

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Gene xx (2006) xxx–xxx

GENE
 SECTION
 EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

In vitro selection of halo-thermophilic RNA reveals two families of resistant RNA[☆]

 Jacques Vergne^a, Jean A.H. Cognet^b, Eörs Szathmáry^c, Marie-Christine Maurel^{a,*}
^a *Biochemistry of Evolution and Molecular Adaptability, Institute Jacques-Monod, CNRS, University Paris VI, Tour 43, 2 place Jussieu, 75251 Paris Cedex 05, France*
^b *Laboratoire de Biophysique Moléculaire, Cellulaire et Tissulaire CNRS, University Paris VI, UMR 7033, Genopole Campus 1, RN7, 91030 Evry Cedex, France*
^c *Group in Theoretical Biology and Ecology, Department of Plant Taxonomy and Ecology, Eötvös University and Collegium Budapest, 2 Szentháromság utca, 1014 Budapest, Hungary*

Received 20 September 2005; received in revised form 9 November 2005; accepted 17 November 2005

Abstract

The “RNA world” hypothesis proposes that early in the evolution of life, RNA was responsible both for the storage and transfer of genetic information and for the catalysis of biochemical reactions. One of the problems of the hypothesis is that RNA is known to be temperature sensitive. Nevertheless, different types of sequences with a thermostable phenotype may exist. In order to test this possibility, we applied an in vitro evolution method (SELEX) to isolate RNA molecules that are resistant at high temperatures (80 °C for 65 h) and high salt concentrations (2 M NaCl). The sequences of the resulting cloned halo-thermophilic RNAs can be grouped in two families (I and II) possessing very different thermal and chemical stabilities and very different secondary structures. The selected RNA molecules illustrate two different possibilities leading to thermal resistance which may be related to primitive conditions. We propose that members of family I constitute a good means of storing sequence information while members of family II are less efficient but replicate faster in early steps of the SELEX. These selected RNA behaviors may be related to primitive conditions and could allow to define limits for survival, and demonstrate that what is at stake for RNA molecules, as for living organisms, is survival and reproduction.

© 2006 Elsevier B.V. All rights reserved.

Keywords: RNA world; High temperature; High salt concentration; RNA stability; Amplification; Early evolution; Origins of life

1. Introduction

An important issue in the problem of the origins of life is whether or not an RNA world may be compatible with extreme primordial conditions. This scenario of evolution postulates that an ancestral molecular world is common to all present forms of

life; the functional properties of nucleic acids and proteins as we see them today would have been carried out by molecules of ribonucleic acids. 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 example, throughout its life cycle, the cell produces deoxyribonucleotides required for the synthesis of DNA that are derived from ribonucleotides. Thymine, a base specific of DNA, is obtained by transformation (methylation) of uracil, a base specific of RNA, and RNAs serve as obligatory primers during DNA synthesis. Furthermore, the current characterization of catalytic RNA molecules strengthens the RNA world hypothesis (Gilbert, 1986; Gesteland et al., 1999; Joyce, 2002) and is an additional argument in favor of the presence during evolution of RNAs before DNA.

Several investigations have corroborated the possible role of RNA in early life (Woese, 1967; Crick, 1968; Orgel, 1968),

Abbreviations: SELEX, Systematic Evolution of Ligands by Exponential enrichment; PCR, Polymerase Chain Reaction; dNTP, 2'-deoxyribonucleotide 5'-triphosphate; rNTP, ribonucleotide 5'-triphosphate; Tris, tris(hydroxymethyl) aminomethane; EDTA, ethylene diamine tetracetate; DTT, dithiothreitol; M-MLV, moloney murine leukemia virus; cDNA, DNA complementary to RNA; CD, circular dichroism; NIH, National Institutes of Health; bp, base pairs; RT-PCR, reverse transcription-polymerase chain reaction; T_m, melting temperature.

[☆] Colour version of the article is available online in GENE at www.sciencedirect.com.

* Corresponding author. Tel.: +33 1 44 27 40 21; fax: +33 1 44 27 99 16.

E-mail address: maurel@ijm.jussieu.fr (M.-C. Maurel).

including data on template-directed synthesis of macromolecules (Joyce and Orgel, 1999), oligomerisation on mineral surfaces (Ferris et al., 1996), and possible transitions from an alternative genetic system to RNA (Schmidt et al., 1997; Kozlov et al., 1999; Chaput et al., 2003). DNA has been 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.

From a fundamental point of view, it is interesting to study the behavior of nucleic acids under extreme conditions (Tobé et al., 2005). It is generally assumed that the upper temperature limit for life is related to the instability of essential key molecules such as nucleic acids (Marguet and Forterre, 1994). Many RNA molecules including members of a growing list of artificially selected aptamers are single-stranded, and much more labile than single-stranded DNA. RNA is thermosensitive creating a problem if early life was thermophilic as is assumed in some scenarios (Wächtershäuser, 1992), compatible with some views on the last common ancestor (Di Giulio, 2000, 2003a,b; Schwartzman and Lineweaver, 2004).

In addition, there are indications that primitive oceans on earth were highly saline (Knauth, 1998), and early life has been discussed in terms of hypersaline conditions (Dundas, 1998). Salt and water are tightly linked: salt is abundant on Mars, and on Europa one of the satellites of Jupiter (Kargel, 1998). Finally, very ancient salt-tolerant bacteria have been extracted from brine inclusions in salt crystals (Vreeland et al., 2000). These arguments are all in favor of an extreme saline primordial environment.

The fragility of RNA structures and activity however, is the subject of controversy and requires direct experimentation. In vitro selection has been used for creating and identifying molecular interactions. It has been successfully used to isolate and characterize thermodynamically stable tri- and tetraloop RNA hairpins (Bevilacqua and Bevilacqua, 1998; Shu and Bevilacqua, 1999; Proctor et al., 2002). A selection has been devised to select RNAs based on their local folding stability. A large RNA molecule with 19 randomized nucleotides shows dependence upon tetraloop identity and on mutations in the nonrandomized region of the RNA (Juneau and Cech, 1999; Guo and Cech, 2002). Regarding the abundance of halite minerals in the early environment, it is important to consider the role of high salt concentrations on macromolecules. We have already shown the effect of high salt concentrations on the structural integrity of tRNAs and on maintaining residual specific charging capacity after heat treatment of the tRNAs for 30 h at 82 °C (Tehei et al., 2002). These studies have provided support for the importance of salt in the protection of macromolecules against thermal degradation allowing their activity to be maintained. In order to learn more about the biochemical properties of RNA molecules, we studied the stability of such molecules selected in vitro at high temperatures in the presence of salt using the SELEX (systematic evolution of ligands by exponential enrichment) method (Tuerk and Gold,

1990). Our results, which brought to light two families of thermal resistant RNA molecules could also be useful in the search for traces of life in ancient sediments and in planetary exploration.

2. Materials and methods

2.1. Preparation of the starting RNA pool

Single-stranded DNA template and primers (Marshall and Ellington, 2000; Meli et al., 2002, 2003) were chemically synthesized (Proligo and MWG-Biotech). The sequence of the forward primer 1 (P1) is 5'-**GGTAATACGACTCACTA-TAGGGAGATACCAGCTTATTCAATT**-3' (T7 promoter sequence in bold), and of the reverse primer 2 (P2) is 5'-AGATTGCACTTACTATCT-3'. The 86 nt-long template (T) 5'-ATACCAGCTTATTCAATT-(N)₅₀-AGATAGTAAAGTGCATCT-3' consists in a 50 nt-long random sequence where N stands for any nucleotide, flanked by two constant regions of 18 nt each for PCR amplification.

A 2 ml PCR (Invitrogen) reaction mixture containing 1 μM of each primer P1 and P2, 0.085 μM template T (10¹⁴ random molecules), 0.2 mM dNTPs, 20 mM Tris-HCl pH 8.4, 1.5 mM MgCl₂, 50 mM KCl, and 0.025 u/μl of Taq DNA polymerase, was treated using the following cycling parameters: denaturation at 94 °C for 30 s, annealing at 40 °C for 2 min and elongation at 68 °C for 2 min.

