The human RNA 3'-terminal phosphate cyclase is a member of a new family of proteins conserved in Eucarya, Bacteria and Archaea

Pascal Genschik¹, Eric Billy,
Michal Swianiewicz and Witold Filipowicz²

Friedrich Miescher-Institut, PO Box 2543, 4002 Basel, Switzerland

¹Present address: Institut de Biologie Moléculaire des Plantes du CNRS, 12 rue du Général Zimmer, 67084 Strasbourg, France

²Corresponding author

RNA 3'-terminal phosphate cyclase catalyses the ATP-dependent conversion of the 3'-phosphate to a 2',3'-cyclic phosphodiester at the end of RNA. The physiological function of the cyclase is not known, but the enzyme could be involved in the maintenance of cyclic ends in tRNA splicing intermediates or in the cyclization of the 3' end of U6 snRNA. In this work, we describe cloning of the human cyclase cDNA. The purified bacterially overexpressed protein underwent adenylation in the presence of [α -³²P]ATP and catalysed cyclization of the 3'-terminal phosphate in different RNA substrates, consistent with previous findings. Comparison of oligoribonucleotides and oligodeoxyribonucleotides of identical sequence demonstrated that the latter are ~500-fold poorer substrates for the enzyme. In Northern analysis, the cyclase was expressed in all analysed mammalian tissues and cell lines. Indirect immunofluorescence, performed with different transfected mammalian cell lines, showed that this protein is nuclear, with a diffuse nucleoplasmic localization. The sequence of the human cyclase has no apparent motifs in common with any proteins of known function. However, inspection of the databases identified proteins showing strong similarity to the enzyme, originating from as evolutionarily distant organisms as yeast, plants, the bacterium *Escherichia coli* and the archaeon *Methanococcus jannaschii*. The overexpressed *E.coli* protein has cyclase activity similar to that of the human enzyme. The conservation of the RNA 3'-terminal phosphate cyclase among Eucarya, Bacteria and Archaea argues that the enzyme performs an important function in RNA metabolism.

Keywords: adenylation/cyclic phosphate/RNA ligation/RNA modification/RNA processing

Introduction

RNA 3'-terminal phosphate cyclase, an enzyme that catalyses conversion of a 3'-phosphate group to the 2',3'-cyclic phosphodiester at the 3' end of RNA, has been identified in extracts of HeLa cells and *Xenopus* nuclei (Filipowicz and Shatkin, 1983; Filipowicz *et al.*, 1983). The HeLa cell cyclase has been purified and its mechanism of action studied (Filipowicz *et al.*, 1985; Reinberg *et al.*, 1985; Vicente and Filipowicz, 1988; reviewed by

Filipowicz and Vicente, 1990). The cyclization of the 3'-terminal phosphate, catalysed by the enzyme, occurs in three steps:

- (i) Enzyme + ATP → Enzyme-AMP + PP_i
- (ii) RNA-N^{3'}p + Enzyme-AMP → RNA-N^{3'}pp^{5'}A + Enzyme
- (iii) RNA-N^{3'}pp^{5'}A → RNA-N>p + AMP

Support for step (i) comes from identification of the covalent cyclase-AMP complex and the ability of 3'-phosphorylated RNA but not the 3'-OH-terminated RNA to release AMP from the preformed cyclase-AMP complex (Filipowicz *et al.*, 1985; Reinberg *et al.*, 1985; Vicente and Filipowicz, 1988). Step (ii) is inferred from experiments demonstrating accumulation of the RNA-N^{3'}pp^{5'}A molecules when the ribose at the RNA 3'-terminus is replaced with the 2'-deoxy- or 2'-O-methylribose (Filipowicz *et al.*, 1985). Reaction (iii) probably occurs non-enzymatically as the result of nucleophilic attack by the adjacent 2'-OH on the phosphorus in the phosphodiester linkage.

The biological role of the cyclase remains unknown, but the enzyme likely functions in some aspects of cellular RNA processing. The anabolic function of the 2',3'-cyclic phosphate in RNA first emerged when it was found that eukaryotic RNA ligases require 2',3'-cyclic ends for RNA ligation (Konarska *et al.*, 1981, 1982; Filipowicz and Shatkin, 1983; Filipowicz *et al.*, 1983; Furneaux *et al.*, 1983; Greer *et al.*, 1983a; Schwartz *et al.*, 1983; Perkins *et al.*, 1985; Pick *et al.*, 1986; reviewed by Filipowicz and Gross, 1984; Westaway and Abelson, 1995). This requirement applies to both the non-organellar RNA ligases characterized to date, one of which ligates RNA ends via the unusual 3',5'-phosphodiester, 2'-phosphomonoester linkage, while the other joins the ends via the regular 3',5'-phosphodiester (reviewed by Filipowicz and Gross, 1984; Phizicky and Greer, 1993; Westaway and Abelson, 1995). The involvement of the two RNA ligases in nuclear pre-tRNA splicing is well documented (Filipowicz *et al.*, 1983; Gegenheimer *et al.*, 1983; Greer *et al.*, 1983a; Laski *et al.*, 1983; Stange and Beier, 1987; Zillmann *et al.*, 1991; Phizicky *et al.*, 1992) but these enzymes might also function in the ligation of other natural RNA molecules such as virusoids and viroids (Branch *et al.*, 1982; Kikuchi *et al.*, 1982; Kiberstis *et al.*, 1985). This possibility is supported by the observation that, while yeast RNA ligase shows a strong preference for tRNA halves (Greer *et al.*, 1983a; Phizicky *et al.*, 1986), plant and mammalian RNA ligases also efficiently ligate artificial non-tRNA substrates (Konarska *et al.*, 1981, 1982; Filipowicz *et al.*, 1983; Furneaux *et al.*, 1983; Schwartz *et al.*, 1983; Perkins *et al.*, 1985; Pick *et al.*, 1986). Although the splicing endonucleases directly generate 5'-tRNA halves carrying 2',3'-cyclic phosphate, during tRNA splicing reactions (Peebles *et al.*, 1983; Gandini-Attardi *et al.*, 1985; Stange

and Beier, 1987; Rauhut *et al.*, 1990), it is possible that other putative substrates depend upon the action of RNA 3'-terminal phosphatase cyclase to form the cyclic phosphodiester ends. It is of interest, that the only cellular RNA ligase identified to date in bacteria also requires 2',3'-cyclic ends for ligation (Greer *et al.*, 1983b; Arn and Abelson, 1996).

Another finding uncovering the potential role of the 2',3'-cyclic phosphate in RNA metabolism was the demonstration that the U6 spliceosomal snRNA in species as diverse as humans, fruit fly and soybean has a cyclic 2',3'-phosphodiester at the 3'-terminus. The mechanism and enzymes responsible for this modification are not known, but RNA 3'-phosphate cyclase is one obvious candidate (Lund and Dahlberg, 1992).