After PCR, the starting random dsDNA pool was ethanol precipitated and submitted to in vitro transcription: the 1.65 ml reaction mixture contained 2 mM each rNTP, 40 mM Tris-HCl pH 7.9, 2 mM spermidine, 6 mM MgCl₂, 10 mM NaCl and 10 mM DTT (Fermentas buffer), 0.02 mg/ml of dsDNA pool and 0.6 u/μl T7 RNA polymerase (Fermentas). After overnight incubation at 37 °C, it was ethanol-precipitated and the DNA template degraded by DNase 1 (RQI RNase-free DNase, Promega). The RNA was deproteinized, ethanol-precipitated, and purified on a 6% denaturing polyacrylamide gel (19/1 acrylamide/bisacrylamide, Tris-borate-EDTA buffer pH 8.3 and 7 M urea). The RNA was eluted from the gel, treated with phenol, phenol/chloroform/isoamyl alcohol, twice with chloroform/isoamyl alcohol and precipitated three times with ethanol. Finally the RNA was dissolved in water.

2.2. Selection procedure

Each selection round consisted in heating the RNA (140 μl containing 0.28 mg/ml of RNA) to 80 °C in 2 M NaCl for 65 h in a tightly closed microtube. Prior to heating, a 10 μl control sample was set aside as the “t₀”. The integrity of the RNA was checked after each round by electrophoresis of a 10 μl sample on a 2% native agarose gel in Tris-acetate-EDTA buffer pH 8.3 plus ethidium bromide. The mixture was desalted by ultrafiltration (washing with water on a Microcon YM-10, Amicon), and the thermoresistant molecules were purified by electrophoresis on a 6% denaturing polyacrylamide gel as above. The full-length RNA molecules were recovered from their expected migration position, concentrated, and washed

with water by ultrafiltration for the subsequent rounds of amplification.

2.3. Amplification of selected RNAs

The pooled full-length RNAs were used for reverse transcription using 10 u/μl M-MLV Reverse Transcriptase (Invitrogen) in a 100 μl reaction mixture with 1.25 μM primer P2, 0.5 mM dNTPs, 50 mM Tris–HCl pH 8.3, 3 mM MgCl₂, 75 mM KCl, and 10 mM DTT at 37 °C for 50 min, then stopped at 70 °C for 15 min. The cDNA synthesized was amplified by PCR (80 μl of the cDNA reaction mixture for 2 ml PCR reaction mixture as above) to produce the DNA template for the next transcription round. The new RNA pool purified as above was then ready for the following selection cycle.

2.4. Cloning and sequencing

After conversion of the RNAs to DNA by reverse transcription and PCR, the RNAs selected were screened by cloning (TOPO TA Cloning kit, Invitrogen). Plasmids were prepared (Plasmix minipreps, Talent kit) from selected clones and sequenced to identify the DNA inserted (Genome Express Service). After PCR amplification of the inserted DNAs and transcription, individual RNAs were prepared as above for further analyses.

2.5. Modeling of secondary structures

Secondary structures of isolated RNAs were predicted with the mfold web server 2.3 version, <http://www.bioinfo.rpi.edu/applications/mfold/old/rna/> (Zuker, 2003).

2.6. Thermal melting curves

Thermal melting curves were performed on a UVIKON XL spectrophotometer. Absorbance at 260 nm of the RNA (0.5 μM), in 10 mM sodium cacodylate pH 7.0 with 0 to 2 M NaCl, was recorded while heating (0.4 °C/min) from 20 to 95 °C.

2.7. CD measurements

CD spectra (Bajji and Davis, 2002) were recorded on a JASCO-J810 spectrophoto-polarimeter using a cell with a 1 mm path-length. Before measurements, the RNA (10 μM) in 2 mM sodium cacodylate pH 7.0, 0.1 mM EDTA, and 2 M NaCl was heated at 95 °C for 3 min and slowly cooled to 20 °C (3 °C/min). CD spectra were recorded at several temperatures from 20 to 80 °C, each one differing by 10 °C.

2.8. Thermostability of cloned RNAs

After heating to 80 °C, the thermostability of the cloned RNAs was evaluated by electrophoresis on native agarose gels and on denaturing polyacrylamide gels as indicated above to clearly analyze the integrity and size of the RNAs; the RNAs

stained with ethidium bromide were quantified using the NIH (National Institutes of Health) image software.

2.9. Efficiency of DNA production of T25 and T24 in the RT-PCR reaction

Initial RNA concentrations of T24 and T25 were in the range of 0.5×10^{-11} to 0.5×10^{-17} M for reverse transcription with 10 u/μl M-MLV Reverse Transcriptase (Invitrogen) in a 100 μl reaction mixture, 1.25 μM primer P2, 0.5 mM dNTPs, 50 mM Tris–HCl pH 8.3, 3 mM MgCl₂, 75 mM KCl, and 10 mM DTT at 37 °C for 50 min, then stopped at 70 °C for 15 min. The cDNA synthesized was amplified by PCR (8 μl of the cDNA reaction mixture for 200 μl PCR reaction, mixture as above). Products of amplification were analyzed on polyacrylamide 10% (19/1) native gel electrophoresis, stained with ethidium bromide, and quantified using the NIH image software.

3. Results

3.1. Isolation of selected thermoresistant RNA molecules

The selection began with a large population of RNA (10^{14} diversity) produced by transcription of a 50 nucleotide (nt)-long random sequence flanked by two constant sequences (23 and 18 nt) required for primer binding in the SELEX method. This yielded an RNA pool of 5'-GGGAGAUACCAGCUUAUUCAAUU-(N)₅₀-AGAUAGUAAGUGCAAUCU-3'.

The molecules were heated to 80 °C for 65 h in the presence of 2 M NaCl. Most were degraded in the first cycles, whereas unheated RNA molecules produced a single intense band. After 10 to 11 cycles, analysis of the heated RNA by native agarose gel electrophoresis still presented a smear of degraded molecules (Fig. 1). In selection cycles 12 and 13, the intensity of the smear moved towards the region of undegraded material, and became a “single” intense band migrating near the unheated sample as of the 14th cycle. This last step was judged appropriate for cloning of the RNA species contained in the resistant material.

3.2. Cloning and sequencing show two different families

The RNA from the 14th cycle was cloned as indicated in Section 2.4. Sequencing of 50 plasmids revealed 30 different

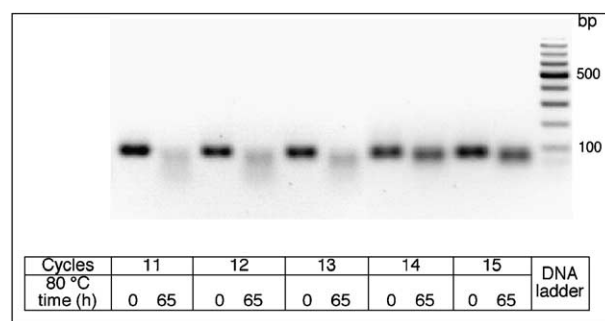


Fig. 1. Selection evolution during cycles 11 to 15 of the SELEX procedure. Analysis by native agarose gel electrophoresis of RNA integrity, at time 0 and after 65 h at 80 °C in 2 M NaCl.

sequences. Multiple sequence alignments and an unrooted dendrogram (from CLUSTALW software) showed two very distinct families of RNA molecules with equal numbers of sequences (Table 1 and Fig. 2). The sequences display a great dispersion in family I, while the sequences in family II exhibit only minor point variations. In each family, point variations are presented as shadowed letters. Most members of family I contain a very long stretch of consecutive nt (Table 1: red in the variable region) complementary to the 5' constant region of 23 nt, and some contain an additional shorter stretch of 5–6 consecutive nt (highlighted in yellow) complementary to a small sequence located within the 3' constant region.

Table 1 reveals two subgroups in family I, subgroup A: T42–T8, and subgroup B: T46–T21. Sequence data suggest they are formed of a long hairpin composed of a loop 4–10 nt

long (black) from the variable region and a stem (red) whose arms derive essentially from the 5' constant region and the variable region. In subgroup B one arm of the stem is composed of 23 nt of the 5' constant sequence minus 2 unpaired nt (AU) and 8 nt of the variable region and is complementary to 29 nt of the variable region that constitutes the other arm. Subgroup B also contains a shorter stem (stem 2 highlighted in yellow) of 5–6 nt whose arms are composed of nts from the 3' constant region and the variable region. An equivalent shorter stem is absent from subgroup A. Family II features two major stems. The one located just beyond the 5' constant region contains 8 complementary nt (red) from the variable region. The shorter stem located close to the 3' end of the molecule contains 5–7 consecutive nt (highlighted in yellow) from the variable region complementary to a small

Table 1
Sequences from the constant regions at the 5' and 3' ends are underlined

	5'	3'	
Family I	T42	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> CCUCUGUC <u>CCUCUUAUA</u> GACGGAGGAGAGUUGAGUAGGC <u>GGCCCCCAGAUAGUAAGUGCAAUCU</u>	Subgroup A
	T52	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> CCUGCCGGAUAGUGCGGGGAGUUGGUGAGUUGGGUCUCCUUGGGUGCUAGAUAGUAAGUGCAAUCU	
	T18	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> CCUGCCGGAUAGCGCGGGGAGUUGGUGAGUUGGGUCUCCUUGGGUGCUAGAUAGUAAGUGCAAUCU	
	T39	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGCGCCCGUCUGCGGAGCCCGAUUGGGUAGUUGGGUCUCCCGCCUCUAGAUAGUAAGUGCAAUCU	
	T13	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> AGUCUCGACCCGCGGGGCGCCAGUUGGGUAGGCUGGGCUGUCUGGCUAGAUAGUAAGUGCAAUCU	
	T37	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> CGGUGUUCUGAUUACCGGGGGCCCGGGUUGGGUAGUUGAAUUGAUUCCCGAGAUAGUAAGUGCAAUCU	
	T10	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> CGCCACGUGGUUCCGUGGGGUAAGUUGAGUAGGUUGGGCUCUAGAUAGUAAGUGCAAUCU	
	T8	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGUCUGGGCCGAUUGGGUAGGUUAGUAUCCUGCGCUCCUAGAUAGUAAGUGCAAUCU	
Family I	T46	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGACCGUUUUUGCGACGGUUCGAUUGGGUAGGUUGGUCUCCGUGCGCUAGAUAGUAAGUGCAAUCU	Subgroup B
	T45	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGGCCGUUUUUUGCGACGGUUCGAUUGGGUAGGUUGGUCUCCCGCGCGCUAGAUAGUAAGUGCAAUCU	
	T36	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGGCCGUUUUUUGCGACGGUUCGAUUGGGUAGGUUGGUCUCCCGUGCGCUAGAUAGUAAGUGCAAUCU	
	T28	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGGCCGUUUUUUGCGACGGUUCGAUUGGGUAGGUUGGUCUCCCGUGCGCUAGAUAGUAAGUGCAAUCU	
	T24	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGGCCGUUUUUUGCGACGGUUCGAUUGGGUAGGUUGGUCUCCCGUGCGCUAGAUAGUAAGUGCAAUCU	
	T26	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGGCCGUUUUUUGCGACGGUUCGAUUGGGUAGGUUGGUCUCCCGUGCGCUAGAUAGUAAGUGCAAUCU	
	T21	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> GGGCCGUU–UUUGCGACGGUUCGAUUGGGUAGGUUGGUCUCCCGUGCGCAGAUAGUAAGUGCAAUCU	
	Family II	T40	
T48		-	U U U A C
T4			U UU UG CUC
T32			U U G C
T14			U G C
T23			G C
T22			G C
T41			U U G C
T6		U	U G
T17			U U G U
T49			U U
T2			U G U
T25	GGGAGAU <u>ACCAGCUUAUUCAAUUC</u> –GGCGGUU <u>CACCU</u> CCUAGCCUCUCUA <u>CACUUC</u> CUAGAUAGUAAGUGCAAUCU		
T38		U G	
T47		G	

10 20 30 40 50

1–23456789 123456789 123456789 123456789 123456789

Constant region (23 nt) Variable region (48–51 nt) Constant region (18 nt)

Nucleotides shown in black correspond to bulges, loops or unassigned regions. Sequences are shown in different colors to highlight each double-stranded stem region. Family I is composed of subgroups A and B. In subgroups A and B, the long stretch in the variable region (up to 29 nt, red) is complementary to the 5' constant region composed of the 23 nt minus 2 unpaired nt of the underlined constant region plus 8 nt from the variable region. In subgroup B the small stretch in the 3' part of the variable region, is complementary to nt (both highlighted in yellow) of the 3' constant region. Shadowed letters are point mutations as compared to T24 taken as reference. For family II, which shows few variations, T25 is chosen as the reference. Point mutations in other members of family II as compared to T25 are indicated by shadowed letters. The longest complementary sequences are shown in red, and highlighted in yellow. Other colours correspond to possible pairings, blue with blue or green with green (see also Fig. 3A). (For interpretation of the references to colour in this Table, refer to the web version of this article.)

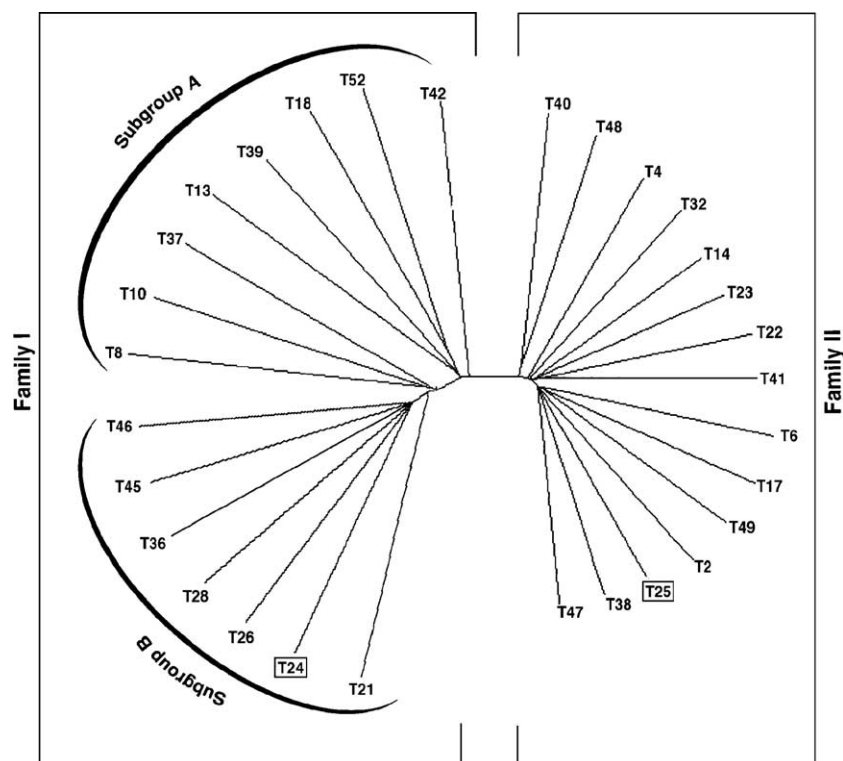


Fig. 2. Unrooted dendrogram of the 30 different sequences obtained by SELEX. In family I, members of subgroups A and B are highlighted. The boxed clones T24 and T25 serve as examples of the results described.

sequence located within the 3' constant region. Most point mutations shown in shadowed letters are located outside these double-stranded regions.