With the long-term aim of determining the function of the cyclase in cellular RNA metabolism we have cloned the cDNA encoding the human cyclase. We have analysed the expression pattern and subcellular localization of the enzyme in mammalian cells. Furthermore, we report that proteins showing strong similarity to the human cyclase are not only present in eukaryotes, but also in Bacteria and Archaea. We show that the bacterially overexpressed *Escherichia coli* protein has RNA 3'-phosphate cyclase activity similar to that of the human enzyme.

Results

Cloning of the human cyclase cDNA

The cyclase was purified from HeLa cells by a modification of the procedure described previously (Filipowicz and Vicente, 1990; see Materials and methods). Purified protein was subjected to tryptic digestion and four peptides were microsequenced. Different combinations of degenerate oligodeoxyribonucleotide primers were used in order to clone, using a PCR approach, a cDNA encoding the enzyme. One amplified 668 bp DNA fragment contained an ORF encoding the peptides pep2 and pep3 (Figure 1) not used for the design of PCR primers. This partial cDNA was used as a probe to screen a HeLa λ gt11 cDNA library. Out of 1×10^6 recombinant phages, eight positive clones were isolated. Their inserts were analysed by restriction mapping and sequencing of the ends. The longest cDNA obtained from this screening [1349 nt, not including the poly(A) tail] extended to position 191 (Figure 1). The upstream coding and non-coding sequence of the clone was obtained by: (i) characterizing additional PCR-generated clones using the λ gt11 library DNA as a template; (ii) sequencing the human CpG island genomic clone (DDBJ/EMBL/GenBank accession number Z57130; Cross *et al.*, 1994) corresponding to the upstream region of the cyclase gene; and (iii) identifying an EST clone (Z42277) representing the upstream portion of the cDNA (see Materials and methods).

Conceptual translation of the cDNA predicted a 39.4 kDa protein of 366 amino acids with an isoelectric point of 7.8. All the microsequenced peptides are present in the deduced protein (Figure 1). The sequence surrounding the initiation codon (TCCCCCATGG) is similar to the consensus (GCCACCATGG) established for vertebrates mRNAs (Kozak, 1987). The 170-nt 5'-terminal leader sequence contains one additional ATG, present in a much less favourable context, CCAGGCATGA. Translation

initiated at this ATG would terminate five codons downstream (Figure 1).

Activity of the overexpressed human protein

The coding region of the cDNA was subcloned into an inducible expression vector, pGEX-2T, to express the cyclase as a fusion protein with glutathione *S*-transferase at the N-terminus. The recombinant protein was overproduced in *E. coli* and purified using glutathione-Sepharose 4B resin (Figure 2A). The purified fusion protein, migrating at ~60 kDa, still contains material with a lower molecular mass (~30 kDa) which likely corresponds to polypeptides prematurely terminated at rare codons present in the region encoding the N-terminal portion of the cyclase (see Figure 2A and its legend).

Several lines of evidence indicated that the overexpressed fusion protein has RNA 3'-terminal phosphatase cyclase activity. First, incubation with [α - 32 P]ATP resulted in the time-dependent labelling of the ~60-kDa protein (Figure 2B), consistent with previous findings demonstrating formation of a covalent cyclase-AMP complex, an intermediate in the cyclization reaction (see Introduction). Second, two different oligoribonucleotides, CCCCAC-CCCG $^{3'}$ p* and AAAAUAAAAG $^{3'}$ p*, both radiolabelled at the 3'-terminal phosphate, were tested as cyclization substrates (Figure 3A). Incubation of either oligoribonucleotide with increasing amounts of the fusion protein resulted in formation of molecules with a 3'-terminal phosphate resistant to the action of calf intestine phosphatase (CIP), a property expected for 2',3'-cyclic phosphodiester ends. The purine-rich and pyrimidine-rich substrates were modified with comparable efficiencies, in agreement with the previous findings that cyclase can utilize molecules with different sequences and base composition as substrates (Filipowicz *et al.*, 1983; Reinberg *et al.*, 1985; Filipowicz and Vicente, 1990). Third, it was found previously that, for the cyclase purified from HeLa cells, ATP is the best cofactor, but GTP, CTP and UTP (but not dATP) can also act as cofactors in the reaction, although much less efficiently (Vicente and Filipowicz, 1988). As shown in Figure 3B, the overexpressed protein has a similar nucleotide specificity. Fourth, TLC analysis was performed to demonstrate directly that incubation of AAAAUAAAAG $^{3'}$ p* with the overexpressed protein results in formation of the 2',3'-cyclic phosphate at the terminus. Following incubation with either the recombinant cyclase or the enzyme purified from HeLa cells, the oligoribonucleotide was digested with nuclease P1. In both cases, the resulting product co-migrated with the pG>p marker (Figure 4A, lanes 4 and 6) while treatment of the unreacted substrate released P_i (lane 2). The presumptive pG>p* from a reaction similar to that shown in Figure 4A, lane 6 was isolated and characterized further. As expected, it was resistant to digestion with nuclease P1 (Figure 4B, lane 1) but was converted to G>p* (lane 2) and pG $^{3'}$ p* (lane 3) by treatment with CIP and RNase T2, respectively. Treatment with RNase T2 followed by nuclease P1 liberated P_i (Figure 4B, lane 4). Digestion of pG>p* with the 2',3'-cyclic nucleotide 3'-phosphodiesterase from brain (CNPase) resulted in the formation of pG $^{2'}$ p* (lane 5) which, as expected, was resistant to treatment with nuclease P1 (lane 6). Presumptive pG>p* was also found to co-migrate with authentic pG>p during

```

      10      30      50      70
GCTGACTCCAGTGTCCCGAGAGGCGCCGCTTCTCCGCTTTCTCGTCAGGCTCCTGCAACCCAGGCATG
      90      110      130      150
AACCAAGGTTTCTGAACTACTGGGCGGGAGCCAACGTCTCTTCTTCTCCGCTCTGGCGGAGGCTTTGT
      170      190      210
CGCTGCGGGCTGGGCCCCAGGGTGTCCCCATGGCGGGGCGCGGGTGGAGGTTCGATGGCAGCATCATGG
      230      250      270      290
AAGGGGGCGGCCAGATCTCTGAGAGTCTCTACGGCCTTGAGCTGTCTCTAGGCCTCCCTTGC GGGTGCA
      310      330      350
G G G Q I L R V S T A L S C L L G L P L R V O
      370      390      410      430
GAAGATCCGAGCCGCGGAGCAGCCAGGCCTGAGGCCTCAACATTATCTGGACTGGAAATGATTCTGA
      450      470      490
K I R A G R S T P G L R P Q H L S G L E M I R
      510      530      550      570
GATTTGTGTGATGGGCAACTGGAGGGGGCAGAAATTGGCTCAACAGAAATAACCTTTACACCAGAGAAGA
      590      610      630
D L C D G Q L E G A E I G S T E I T F T P E K I
      650      670      690      710
TCAAAGGTGGAATCCACACAGCAGATACCAAGACAGCAGGGAGTGTGTGCCTCTTGATGCAGGTCTCAAT
      730      750      770
K G G I H T A D T K T A G S V C L L M Q V S M
      790      810      830      850
GCCGTGTGTTCTTCTTGCTGCTTCTCCATCAGAACTTCATTTGAAAGGTGGAATAATGCTGAAATGGCA
      870      890      910
P C V L F A A S P S E L H L K G G T N A E M A
      930      950      970      990
CCACAGATCGATTATACAGTGTGGTCTTCAAGCCAATTGTTGAAAAATTTGGTTTCATATTTAATTGTG
      1010      1030      1050
P Q I D Y T V M V F K P I V E K F G F I F N C D
      1070      1090      1110      1130
ACATTAAACAAGGGGATATTACCCAAAAGGGGGTGGTGAAGTGATTGTTGGAATGTCCAGGTTAAACA
      1150      1170      1190
I K T R G Y Y P K G G G E V I V R M S P V K Q
      1210      1230      1250      1270
ATTGAACCCCTATAAATTTAAGTGAAGCTGAGCGTGGCTGTGTGACTAAGATATATGGAAGAGCTTTCTGTTGCTGGT
      1290      1310      1330
L N P I N L T E R G C V T K I Y G R A F V A G
      1350      1370      1390
GTTTTGCCATTTAAGTAGCAAAAGATATGGCAGCGGCAGCAGTTAGATGCATCAGAAAGGAGATCCGGG
      1410      1430      1450      1470
V L P F K V A K D M A A A A V R C I R K E I R D
      1490      1510      1530
ATTTGTATGTTAATCCAGCCTGTTCAAGAACCTAAAGACCAAGCATTGGAATGGAATGGAATAAT
      1550      1570      1590
L Y V N I Q P V Q E P K D Q A F G N G N G I I
      1610      1630      1650      1670
AATTATGCTGAGACCTCCACTGGCTGTTTGTGTTGGATCATCGCTTGGTAAACGAGGTGTAATGCA
      1690      1710      1730      1750
I I A E T S T G C L F A G S S L G K R N A
      1770      1790      1810      1830
GACAAAGTTGGAATTGAAGCTGCCGAAATGCTATTAGCAATCTTAGACATGGTGGTACTGTGGATGAGT
      1850      1870      1890      1910
D K V G I E A C E M L L A N L R H G G T V D E Y
      1930      1950      1970      1990
ATCTGCAAGACCAGCTGATTGTTTTCATGGCATTAGCCAATGGAGTTTCCAGAATAAAAAACAGGACCAGT
      2010      2030      2050      2070
L Q D Q L I V F M A L A N G V S R I K T G P V
      2090      2110      2130      2150
TACACTCCATACGCAAAACCGGATACATTTTGTGTAACAAATAGCAAGGCTAAATTTATTGTGAAGAAA
      2170      2190      2210      2230
T L H T Q T A I H F A E Q I A K A K F I V K K
      2250      2270      2290      2310
TCAGAAGATGAAGAAGACGCCGCTAAAGATACTTATATTATTGAATGCCAAGGAATTGGGATGACAAATC
      2330      2350      2370      2390
S E D E E D A A K D T Y I I E C Q G I G M T N P
      2410      2430      2450      2470
CAAATCTATAGAGTATTTGCCTCTTAAATGATACCTCATTGATATATTGCACTATTTTATAAATACTATA
      2490      2510      2530      2550
N L *
      2570      2590      2610      2630
AAATAATGACTAGGAAGTAACTTATTAAAGGCTATGACTTAAATTTGAAGATGAAGTACAGTGTCTAGG
      2650      2670      2690      2710
      2730      2750      2770      2790
TTTGCTGAGAGGCTTCATTAAATTAATCTCACTTTGAATATCTCCTGAGAGATGGACAATGAAATATCA
      2810      2830      2850      2870
GTTGGTGGATATGTGTGATAGCTGATTTCAATATTGAAGTATTGAAATAAAATATTCTTTACACCTGAGA
      2890      2910      2930      2950
AAAAAAAA