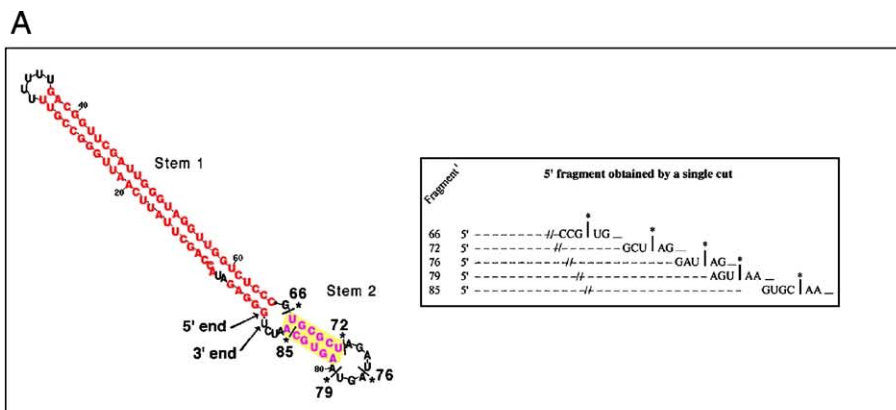
3.3. Modeling secondary structures suggests a rod-like family and a family with a complex stem-loop structure

The basic secondary structures obtained above by inspection of the sequences is detailed by thermodynamical modeling (RNA mfold) in 1 M NaCl (Fig. 3A). It suggests for members of family I two probable stems, 1 and 2, at 80 °C, and for members of family II a highly unpaired structure with three short stems at 65 °C. We observed that experimental melting curves of two representative members are close at 1 M and 2 M NaCl. At 37 °C and at 80 °C, modeling of members of family I reproduces essentially the same very long stem structure at the 5' end of each RNA, with few internal loops and bulges; it is built in part from a region of primer P1 and of a matching sequence present in the variable region (not shown). In addition, a small stem (stem 2) partly formed with the transcribed reverse primer P2 is present at the 3' end.

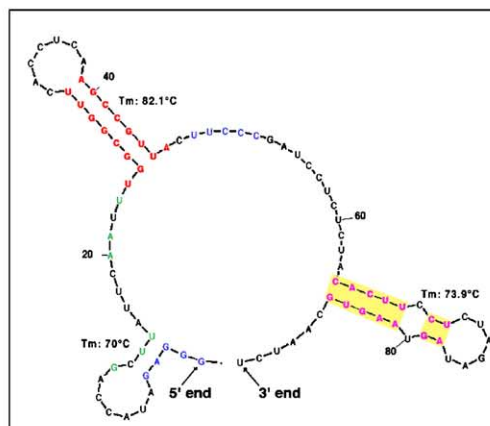
Half of the RNAs belonging to family I, (T46, T45, T36, T24, T26, and T21) are very stable and retain their complete secondary structure at 80 °C. Among the other RNAs of family I, pairings at 80 °C are disrupted essentially in stem 2 in T52, T18, T13 and T28 but also locally in stem 1 where an internal bulge arises in T42, T37, T10 and T8. In T39 a few punctual unpaired nucleotides appear at weak points.

Among the selected RNAs, those of family I with long double-stranded regions have the most stable secondary structures. In T24 taken as a model for subsequent studies, it is remarkable that to pair the seven G of the constant sequences in stems 1 and 2 (29 and 6 nt long, respectively), six C and one U were selected in the variable region, indicating a preference for the more thermostable base pair (G·C rather than G·U). Likewise, to pair seven U of the same regions, five G and two A were selected, indicating a clear preference for G. In the remaining upper part of stem 1, (GGGCCGUU)·(GACG-GUUC), in which both strands are from the random region, seven out of eight base pairs (bp) contain a G (Fig. 3A). Thermodynamic parameters have been reported and tested for RNA hairpin formation (Dale et al., 2000; Proctor et al., 2002; Vecenie and Serra, 2004). The sequence of the 5' loops of members of subgroup B are UUU(G or U) (C or U). They may be a variant of a tetraloop motif with a one-nucleotide addition (Varani, 1995). The tridimensional structures of UUCG tetraloops are well characterized (Santini et al., 2003) but the structure of pentaloops have not been solved.

At 37 °C, members of family II form variable complex structures with various short stems, loops and bulges (not shown). Modeling at 80 °C indicates that increasing the temperature eliminates most of the secondary structures, leaving only a small stable hairpin near the 5' end of the molecule with a stem of 7 bp (red) and a highly conserved loop of 7 bases (black). Only T6 features a point mutation in this loop. Thus the loop with the sequence CA(C or U)CUCA may adopt an unusual stable 3D structure. Family II can be considered as



Family I (T24)



Family II (T25)

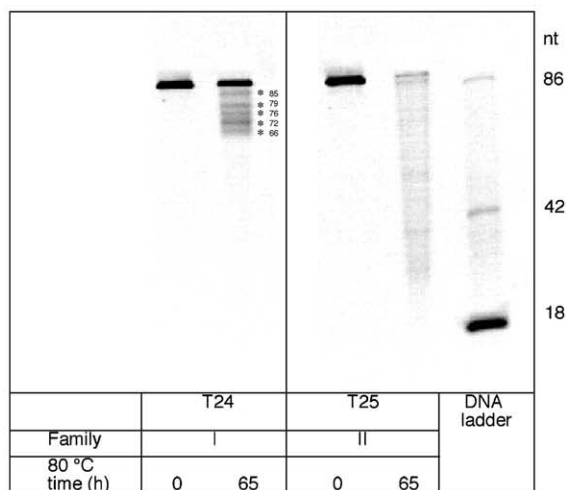
B

Fig. 3. Regions of most stable secondary structures in relation to analysis of degradation by electrophoresis. (A) Typical regions of most stable secondary structures proposed by mfold for the representative members T24 and T25 of families I and II, respectively, are shown with different colors. T24 presents the most extensive region of stable secondary structure at 80 °C in 1 M NaCl (red and red–yellow). Inset¹: estimated size of the 5' fragment. Bars with asterisks indicate sites of presumed cleavages. T25 is shown as composed of a group of the most stable secondary structures predicted by mfold at 65 °C in 1 M NaCl (red and red–yellow). Other possible pairings appearing at 37 °C are shown in blue and green. (B) Denaturing gel electrophoresis of T24 and T25. The thermoresistance to degradation after heating to 80 °C for 0 or 65 h in 2 M NaCl was evaluated by analysing the integrity of the cloned RNAs. After migration the products were quantified using the NIH image software. T24 produced discrete bands of about 85, 79, 76, 72 and 66 nt (asterisks) which correspond essentially to degradation of stem 2; T25 appeared highly degraded forming a long smear. (For interpretation of the references to colour in this figure, refer to the web version of this article.)

composed of a population of RNAs with a short secondary structure (red) resistant up to 80 °C, and with no known large stable secondary conformation, but with a remarkably well conserved sequence. Modeling at 65 °C of T25 used as representative of family II, however suggests a second stable hairpin with a seven base pair stem (highlighted in yellow) with a C, U mismatch, and a 6 nt loop (Fig. 3A).

3.4. Thermal melting

The results of melting studies of two representative clone sequences at low and high salt concentrations (Fig. 4) are summarized in Table 2. Melting of T24, in buffer without NaCl, displays a small amplitude sigmoidal curve between 20 and 45 °C with an inflection point (T_m 1) at 35 °C, followed by a large amplitude and sharp sigmoidal curve between 45 and 60 °C with an inflection point (T_m 2) at 54 °C. The curve reaches a plateau at 65 °C. The temperature and amplitude of the first transition would correspond to melting of the short stem 2, and second transition to melting of the long stem 1, which

correspond to about 80% of the total double stranded structure. The thermal melting curve of T25 in buffer without NaCl shows a large and gradual sigmoidal curve with one inflection point (T_m 32 °C), probably due to the presence of many small secondary structures melting closely. For both families addition of NaCl shifts the curve towards higher temperatures. For T24 increasing the NaCl concentration (from 0 to 2 M) shifts T_m 1 value from 35 to 72 °C and T_m 2 from 54 to 84 °C, indicating that two substructures of family I are stabilized. For T25 the T_m shifts from 35 to 58 °C with broader sigmoidal transitions indicating a disparity in the effect of salt on the small component structures. Stability structure of T24 and T25 was also analyzed by circular dichroism (Fig. 5). CD spectra were recorded at different temperatures from 20 to 80 °C in 2 M NaCl. Increasing temperature from 20 to 80 °C induces a 30% decrease of the main peak at 265 nm for T24, and a 50% decrease for T25. Therefore, melting T24 is more difficult than melting T25, which confirms that structure of T24 is more stable than structure of T25.

3.5. Properties of selected RNAs: thermal resistance, degradation kinetics and amplification

The cloned RNAs were checked for their resistance to degradation by heating to 80 °C for 65 h in 2 M NaCl. The products were analyzed by denaturing and native gel electrophoresis. In denaturing gels (Fig. 3B), T24 produced about 45% of intact molecules and a set of well defined discrete bands (85–66 nt) indicating a very resistant core in T24, which bring up to 85% the total resistant material. These discrete bands correspond closely to the sizes of the fragments produced by the degradation of the stem-loop 2 shown in Fig. 3A and summarized in Table 2. In contrast T25 appeared highly degraded forming a long smear, with only about 10% undegraded molecules. On native gels (Fig. 6) the RNAs of family I produced an intense narrow band and were far more stable than the RNAs of family II.

In order to evaluate the enrichment in the stable RNA molecules, degradation kinetic studies were performed at 80 °C, comparing T24 to the initial population (Fig. 7). T24 appeared very stable even after 65 h, while degradation of the initial RNA population was virtually complete after 65 h. The majority of T24 was full-length (Fig. 7A), and from the kinetic curves (Fig. 7B), a 20 times relative degradation rate ratio appeared between the initial population and T24.

The outcome of this SELEX is a distribution of cloned sequences in two families equally represented. Family I characterized by very long (up to 29 bp) thermodynamically and chemically stable hairpin structures terminated by a short loop. Consequently, up to 85% of the material remains undegraded after 65 h at 80 °C. Family II is a family of sequences with only point mutations, very short secondary structures probably associated with lower thermal stability and with much lower resistance to thermal degradation. Only 10% of the initial material remains undegraded after 65 h at 80 °C. Despite this low level of stable molecules, family II is highly represented after 14 cycles of selection. To compensate for this

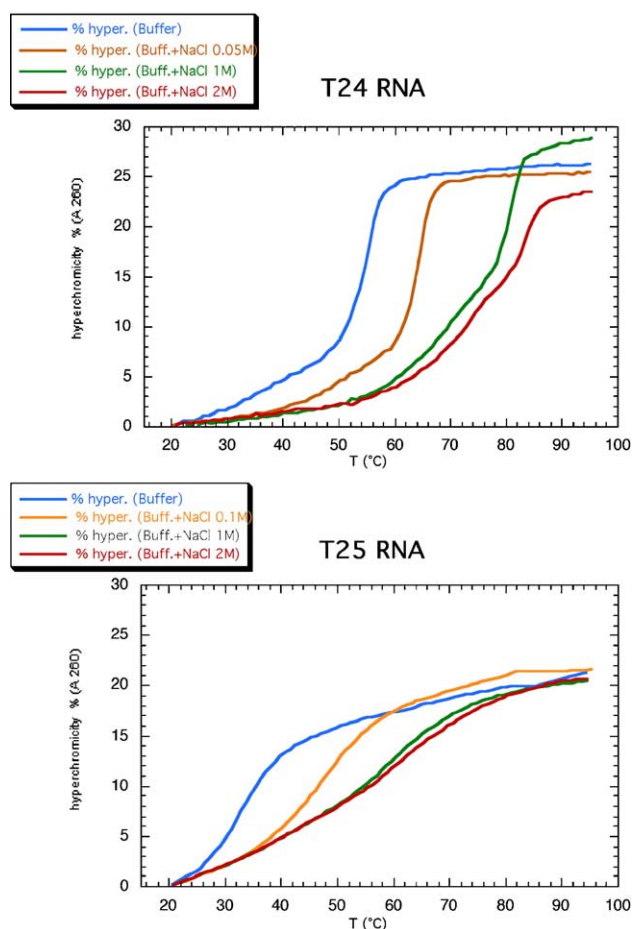


Fig. 4. Thermal melting study by UV absorbance. Melting of T24 and T25 RNA were followed by UV absorbance at 260 nm as a function of temperature. Structural stabilization by salt was studied by adding NaCl from 0 to 2 M to the buffer (10 mM sodium cacodylate pH 7.0). Results are expressed as percentage of hyperchromicity, resulting from unfolding, as the temperature was increased from 20 to 95 °C. (For interpretation of the references to colour in this figure, refer to the web version of this article.)

Table 2
Results obtained from thermodynamical calculations, gel electrophoresis, melting curves and CD for a representative clone of each family I and II, T24 and T25, respectively

Representative sequences (total length)	Proposed secondary structure and single strand cut positions (cf. Fig. 3A)	Observed single strand cut positions from denaturing gel electrophoresis (cf. Fig. 3B)	Thermostability from Tm and CD	Thermo-degradation (after 65 h at 80 °C in 2 M NaCl) (cf. Fig. 3B)
T24 (family I) (90 nt)	Major stem-loop 1, minor stem-loop 2 *66 *72 *76 *79 *85	Well identified bands 66 nt: –24 nt from 3' end 72 nt: –18 nt from 3' end 76 nt: –14 nt from 3' end 79 nt: –11 nt from 3' end 85 nt: –5 nt from 3' end	–Tm: two transitions. Tm1 (35 to 72 °C for [NaCl]: 0 to 2 M), attributed to melting of stem 2. Tm2 (54 to 84 °C for [NaCl]: 0 to 2 M), attributed to melting of stem 1 –CD analysis at 2 M NaCl. From 20 to 80 °C, 30% decrease of ellipticity at 265 nm	15% degraded
T25 (family II) (91 nt)	Many possible small stems	Continuous smear over the range from 91 to ~28 nt	–Tm: one transition. Tm (35 to 58 °C for [NaCl]: 0 to 2 M) attributed to melting of a set of small stems –CD analysis at 2 M NaCl. From 20 to 80 °C, 50% decrease of ellipticity at 265 nm	90% degraded

* Single strand cut positions (cf. Fig. 3A).

lack of stability, the selected sequences may replicate better than the very stable sequences of family I. This compensation suggests that the global replication over the entire SELEX

process is on the order of five times greater for members of family II than for those of family I. In order to study this hypothesis, the amount of T24 and of T25 DNA obtained from the RT-PCR reactions was quantified for different initial RNA concentrations close to those of the first cycles of SELEX (about 10^{-19} M for individual member of the initial population). In the range 0.5×10^{-11} to 0.5×10^{-17} M, the quantification of the ratio of DNA obtained for T25/T24 after PCR is observed to increase from 1 to about 8 folds at very low concentrations (Fig. 8). This result reveals that the two RNA families display different behaviors with respect to the RT-PCR enzymes at the beginning of the SELEX process. When RNA concentration is very low, production of T24 DNA is lower than that of T25 DNA suggesting that uptake of T24 by the RT-PCR enzymes is less than that of T25.

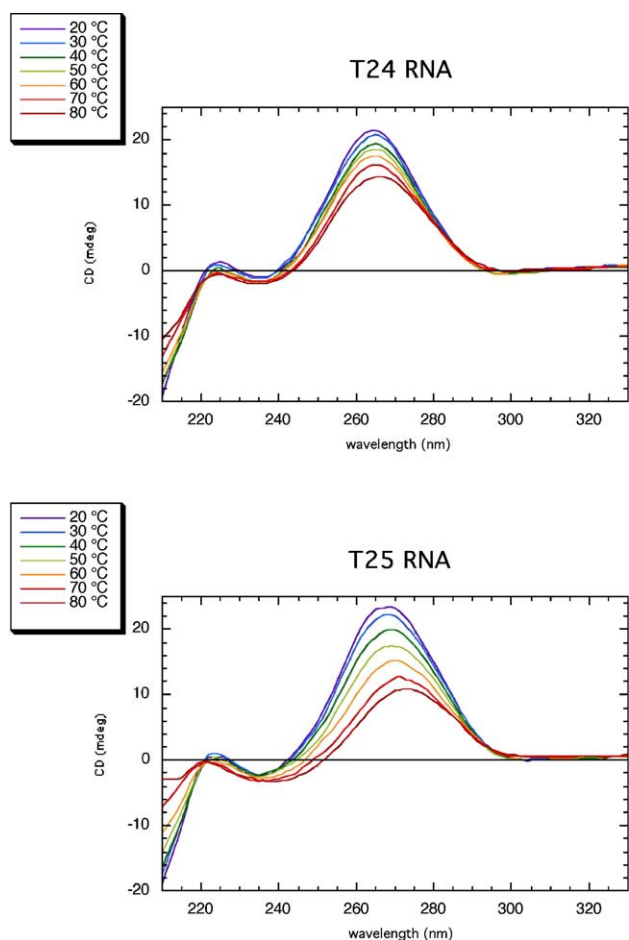


Fig. 5. Thermal melting study with CD. Influence of the temperature from 20 to 80 °C on the CD spectra (ellipticity measurements, mdeg) of T24 and T25 RNA in high salt concentration conditions: 2 M NaCl, 2 mM sodium cacodylate pH 7.0, 0.1 mM EDTA. Wavelength is in the range 210 to 330 nm. (For interpretation of the references to colour in this figure, refer to the web version of this article.)