```

Fig. 1. The nucleotide and deduced amino acid sequences of the human cDNA encoding RNA 3'-terminal phosphate cyclase. The short (5-amino-acid) ORF in the 5' leader region is underlined. Termination codons in the leader sequence are overlined. The putative polyadenylation signal is double underlined. Amino acid sequences of the sequenced peptides (pep1, pep2, pep3 and pep4) are shaded. It has not been directly established that position 1 indeed represents the start of transcription.

TLC in solvent B (data not shown). Furthermore, in another set of experiments, it was found that oligoribonucleotides containing p^{*}Cp ligated at the 3' end, N_np^{*}Cp, were converted into N_np^{*}C>p products upon incubation with the overexpressed protein (data not shown). All these results support the conclusion that the recombinant protein has RNA 3'-terminal phosphate cyclase activity.

Comparison of RNA and DNA molecules as substrates

We have previously shown that oligoribonucleotides containing terminal 2'-deoxy- or 2'-*O*-methylribose can be converted, upon incubation with the cyclase and ATP, into products bearing the 3'-terminal structures dN^{3'}pp^{5'}A and N^{m3'}pp^{5'}A, respectively (Filipowicz *et al.*, 1985). These

experiments led to the proposal that cyclization of the 3'-phosphate in RNA proceeds via formation of a terminal N^{3'}pp^{5'}A intermediate. However, these findings also raised the possibility that DNA molecules bearing a phosphate at the 3'-terminus might be physiological substrates for the enzyme.

Two different assays were used to compare the ability of the 3'-phosphorylated oligoribonucleotide AAAUAA-AAG^{3'}p, referred to as RNA^{3'}p, and the oligodeoxyribonucleotide of equivalent sequence AAAATAAAAG^{3'}p, referred to as DNA^{3'}p, to act as substrates for the enzyme. Using a competition assay (Figure 5A), it was found that an ~500-fold higher concentration of DNA^{3'}p than RNA^{3'}p (or another oligoribonucleotide, CCCCACCCG^{3'}p), is required to compete with the cyclization of the radio-

Fig. 2. Purification and labelling with [α - 32 P]ATP of the overexpressed human cyclase. (A) Purified and overexpressed human cyclase (GST fusion, 1.8 μ g; lane 2) analysed by SDS-PAGE and Coomassie blue staining. The ~30 kDa material most probably corresponds to polypeptides prematurely terminated at rare codons present in the upstream part of the cyclase cDNA. Since cyclase expression levels in bacteria were very low, further purification of the protein was not attempted. Extracts prepared from mock-transfected *E.coli* cells did not yield either the ~60 or ~30 kDa bands upon fractionation on the glutathione-Sepharose 4B column (data not shown). (B) Labelling of the cyclase-GST fusion protein with [α - 32 P]ATP. A total of 20 ng of cyclase was incubated under assay conditions in the presence of [α - 32 P]ATP for 1 min (lane 1), 30 min (lane 2), 1 h (lane 3), 3 h (lane 4) or 6 h (lane 5) at 25°C. Products of the reaction were analysed by SDS-PAGE and autoradiography. Positions of protein size markers (in kDa) are indicated.

labelled AAAAUAAAAG $^{3'}p^*$ substrate. A mixture of 3'-phosphorylated oligodeoxyribonucleotides [(dN) $_n$ pN $^{3'}p$, $n = 8-14$], obtained by a limited digestion of a synthetic 80-mer oligodeoxyribonucleotide with micrococcal nuclease was also a poor competitor in the reaction (Figure 5A). A 3'-hydroxyl-terminated oligodeoxyribonucleotide (DNA $^{3'}OH$) did not compete with the cyclization of RNA $^{3'}p^*$ even when added at 10 000-fold excess (Figure 5A), while a similar excess of RNA $^{3'}OH$ inhibited the reaction by ~40%. This small degree of inhibition observed in the presence of a large excess of RNA $^{3'}OH$ is most probably due to the traces of hydrolysis of the oligoribonucleotide, resulting in the formation of fragments bearing the 3'-phosphate group.