4. Discussion

The “RNA world” is now recognized as one of the main hypotheses put forward to explain a critical phase in the origins of life. Yet, to date little has been done to investigate the behavior of RNA under the extreme conditions that could have

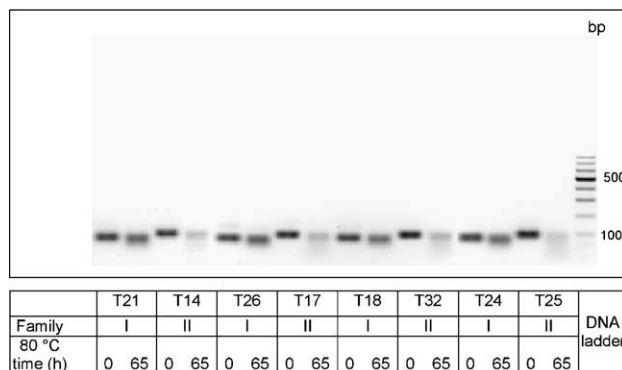


Fig. 6. Native gel electrophoresis of isolated RNAs. After cloning, isolated RNAs of each family were checked for their resistance at time 0 and after 65 h at 80 °C in the presence of 2 M NaCl.

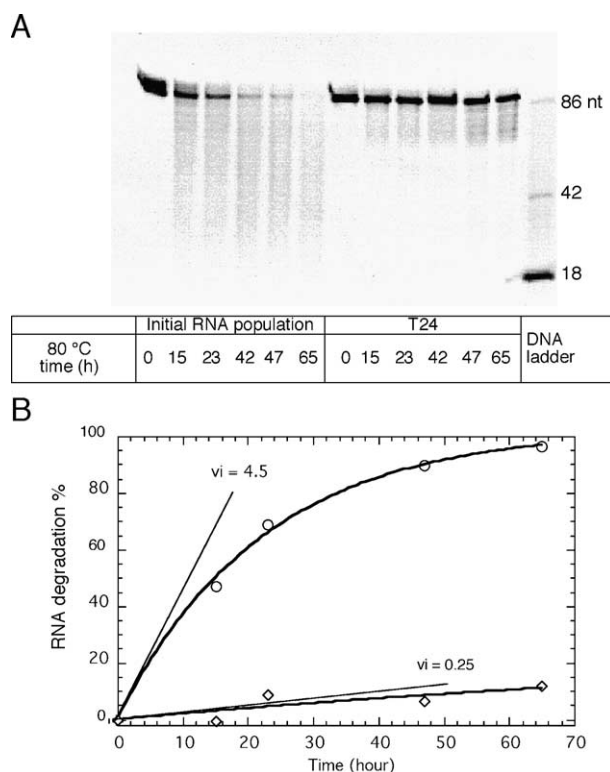


Fig. 7. Kinetic degradation profiles of T24 and of RNA initial population. The degradation profiles were followed by denaturing gel electrophoresis. (A) The initial RNA population and T24 were heated at 80 °C in the presence of 2 M NaCl from 0 to 65 h. The resistant material appearing in the gel was quantified as a function of time (NIH image software). The T24 resistant molecules were either undegraded or partly degraded leaving an undegraded core of about 66 nt. As visible in Fig. 3A (labelled with asterisks and bars), subfragments of about 85, 79, 76, 72 and 66 nt were produced. (B) Enrichment stability was evaluated by comparing the initial degradation rates (v_i) of the initial population (O) and of T24 (◇).

prevailed at that time, such as high salt concentrations and high temperatures.

This is the first time that the SELEX method has been applied to test the outcome of exposure of RNAs to such extreme conditions. The data obtained from sequencing the cloned RNA, thermal melting curves and CD, and from gel electrophoresis, compared to thermodynamical secondary structure calculations, agree with a stabilizing effect of the RNAs by formation of stem loop substructures. All the results are consistent with the minimal conformations proposed in Fig. 3 and summarized in Table 2.

Studies of thermal stability and of secondary structures by thermodynamical calculations show an important difference between the families I and II. The secondary structures of family I RNAs contain longer and more stable double-stranded stems than the proposed secondary structures of family II. Experimental results show that, after heating to 80 °C in 2 M NaCl, the integrity of the RNAs of family I is greater than that of family II (Figs. 3B and 6). Consequently, formation of stem loop and base pairing is therefore the main explanation for the higher thermal and chemical stability of family I.

Since pairing is essential for resistance to high temperatures, selection led as expected to RNAs with a base composition rich

in G and in G-C pairs. When one of the strands of the stem is from the constant region and the other strand is from the variable region, G-C base pairs are favored over G-U base pairs. When both strands are from the variable region, the stem is very rich in G-C and G-U base pairs as observed for families I and II (Fig. 3A, red).

Thermal melting profiles are in good agreement with the proposed minimal secondary structures calculations. The structure of T24, (Fig. 3A) representative of subgroup B in family I (Table 1), is mainly composed of the long stable stem 1 (29 bp) and the short less stable stem 2 (6 bp). The melting curves of T24 correspond to the sum of melting two different helical structures with different T_m values (Table 2). Increasing the salt concentration stabilizes both stems, shifting their T_m values to higher temperatures. The melting curve obtained in the presence of 2 M NaCl, suggests that stem 2 would be unpaired at 80 °C while stem 1 would partly persist. CD analysis between 20 and 80 °C shows also a partial melting for T24 less than for T25 which confirm a global more stable structure for T24 than for T25. The single-stranded region resulting from unpairing of stem 2 would be sensitive to statistical degradation from the 3' end and corresponds to the electrophoretic results obtained (Table 2). The melting curve of T25 (family II) is in agreement with the melting of complex secondary structures containing short stems and loops. Therefore a small well-conserved secondary structure is important for the chemical stability of RNA at high temperatures and high salt concentrations (Fig. 3A, red). Salt has a protective effect by decreasing the repulsive interactions between negative phosphate groups in double stranded structures. Negative charges are neutralized by sodium cations, but other cations like ammonium or potassium could be used.

The chemical susceptibility of RNA phosphodiester to degradation has been investigated thoroughly as a function of pH, K^+ or Mg^{2+} concentrations, and temperature (Li and Breaker, 1999). The projected reaction rate constant for r(ApG) at 23 °C, pH 7.0, 2 M K^+ is $1.7 \times 10^{-8} \text{ min}^{-1}$. At 80 °C, it is $1.7 \times 10^{-4} \text{ min}^{-1}$. Therefore RNA is expected to be degraded 10,000 times faster at 80 °C than at 23 °C under these conditions. Assuming to first approximation that the reaction rate constant is sequence independent, for an RNA of 90 nucleotides, the rate would be $1.5 \times 10^{-2} \text{ min}^{-1}$. As a result, after 65 h at 80 °C, 2 M KCl, one would expect 78% of dinucleotide r(ApG) to remain intact, and $1.5 \times 10^{-8}\%$ of an RNA of 90 nucleotides, i.e. complete degradation. These crude estimations are based on the chemical susceptibility of the RNA phosphodiester, r(ApG) to degradation. Our results are in qualitative agreement with these studies of Li and Breaker.