In the second assay, the oligoribo- and oligodeoxyribonucleotides were compared for their ability to release AMP from the preformed adenylylated enzyme complex (Reinberg *et al.*, 1985). The complex was formed by preincubation of the fusion protein with [α - 32 P]ATP. Incubations were then continued in the presence of increasing quantities of different oligonucleotides. Addition of 22 fmol RNA $^{3'}p$ decreased the amount of the complex by >95% (Figure 5B, lane b) while no complex was detected when 220 or 2200 fmol RNA $^{3'}p$ was added in the second incubation (Figure 5, lanes c and d). In contrast, incubation in the presence of 220 or 2200 fmol of either RNA $^{3'}OH$, DNA $^{3'}p$ or DNA $^{3'}OH$ (lanes e and f) did not result in the release of the label from the preformed cyclase-AMP complex. In the presence of still higher amounts (20 and 200 pmol), DNA $^{3'}p$ but not DNA $^{3'}OH$ resulted in AMP release from the complex (Figure 5B, lanes e and f). In another set of experiments we have directly demonstrated that prolonged incubation of 3'-phosphorylated oligo-

Fig. 3. Cyclization of the 3'-terminal phosphate catalysed by the overexpressed human cyclase measured by the CIP/Norit assay. (A) Comparison of CCCCACCCCG $^{3'}p^*$ and AAAAUAAAAG $^{3'}p^*$ substrates. (B) Effect of different NTPs. Except when indicated otherwise, assays were performed as described in Materials and methods with 90 (A) and 50 (B) fmol of the substrate added per reaction. Amounts of added cyclase in (A) are as indicated. In (B), the assays contained 600 pg of cyclase and 0.2 mM NTPs.

deoxynucleotides in the presence of an excess of the cyclase generates low amounts of products bearing a dN $^{3'}pp^{5'}A$ terminus (P.Genschik, W.Filipowicz and O.Vicente, unpublished results). Taken together, these results indicate that 3'-phosphorylated oligodeoxyribonucleotides are ~500-fold poorer substrates for the cyclase than oligoribonucleotides.

Intracellular localization of the cyclase

The intracellular localization of the cyclase was determined by an epitope-tagging approach combined with indirect immunofluorescence. The coding region of the cyclase cDNA was cloned in the vector pBact-myc to express the enzyme containing a *myc* epitope fused in frame at the N-terminus. The plasmid expressing the tagged protein was transfected into HeLa cells, rat glioma C6 cells and mouse embryonal carcinoma P19 cells. As a control, HeLa cells were transfected with a plasmid expressing the splicing factor ASF/SF2 (Manley and Tacke, 1996) fused in frame to the same epitope. The cells were processed for immunofluorescence microscopy, using a mouse anti-*myc* monoclonal antibody and FITC-conjugated goat anti-mouse antibody (Figure 6). Images were analysed with the help of the confocal laser scanning microscope. Indirect

Fig. 4. Analysis of the reaction products by TLC. Incubations with the cyclase were essentially as described in Materials and methods and contained 0.2 mM ATP, 60 fmol AAAAUAAAAG^{3'}p* and either 50 ng of partially purified HeLa cell cyclase [the poly(A)-Sephacryl fraction (Vicente and Filipowicz, 1988); panel A, lanes 3 and 4] or 40 ng of the overexpressed protein (panel A, lanes 5 and 6, and panel B). Incubations were for 20 min at 25°C. In (A) aliquots of the reactions were either applied to TLC without any additional treatment (lanes 1, 3 and 5) or were digested with nuclease P₁ (lanes 2, 4 and 6). (B) Putative pG²¹p*, obtained by digestion of the reacted AAAAUAAAAG^{3'}p* substrate with nuclease P₁ and isolated by cellulose TLC in solvent B, was subjected to treatment with the following enzymes: none (lane 1); CIP (lane 2); RNase T2 (lane 3); RNase T2 followed by nuclease P₁ (lane 4); CNPase (lane 5); CNPase followed by nuclease P₁ (lane 6). Chromatography was in solvent A. Positions of nucleotide markers are indicated.

immunofluorescence indicated that in HeLa and C6 cells, 98–99% of cyclase localizes to the nucleus and shows a diffuse distribution throughout the nucleoplasm with the protein being excluded from the nucleoli. The remaining 1–2% of fluorescence was seen in the cytoplasm (Figure 6, panels A–D; and the legend). The ratio of nuclear to cytoplasmic staining was independent of the amount of plasmid used for transfection or the time (24–48 h) at which cells were collected after transfection (data not shown). Assessment of the significance of the low cytoplasmic staining seen with the HeLa and C6 cells will require experiments with specific anti-cyclase antibodies. With the mouse P19 cells, the protein was found exclusively in the nucleoplasm and no cytoplasmic staining was observed (Figure 6, panel E). With all three cell lines, no signal was obtained in mock-transfected cells (data not shown) or in the non-transfected cells present in the same field as transfected cells (Figure 6, panels A–E). Likewise, no fluorescence was detectable when the anti-myc antibody was omitted (not shown). The diffuse nucleoplasmic staining seen in the cyclase cDNA-transfected cells was clearly different from the characteristic speckled staining observed with HeLa cells expressing the splicing factor ASF/SF2 (Figure 6, panel F; A.Krainer, personal communication). The monoclonal mouse antibody against the spliceosomal protein U2B', followed by FITC-conjugated goat anti-mouse antibody, also yielded a speckled immunofluorescence pattern similar to that seen for ASF/SF2 (data not shown).

Cyclase mRNA in different human tissues and cell lines

Expression of mRNA encoding the cyclase was determined in various cell lines (Figure 7A) and different human tissues (Figure 7B) by Northern blot analysis. Cyclase is expressed ubiquitously. Two hybridizing RNA species, of ~1.8 and 3 kb, were detected. The 1.8-kb mRNA corresponds in size to the cDNA characterized in this work; the

identity of the longer RNA is unknown. The ratio between the two RNAs varies among tissues and cell lines. The 3-kb transcript is present at a relatively low level in Namalwa cells, HeLa cells, and in heart and placenta (between 21% and 30% of total transcripts as determined by PhosphorImager quantification). Among the tissues analysed, the highest cyclase mRNA level was observed in skeletal muscle.