In most SELEX experiments the selection pressure is applied to obtain molecules with definite functional structures. In the SELEX experiments reported here we investigate the structural properties of resistant RNA molecules without specific functional requirements. As observed before, in vivo or in vitro molecular evolution tends to favor increased thermodynamical stability with secondary structures (Schultes et al., 1999; Gulyaev et al., 2002). General investigations of naturally and artificially selected RNA secondary structures

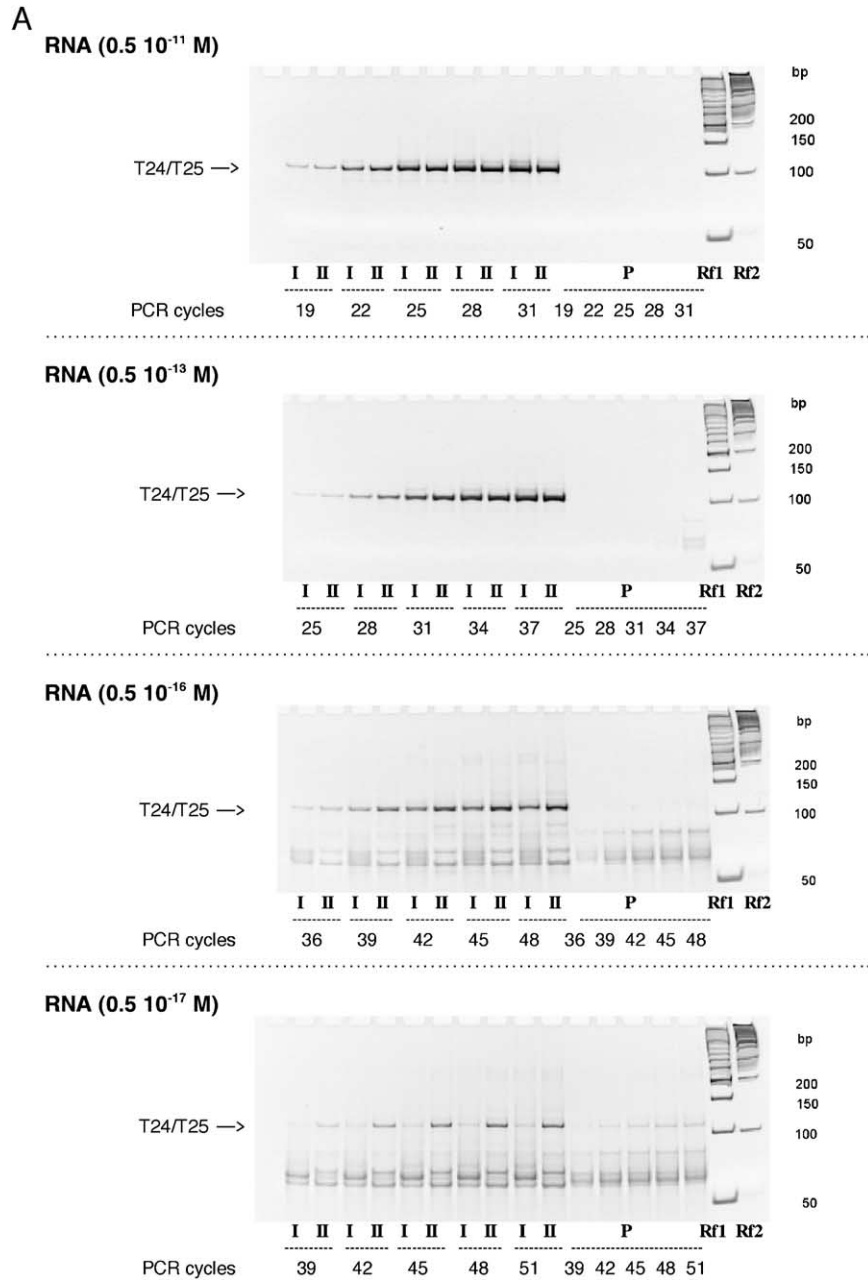


Fig. 8. Efficiency of DNA production of T25 and T24 in the RT-PCR reaction. This study was performed as a function of initial RNA concentration (0.5×10^{-11} , 0.5×10^{-13} , 0.5×10^{-16} and 0.5×10^{-17} M) used in the RT step. (A) After amplification by PCR, products were analyzed by electrophoresis on polyacrylamide 10% native gel (lanes I, T24 and lanes II, T25); sets of RT-PCR with primers and without RNA template were run as nonproductive T24, T25 DNA controls (lanes P); DNA ladders Rf1 and Rf2 are from Fermentas and Promega; gels were stained with ethidium bromide. (B) Histograms of T24 and T25 DNA produced for several PCR cycles. Quantification was obtained with the NIH software as pic area from band scannings. (For interpretation of the references to colour in this figure, refer to the web version of this article.)

have been reported and the importance of replication rate has been emphasized (Meyers et al., 2004).

In vitro selection for thermohalophilic RNAs produced rod-like structures, in which strong double-strand pairing appears to be the most important feature for stability. Our results highlight structures of family I that would resemble small viroid-like molecules of 90 nt. Viroïds are subviral pathogens with a compact rod-like secondary structure (246–574 nt) (Riesner and Gross, 1985; Diener, 2003). They do not code for protein and it has been suggested that they

represent “living fossils” of a precellular RNA world, somewhere at the frontiers of life. We analyzed the resistance of Potato Spindle Tuber Viroid (PSTVd) in the same conditions as T24 (data not shown). After 65 h at 80 °C in the presence of 2 M NaCl, the viroid was completely degraded. It is important to emphasize that unlike T24, PSTVd presents many small internal loops that weaken the secondary structure at high temperatures. Nevertheless, viroid could have collected from the RNA world a series of structural motifs. It should be pointed out that in contrast with the numerous RNA

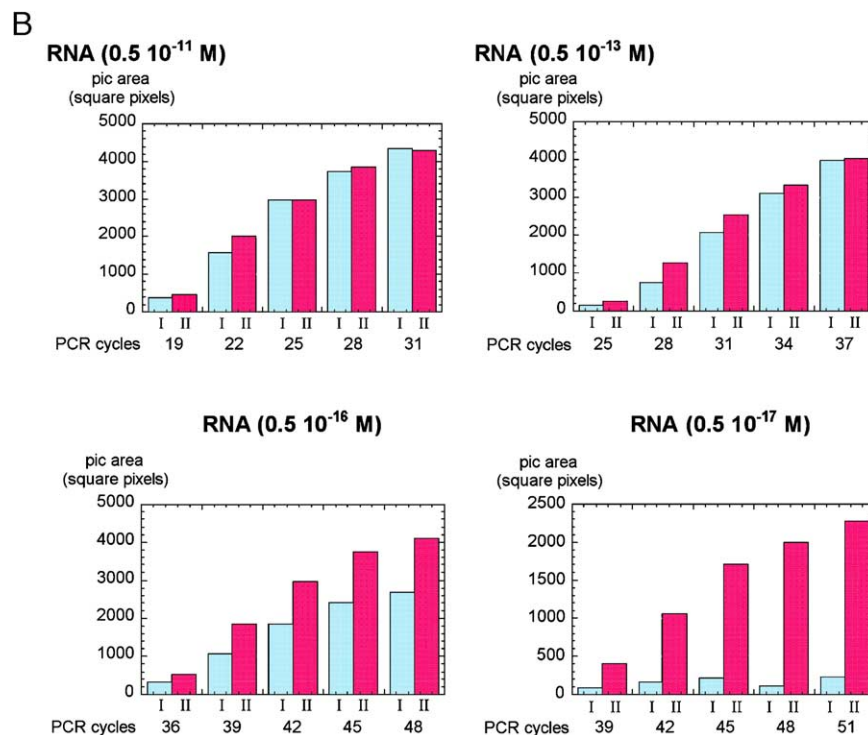


Fig. 8 (continued).

viruses that occur in higher eukaryotic organisms there are few (less than 10%) RNA bacteriophages. It would be interesting to search further for phages in halothermophilic archaeobacteria (Ackermann, 2003).

The detailed experiments presented here reveal that starting a selection with a RNA population at very low concentration can prevent loss of information at high temperatures and high salt concentrations for low stable structures. Sequences of family I could represent a very good means of storing sequence information and might well preserve their information contents. Many different representatives possess a high percentage of stable secondary structures, helical RNAs and stable hairpins. These molecules are badly involved in the enzymatic process of the RT-PCR at low RNA concentration. Family II with very closely related representatives, possess little secondary structure and can be degraded at high temperature, but the information content is such that amplification is easier. Low RNA concentration would favor family II amplification compared with family I, because of a better enzymatic uptake process in the RT-PCR steps.

The balance between survival and fecundity is a common pattern in evolutionary biology. Since fitness is a combination of these, it is not surprising that alternatives can equally fit with different combinations of fecundity and survival, which also lead to and increase diversity, a welcome outcome given the simple competitive tendency of molecular replicators. It is reassuring that a trade-off such as this one comes to light in a simple in vitro genetic experiment.