Cyclase is conserved from bacteria to humans

The sequence of the cloned human cyclase was used for database searches. Several ORFs encoding proteins of unknown function, but sharing significant similarity with the cloned protein, were identified in different organisms. These organisms include *Drosophila melanogaster*, *Caenorhabditis elegans*, *Schizosaccharomyces pombe*, *Saccharomyces cerevisiae*, *Toxoplasma gondii*, the bacterium *E.coli* and the archaeon *Methanococcus jannaschii*. Moreover, we have identified EST clones encoding cyclase-like proteins in mouse, *Arabidopsis thaliana* and zebrafish, and also an EST encoding another cyclase-like protein in humans (see legend to Figure 8). The human EST clone has been sequenced and the deduced protein encoded by it, although not full length, is included in the alignment shown in Figure 8. The two human proteins, referred to in Figure 8 as Hs1 and Hs2 show 30% identity and 52% similarity. The identified mouse ESTs encode counterparts of each of the two human proteins (data not shown). The human cyclase characterized in this work and the other cyclase-like proteins listed in Figure 8 have no apparent structural features or motifs in common with proteins of known function deposited in various databases (see Discussion).

The overexpressed *E.coli* protein has cyclase activity

The protein encoded by the *E.coli* gene having similarity with the human cyclase cDNA has been overexpressed in

Fig. 5. Activity of oligoribo- or oligodeoxyribonucleotides containing either 3'-p or 3'-OH termini in the competition (A) and AMP release (B) assays. (A) Cyclase assays were performed as described in Materials and methods. They contained 360 pg of the overexpressed cyclase, 40 fmol of AAAAUAAAAG^{3'}p* and different unlabelled oligonucleotides, acting as substrates or competitors, at molar excess over AAAAUAAAAG^{3'}p* as indicated. The 3'-OH-containing AAAUAAAAG (DNA^{3'}-OH) was obtained either by chemical synthesis (○) or by dephosphorylation of DNA^{3'}p with CIP (◇). (B) AMP release assays were performed as described in Materials and methods. Different amounts of substrates were added: none (lane N); 2 fmol (lane a); 22 fmol (lane b); 220 fmol (lane c); 2200 fmol (lane d); 20 pmol (lane e) and 200 pmol (lane f).

E. coli as a fusion protein with the 6×His tag at the C-terminus. The protein was purified using the Ni-NTA resin; its purity was >95% as judged after Coomassie blue staining of the gel (Figure 9A). Two oligoribonucleotides, CCCCCACCCG^{3'}p* and AAAAUAAAAG^{3'}p*, used for assaying activity of the human cyclase, were also found to act as substrates for the *E. coli* protein (Figure 9B). TLC analysis directly demonstrated that incubation with the overexpressed *E. coli* protein resulted in the cyclization of the 3'-terminal phosphate in the AAAAUAAAAG^{3'}p* (Figure 10, panels A and B) and AAAAUAAAAGp*^{3'}C^{3'}p (panel C) substrates. In addition to analyses with nuclease P1, RNase T2 and CNPase, similar to those shown in Figure 4, the putative pG>p* obtained by nuclease P1 digestion was characterized by treatment with RNase T1 and alkali. As expected, digestion with RNase T1 (Figure 10B, lanes 5–7) yielded results identical to those seen with RNase T2 (Figure 10B, lanes 2–4). Also as expected, mild alkaline treatment produced a mixture of pG^{2'}p* and pG^{3'}p*. Together, these results demonstrate that the

Fig. 6. Analysis of cyclase localization by indirect immunofluorescence in different cell lines. Cells were transfected with pBact-CYC-myc, encoding the *myc*-epitope-tagged cyclase (A–E) or with pMyc-ASF, expressing the *myc*-epitope-tagged splicing factor ASF/SF2 (F). The transfected cells were treated with the mouse anti-*myc* monoclonal antibody GE10 followed by the FITC-conjugated goat anti-mouse antibody. Slides were examined with a Zeiss Axiophot microscope and a Leica confocal scanning laser microscope. HeLa cells (A and B), C6 rat glioma cells (C and D) and mouse embryonal carcinoma P19 cells (E), expressing the *myc*-tagged cyclase. (F) HeLa cells expressing the *myc*-tagged ASF/SF2. Panels (B) and (D) represent higher intensity images of the fields shown in (A) and (C), respectively, in order to visualize low cytoplasmic staining. Similar overexposure of the fields shown in (E) and (F) did not reveal cytoplasmic fluorescence (not shown). For pBact-*myc*-transfected HeLa and C6 cells, the cytoplasmic fluorescence was calculated as 1.2% (average for eight cells) and 1.8% (average for five cells), respectively.

bacterial protein has RNA 3'-terminal phosphate cyclase activity.

Discussion

RNA 3'-terminal phosphate cyclase activity was previously demonstrated in nuclear extracts from HeLa cells and *Xenopus* oocytes (Filipowicz and Shatkin, 1983; Filipowicz *et al.*, 1983, 1985; Reinberg *et al.*, 1985; reviewed by Filipowicz and Vicente, 1990). Cloning of the cDNA encoding the human cyclase, described in this work, allowed us to identify cDNAs and/or genes originating from different organisms which encode proteins showing significant similarity with the human enzyme. These organisms include diverse eukaryotes such

Fig. 7. Levels of cyclase mRNA in different cell lines (A) and human tissues (B), as analysed by Northern blotting. (A) The following cell lines were used to isolate RNA: Namalwa (Burkitt lymphoma), lane 1; HeLa, lane 2; MCF7 (breast adenocarcinoma), lane 3; K562 (chronic myeloid leukaemia), lane 4; HUT78 and Jurkat (T-cell lymphomas), lanes 5 and 6, respectively; 293T (a transformed primary embryonic kidney line), lane 7. (B) RNA was isolated from the following tissues: heart (lane 1), brain (lane 2), placenta (lane 3), lung (lane 4), liver (lane 5), skeletal muscle (lane 6), kidney (lane 7) and pancreas (lane 8). Sizes of mRNAs (arrows) and markers are indicated.

as mammals, insects, worms, plants and yeast and also the bacterium *E. coli* and the archaeon *M. jannaschii*. Moreover, inspection of the EST clones deposited in the GenEMBL database, followed by additional sequence analyses, revealed that the human and mouse genomes contain at least one additional gene encoding the cyclase-like protein (Figure 8). Apart from the human enzyme, there is still no evidence that other eukaryotic proteins listed in Figure 8 represent RNA cyclases. However, as demonstrated in this work (Figures 9 and 10), the *E. coli* protein efficiently catalysed the conversion of the 3'-phosphate to a 2',3'-cyclic phosphodiester at the end of three different oligoribonucleotide substrates. The biochemical properties and substrate specificity of the bacterial protein are very similar to those of the human enzyme (our unpublished data). These results strongly suggest that other proteins listed in Figure 8 also have RNA cyclase activity. Taken together, the data indicate that RNA cyclase is a widespread enzyme conserved among three kingdoms: Eucarya, Bacteria and Archaea. However, no obvious homologues to the cyclase could be found in the streamlined genomes of *Haemophilus influenzae* and *Mycoplasma genitalium* (Fleischmann *et al.*, 1995; Fraser *et al.*, 1995).