The results presented here raise the question of whether or not RNAs can respond to the combined selection pressure of

high temperature and functional requirements, since positive results in such a case would significantly broaden the applicability of the idea of a metabolically complex RNA world, in which ribo-organisms could occupy hot habitats. Our findings support the importance of salt in the protection of macromolecules against thermal degradation, and could be useful in the search for traces of life in ancient sediments and in planetary exploration.

Acknowledgements

We are grateful to Anne-Lise Haenni for valuable suggestions and careful reading and improvements of the manuscript. We also thank Detlev Riesner and Axel Schmitz for providing the PSTVd. This work was supported by grants from CNRS (Centre National de la Recherche Scientifique), University Paris VI and CNES (Centre National d'Etudes Spatiales), and GEOMEX (National French Research Program).

References

- Ackermann, H.W., 2003. Bacteriophage observations and evolution. *Res. Microbiol.* 154, 245–251.
- Bajji, A.C., Davis, D.R., 2002. Synthesis of the tRNA(Lys,3) anticodon stem-loop domain containing the hypermodified ms2t6A nucleoside. *J. Org. Chem.* 67, 5352–5358.
- Bevilacqua, J.M., Bevilacqua, P.C., 1998. Thermodynamic analysis of an RNA combinatorial library contained in a short hairpin. *Biochemistry* 3, 15877–15884.
- Chaput, J.C., Ichida, J.K., Szostak, J.W., 2003. DNA polymerase-mediated DNA synthesis on a TNA template. *J. Am. Chem. Soc.* 125, 856–857.

- Crick, F.H., 1968. The origin of the genetic code. *J. Mol. Biol.* 38, 367–379.
- Dale, T., Smith, R., Serra, M.J., 2000. A test of the model to predict unusually stable RNA hairpin loop stability. *RNA* 6, 608–615.
- Diener, T.O., 2003. Discovering viroids—a personal perspective. *Nat. Rev. Microbiol.* 1, 75–80.
- Di Giulio, M., 2000. The late stage of genetic code structuring took place at a high temperature. *Gene* 261, 189–195.
- Di Giulio, M., 2003a. The universal ancestor and the ancestor of bacteria were hyperthermophiles. *J. Mol. Evol.* 57, 721–730.
- Di Giulio, M., 2003b. The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. *J. Theor. Biol.* 221, 425–436.
- Dundas, I., 1998. Was the environment for primordial life hypersaline? *Extremophiles* 2, 375–377.
- Ferris, J.P., Hill Jr., A.R., Liu, R., Orgel, L.E., 1996. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 38, 59–61.
- Gesteland, R.F., Cech, T.R., Atkins, J.F. (Eds.), 1999. *The RNA world*, 2nd ed. Monograph, vol. 37. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Gilbert, W., 1986. *The RNA World*. Nature 319, 618.
- Gulyaev, A.P., van Batenburg, F.H., Pleij, C.W., 2002. Selective pressures on RNA hairpins in vivo and in vitro. *J. Mol. Evol.* 54, 1–8.
- Guo, F., Cech, T.R., 2002. Evolution of *Tetrahymena* ribozyme mutants with increased structural stability. *Nat. Struct. Biol.* 9, 855–861.
- Joyce, G.F., 2002. The antiquity of RNA-based evolution. *Nature* 418, 214–221.
- Joyce, G.F., Orgel, L.E., 1999. Prospects for understanding the origin of the RNA world, *The RNA World*, 2nd ed. Monograph, vol. 37. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 49–77.
- Juneau, K., Cech, T.R., 1999. In vitro selection of RNAs with increased tertiary structure stability. *RNA* 5, 1119–1129.
- Kargel, J.S., 1998. The salt of Europa. *Science* 280, 1211–1212.
- Knauth, L.P., 1998. Salinity history of the Earth's early ocean. *Nature* 395, 554–555.
- Kozlov, I.A., Politis, P.K., Pitsch, S., Herdewijn, P., Orgel, L.E., 1999. A highly enantio-selective hexitol nucleic acid template for nonenzymatic oligoguanilate synthesis. *J. Am. Chem. Soc.* 121, 1108–1109.
- Li, Y., Breaker, R.R., 1999. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. *J. Am. Chem. Soc.* 121, 5364–5372.
- Marguet, E., Forterre, P., 1994. DNA stability at temperatures typical for hyperthermophiles. *Nucleic Acids Res.* 22, 1681–1686.
- Marshall, K.A., Ellington, A.D., 2000. In vitro selection of RNA aptamers. *Methods Enzymol.* 318, 193–214.
- Meli, M., Vergne, J., Decout, J.L., Maurel, M.-C., 2002. Adenine-aptamer complexes: a bipartite RNA site that binds the adenine nucleic base. *J. Biol. Chem.* 277, 2104–2111.
- Meli, M., Vergne, J., Maurel, M.-C., 2003. In vitro selection of adenine-dependent hairpin ribozymes. *J. Biol. Chem.* 278, 9835–9842.
- Meyers, L.A., Lee, J.F., Cowperthwaite, M., Ellington, A.D., 2004. The robustness of naturally and artificially selected nucleic acid secondary structures. *J. Mol. Evol.* 58, 681–691.
- Orgel, L.E., 1968. Evolution of the genetic apparatus. *J. Mol. Biol.* 38, 381–393.
- Proctor, D.J., Schaak, J.E., Bevilacqua, J.M., Falzone, C.J., Bevilacqua, P.C., 2002. Isolation and characterization of a family of stable RNA tetraloops with the motif YNMG that participate in tertiary interactions. *Biochemistry* 41, 12062–12075.
- Riesner, D., Gross, H.J., 1985. Viroids. *Annu. Rev. Biochem.* 54, 531–564.
- Santini, G.P.H., Pakleza, C., Cognet, J.A.H., 2003. DNA tri- and tetra-loops and RNA tetra-loops hairpins fold as elastic biopolymer chains in agreement with PDB coordinates. *Nucleic Acids Res.* 31, 1086–1096.
- Schmidt, J.G., Nielsen, P.E., Orgel, L.E., 1997. Information transfer from peptide nucleic acids to RNA by template-directed syntheses. *Nucleic Acids Res.* 25, 4797–4802.
- Schultes, E.A., Hraber, P.T., LaBean, T.H., 1999. Estimating the contributions of selection and self-organization in RNA secondary structure. *J. Mol. Evol.* 49, 76–83.
- Schwartzman, D.W., Lineweaver, C.H., 2004. The hyperthermophilic origin of life revisited. *Biochem. Soc. Trans.* 32, 168–171.
- Shu, Z., Bevilacqua, P.C., 1999. Isolation and characterization of thermodynamically stable and unstable RNA hairpins from a tri-loop combinatorial library. *Biochemistry* 38, 15369–15379.
- Tehei, M., Franzetti, B., Maurel, M.-C., Vergne, J., Hountondji, C., Zaccai, G., 2002. The search for traces of life: the protective effect of salt on biological macromolecules. *Extremophiles* 6, 427–430.
- Tobé, S., Heams, T., Vergne, J., Herve, G., Maurel, M.-C., 2005. The catalytic mechanism of hairpin ribozyme studied by hydrostatic pressure. *Nucleic Acids Res.* 33, 2557–2564.
- Tuerk, C., Gold, L., 1990. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 505–510.
- Varani, G., 1995. Exceptionally stable nucleic acid hairpins. *Annu. Rev. Biophys. Biomol. Struct.* 24, 379–404.
- Vecenie, C.J., Serra, M.J., 2004. Stability of RNA hairpin loops closed by AU base pairs. *Biochemistry* 43, 11813–11817.
- Vreeland, R.H., Rosenzweig, W.D., Powers, D.W., 2000. Isolation of a 250 million-year-old halotolerant bacterium from a primary salt crystal. *Nature* 407, 897–900.
- Wächtershäuser, G., 1992. Groundworks for an evolutionary biochemistry: the iron-sulphur world. *Prog. Biophys. Mol. Biol.* 58, 85–201.
- Woese, C.R., 1967. *The Genetic Code: The Molecular Basis for Gene Expression*. Harper and Row, New York, pp. 156–160.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415.