Proteins belonging to the family of RNA cyclases listed in Figure 8 have no apparent structural features in common with other known proteins. In particular, no sequences corresponding to known nucleotide-binding motifs or RNA-binding domains could be identified. Previous analysis of the stability of the covalent cyclase-AMP complex suggested that AMP is linked to the protein via a phosphamide linkage, possibly involving the ϵ -amino group of a lysine (Reinberg *et al.*, 1985; Vicente and Filipowicz, 1988). However, it should be noted that no lysine is present at conserved position in all proteins aligned in Figure 8. With respect to the ability to transfer the nucleotidyl group from the protein-NMP intermediate to the terminal phosphate or pyrophosphate in nucleic acids, the cyclase resembles RNA and DNA ligases and capping enzymes (reviewed by Shuman and Schwer, 1995), but none of the sequence motifs shared by these enzymes

(Shuman and Schwer, 1995) is apparent in the RNA cyclase.

Hinton *et al.* (1982) have observed that in the absence of its natural 5'-phosphorylated substrate, T4 bacteriophage RNA ligase can inefficiently convert an oligoribonucleotide terminal 3'-phosphate to a 2',3'-cyclic form via a mechanism which is, most probably, similar to that of RNA cyclase. We have tested whether, in the absence of the 3'-phosphorylated end, human cyclase has the potential to activate the 5'-terminal phosphate in pAAAAUAAAG. No evidence of A^{5'}pp^{5'}A formation was found, even when a large excess of the enzyme was used. Similarly, no evidence of cyclase-catalysed inter- or intramolecular ligation of either 5'- or 3'-phosphorylated oligoribonucleotides was obtained (our unpublished results). Sequence comparisons between RNA cyclases and known RNA ligases (Rand and Gait, 1984; Phizicky *et al.*, 1986; Arn and Abelson, 1996) did not reveal significant similarities.

Previous experiments indicated that partially or highly purified HeLa cell cyclase can use a variety of 3'-phosphorylated RNA molecules as substrates, ranging from synthetic oligoribonucleotides such as AUGp, (Up)₁₀pGp and (Ap)_npAp to natural RNAs such as tRNAs, 5S rRNA or tobacco mosaic virus RNA fragments modified by ligation of pGp, pAp or pCp (Filipowicz *et al.*, 1983, 1985; Reinberg *et al.*, 1985; Vicente and Filipowicz, 1988; Filipowicz and Vicente, 1990). Consistent with this, the recombinant RNA cyclase was able to cyclize both purine-rich (AAAAUAAAAGp) and pyrimidine-rich (CCCCAC-CCCGp) oligoribonucleotides with comparable efficiency (Figure 3). Likewise, the 3'-phosphate in AAAAUAAA-GCp, CCCCCCCCCGp and AUGp underwent cyclization in the presence of the recombinant enzyme (data not shown). It was demonstrated previously that nucleoside 3'-phosphates and nucleoside 5',3'-diphosphates do not act as substrates for the cyclase (Filipowicz *et al.*, 1983; Reinberg *et al.*, 1985; Filipowicz and Vicente, 1990). In agreement with this, using competition and AMP release assays similar to those shown in Figure 5, we have

found that these compounds, as well as 3'-phosphorylated diribonucleotides, do not act as substrates for the recombinant enzyme (our unpublished results). Hence, although the cyclase has little sequence or terminal nucleotide specificity, it appears to require a trinucleotide as the minimal substrate length. More systematic studies are

We have shown previously that oligoribonucleotides containing terminal 2'-deoxy- or 2'-*O*-methylribose can be converted, in the presence of cyclase and ATP, into products bearing the 3'-terminal structures dN^{3'}pp^{5'}A and

Fig. 9. Purification and activity of the overexpressed *E. coli* cyclase. (A) Purification of the protein fused to the 6×His tag as monitored by SDS-PAGE. Lane 1, size markers; lane 2, 40 µg of extract from the induced *E. coli* culture; lane 3 and 4, 3 and 9 µg of the purified protein, respectively. (B) Comparison of CCCCCACCCG^{3'}p* and AAAAUAAAAG^{3'}p* substrates. Assays were performed as described in Materials and methods, with 50 fmol of the substrate added per reaction. For other details, see legends to Figures 2 and 3.

Fig. 10. Analysis of the reaction products by TLC. Incubations with the *E. coli* cyclase were essentially as described in Materials and methods and contained 0.2 mM ATP, 60 fmol of either AAAAUAAAAG^{3'}p* (A and B) or AAAAUAAAAGp*Cp (C) and 6 ng of the overexpressed *E. coli* protein. Incubations were for 20 min at 25°C. (A) Aliquots of the reactions were either applied to TLC without any additional treatment (lane 1) or were digested with nuclease P1 (lane 2), nuclease P1 followed by CIP (lane 3) or nuclease P1 followed by RNase T2 (lane 4). (B) Putative pG>p*, obtained by digestion of the reacted AAAAUAAAAG^{3'}p* with nuclease P1 and isolated by the cellulose TLC in solvent B, was subjected to treatment with the following enzymes: none (lane 1); RNase T2 (lane 2); RNase T2 followed by nuclease P1 (lane 3) or CIP (lane 4); RNase T1 (lane 5); RNase T1 followed by nuclease P1 (lane 6) or CIP (lane 7); CNPase (lane 8); CNPase followed by nuclease P1 (lane 9) or CIP (lane 10). Lane 11, partial alkaline hydrolysis. (C) Aliquots of the cyclase reaction containing AAAAUAAAAGp*Cp as substrate were either applied to TLC without any additional treatment (lane 1) or were digested with either nuclease P1 (lane 2), nuclease P1 followed by RNase T2 (lane 3) or nuclease P1 followed by CNPase (lane 4). Chromatography was in solvent A. Positions of nucleotide markers are indicated.

Nm^{3'}pp^{5'}A, respectively (Filipowicz *et al.*, 1985). These findings raised the possibility that 3'-phosphorylated DNA molecules rather than, or in addition to, RNA might be the physiological substrates for the cyclase. Although, to our knowledge, known DNA recombination, repair or ligation pathways do not involve 3'-phosphorylated intermediates (for review, see Kornberg and Baker, 1991), such molecules can be generated under some conditions *in vitro* (Kimball *et al.*, 1993; Christiansen *et al.*, 1994; Latham and Lloyd, 1995; Bhagwat and Gerlt, 1996). For example, eukaryotic topoisomerase I, which transiently attaches to the 3'-phosphate DNA through a phosphoester linkage, can be inefficiently displaced by water at slightly alkaline conditions (pH 7.5–10), yielding 3'-phosphorylated molecules as products (Christiansen *et al.*, 1994). The Flp site-specific recombinase from *S. cerevisiae* is another enzyme which transiently reacts with the 3'-

phosphoryl group at the site of the breaks. In the presence of hydrogen peroxide, the Flp target phosphodiester in DNA can be cleaved hydrolytically, generating 3'-phosphate and 5'-OH ends (Kimball *et al.*, 1993; Sadowski, 1995). Activation of the DNA 3'-phosphate to dN^{3'}pp^{5'}A (catalysed by the cyclase), followed by the nucleophilic attack of the 5'-hydroxyl (catalysed by a hypothetical DNA ligase), could represent a pathway repairing breaks similar to those described above.

To address the possibility that cyclase participates in DNA- rather than RNA-related transactions, we have compared the activity of oligoribo- and oligodeoxyribonucleotides of identical sequence, using competition and AMP release assays. These experiments revealed that oligoribonucleotides are ~500-fold better substrates for the human cyclase than oligodeoxyribonucleotides (Figure 5). Although these results do not exclude the possibility

that DNA molecules of a particular sequence or structure act as substrates for the enzyme *in vivo*, they make it rather unlikely. The findings that, in *in vitro* assays, oligoribonucleotides are much better substrates than oligodeoxyribonucleotides (Figure 5) and that the 3'-phosphate in mono- or diribonucleotides does not undergo cyclization (Filipowicz *et al.*, 1983; Reinberg *et al.*, 1985; Vicente and Filipowicz, 1988; see also above) strongly argue that 3'-phosphate-terminated RNA molecules are indeed the natural substrates of the cyclase.

The physiological role of the cyclase remains unknown. The predominantly nucleoplasmic localization of the enzyme in mammalian cells (Figure 6) suggests that it is involved in RNA processing or other RNA metabolic reactions taking place in the nucleus. Two possible functions can be envisaged. The enzyme might be responsible for cyclization of the 3'-terminal phosphate in U6 snRNA. Lund and Dahlberg (1992) have found that U6 RNA in most eukaryotes investigated contains a 2',3'-cyclic phosphate end. The mechanism and enzymes responsible for this modification are unknown but it has been proposed that conversion of U6 RNA from the form containing the 3'-terminal oligouridylyl extension to that bearing the cyclic phosphate occurs within the spliceosome concurrent with the pre-mRNA splicing reaction (Tazi *et al.*, 1993). RNA ligases involved in pre-tRNA splicing in eukaryotes require 5'-tRNA half molecules terminated with cyclic phosphate (see Introduction). Although it has been shown, for both yeast and vertebrates, that the 2',3'-cyclic phosphate in 5'-tRNA halves is produced as a direct result of cleavage by the splicing endonuclease (Peebles *et al.*, 1983; Gandini-Attardi *et al.*, 1985; Rauhut *et al.*, 1990), it is possible that the cyclase functions to regenerate cyclic ends in 3'-monoester-terminated molecules formed by the action of decyclizing phosphodiesterases. Alternatively, the cyclase could be responsible for cyclizing the 3'-phosphate in substrates other than tRNA halves (see Introduction).

RNA ligases requiring the 2',3'-cyclic phosphate for ligation are not confined to eukaryotes. Greer *et al.* (1983b) have identified an RNA ligase in *E. coli* and other bacteria which ligates the 2',3'-cyclic phosphate and 5'-hydroxyl ends via the 2',5'-phosphodiester linkage. The *E. coli* enzyme appears to be specific for ligation of tRNA halves (Arn and Abelson, 1996) but its physiological substrates and the mechanism of production of their cyclic ends remain unknown. tRNA genes in Archaea contain introns which structurally resemble nuclear tRNA gene introns (reviewed by Westaway and Abelson, 1995). Similarly to eukaryotes, the tRNA-splicing endonuclease of the archaeon *Halobacterium volcanii* produces 5' half molecules containing the 2',3'-cyclic phosphate, but the requirements for ligation of tRNA halves in Archaea have not been characterized (Thompson and Daniels, 1988). The processing of the 23S rRNA intron in the archaeon *Desulfurococcus mobilis* shares many properties with the processing of tRNA gene introns. Interestingly, cleavage of the *D. mobilis* pre-rRNA by the endonuclease appears to generate splicing intermediates bearing the 3'-phosphomonoester end (Kjems and Garrett, 1988). It is not known whether the terminal phosphate has to undergo cyclization before the ligation step.

Finally, it is possible that the cyclase is not a component

of an RNA ligation pathway but modifies the RNA 3' end for a different purpose, for example, activation of RNA for exonucleolytic degradation. It has been shown that the presence of a terminal 3'-phosphate makes an oligonucleotide largely resistant to 3'→5' exonucleolytic degradation by snake venom phosphodiesterase. Consistent with the interpretation that this effect is conferred by a strong negative charge, oligonucleotides bearing a 2',3'-cyclic phosphate were found to be only marginally less active than 3'-OH-terminated substrates (Richards and Laskowski, 1969; Laskowski, 1971). Identification of physiological substrates of the cyclase and its function in cellular RNA metabolism are the subjects of current experimentation.

Materials and methods

Cloning of the human cyclase

Protein purification and peptide sequencing. The cyclase was purified from HeLa cells according to established protocols (Vicente and Filipowicz, 1988) with some modifications. (i) The cyclase was eluted from the heparin-Sepharose column (step 3; Vicente and Filipowicz, 1988) with a linear gradient of 75–450 mM NaCl. Fractions corresponding to pool a (Filipowicz *et al.*, 1985), eluting at 260 mM NaCl, were collected and applied, after dialysis, to a poly(A)-Sepharose column (Vicente and Filipowicz, 1988). (ii) The mono-S step (step 5) was omitted. The poly(A)-Sepharose fraction, concentrated and dialysed against buffer C containing 75 mM NaCl, was directly applied to a Blue-Sepharose column. Material eluting at 0.6 M NaCl (fraction BS-600; Vicente and Filipowicz, 1988) was successively concentrated using Centricon-30 and Microcon-10 filters (Amicon). A total of 10 µg of the protein was separated in 10% SDS-PAGE. Proteins were blotted onto nitrocellulose (Schleicher & Schuell) and stained with Ponceau S. A 39 kDa band, representing the cyclase, was excised and treated with trypsin. Proteolytic peptides were resolved by HPLC and sequenced by Dr W.S. Lane (Harvard MicroChem, Cambridge, MA). Four peptides were sequenced (*pep1* VEVDSIMEGGGQIL; *pep2* GYYPK; *pep3* QLNPINLTER; *pep4* DLYVNIQPVQE).

Cloning of the cyclase cDNA

Partial sequence of the cyclase cDNA was obtained by two consecutive PCR amplifications, using DNA prepared from a λgt11 human HeLa cell cDNA library (Clontech) as a template. For the first PCR, oligonucleotides 1 (CACATGGCTGAATATCGAC, corresponding to the border sequence of the λgt11 *EcoRI* cloning site) and 2 [TC(C/T)TG(C/G)AC(A/G/T)GG(C/T)TG(A/G)AT(A/G)TT], representing a mixture of 96 oligomers complementary to the sequence encoding peptide NIQPVQE, a fragment of *pep4*, were used as upstream and downstream primers, respectively. A 100 µl PCR contained 1 µM oligonucleotide 1, 4 µM oligonucleotide 2, 200 ng of the library DNA, 250 µM of each of the four dNTPs and 2.5 U *Taq* DNA polymerase in Perkin Elmer buffer. Thirty cycles (denaturation at 94°C for 40 s, annealing at 45°C for 1 min, and extension at 70°C for 1.5 min) were performed. For the second PCR reaction, oligonucleotide 3 [GA(C/T)GG(A/C/G/T)TC(A/C/T)AT(A/C/T)ATGGA(A/G)GG], a mixture of 144 oligomers coding for the sequence DGSIMEG of *pep1* and oligonucleotide 4 [GGCTG(A/G)AT(A/G)TT(A/C/G)AC(A/G)TA(C/G)AG(A/G)TC], overlapping with oligonucleotide 2 and representing a mixture of 96 oligomers coding for the sequence DLYVNIQ of *pep4*, were used as primers. The 100 µl PCR contained 4 µM primers, 250 µM dNTPs, 2.5 U *Taq* DNA polymerase and 8 µl of the first PCR. Thirty cycles (40 s at 94°C, 1 min at 52°C, 1.5 min at 72°C) were performed. A 667-bp amplified DNA fragment (positions 195–862, Figure 1) was subcloned into the *SmaI* site of pBluescribe (Stratagene) and sequenced by the dideoxy chain termination method. The fragment contained an ORF encoding the peptides *pep2* and *pep3*. The PCR amplification scheme described above represents the only successful reaction of many similar ones tested with other combinations of primers.

To obtain a longer cDNA, the HeLa λgt11 cDNA library was screened with the PCR-amplified fragment as a probe. Hybridizations were at 42°C in 5×SSPE buffer (SSPE is 0.18 M NaCl, 10 mM NaH₂PO₄, 1 mM Na₂-EDTA, pH 7.7) containing 50% formamide, 5×Denhardt's solution (100×Denhardt's solution is 2% Ficoll, 2% PVP, 2% BSA),

1% SDS and 50 µg/ml denatured salmon sperm DNA. Eight clones were isolated after screening 1×10^6 recombinant phages. The inserts were subcloned in pBluescript II and analysed by restriction mapping and sequencing of the ends. The longest clones were sequenced on both strands. The longest clone extended to position 191 (Figure 1), and was thus incomplete at the 5' end. The upstream sequence of the cyclase cDNA was obtained by a PCR approach using oligonucleotide 5 (CCCTTTTGGGTAATATCC, positions 662–649 in Figure 1), oligonucleotide 6 (AACTGGTAATGGTATCC, corresponding to the border sequence of the *EcoRI* λ gt11 cloning site) and 200 ng of the λ gt11 library DNA. PCRs, cloning of the amplified fragments and sequencing were performed as described above. The longest amplified fragment extended to position 1 of the sequence shown in Figure 1. The authenticity of the human-generated sequence was confirmed by identification of the human cyclase EST clone (DDBJ/EMBL/GenBank accession number Z42277), which starts at position 93 (Figure 1), and by sequencing the CpG island (Cross *et al.*, 1994) human genomic clone (Z57130) which represents the upstream portion of the cyclase gene. The 740-bp sequenced region spans the putative promoter (62 bp), 5'-terminal leader (170 bp) and the coding region which is interrupted by two introns located at positions 215/216 and 316/317 (Figure 1). Sequences of the cyclase cDNA and the genomic CpG island clone are deposited in the EMBL, GenBank and DDBJ databases under accession numbers Y11651 and Y11652, respectively.

Cloning of the *E. coli* cyclase

Two neighbouring ORFs of unknown function, encoded by the *E. coli* K12 chromosome in a region adjacent to *malT* (DDBJ/EMBL/GenBank accession number U18997), were identified in the database search using the human cyclase protein sequence as a query. The region of 1047 bp (positions 336558–337605 in U18997), covering the two ORFs was PCR-amplified using 200 ng of purified *E. coli* K12 genomic DNA (Ausubel *et al.*, 1990) as a template and appropriate oligodeoxynucleotides as primers and cloned into pBluescribe. Resequencing of the insert on both strands revealed an additional C residue between G337194 and A337195, and a sequence GCCGC instead of CGCCG at positions 337219–337223. To further confirm these changes, the cyclase gene fragment was PCR-amplified using DNA prepared from the λ clone DD765 (kindly provided by Drs G.Plunkett and F.Blattner, University of Wisconsin, Madison) as a template and the *Pfu* DNA polymerase (Stratagene). Two clones originating from two independent PCRs were sequenced on both strands, yielding identical results. The corrected sequence encodes a single ORF showing 32% identity and 42% similarity to the human RNA cyclase (Figure 8).

Northern blot analysis

Northern blot analysis was performed using formaldehyde-agarose gels and 15 µg of total RNA isolated (Ausubel *et al.*, 1990) from various cell lines. The RNA was blotted onto GeneScreen membranes and UV cross-linked. The blot containing RNAs originating from different human organs was purchased from Clontech. The cDNA fragment (positions 191–1210, Figure 1A) was used as the cyclase probe. The human β -actin cDNA, furnished by Clontech, was used as a control. The probes were labelled with [α - 32 P]dCTP (3000 Ci/mmol, Amersham) by the random priming method (Feinberg and Vogelstein, 1983). The hybridization was for 16 h at 42°C in 5×SSPE containing 50% formamide, 10% dextran sulfate, 1% SDS and 50 µg/ml denatured salmon sperm DNA. The blots were washed in 2×SSC (SSC is 0.15 M NaCl, 15 mM Na₂citrate) and 0.1% SDS for 30 min at 42°C, and in 0.2×SSC and 0.1% SDS for 30 min at 42°C and then at 60°C.

Overexpression and purification of the human and *E. coli* cyclases

Human RNA cyclase. *Bam*HI sites were introduced on the 5' and 3' side of the cyclase-coding sequence using PCR-based site-directed mutagenesis. The *Bam*HI–*Bam*HI fragment was cloned into the pGEX-2T vector (Pharmacia) for expression in the *E. coli* strain BL161. In the pGEX-2Tyc construct, the *Schistosoma japonicum* glutathione S-transferase is placed in frame at the N-terminus of the fusion protein. The best, though still quite low, yield was obtained by inducing a 0.7 OD₆₀₀ culture (grown at 37°C) with 1 mM IPTG for 2 h at 30°C. The protein remained soluble during expression in *E. coli* and was purified in the native form under non-denaturing conditions on glutathione–Sephadex 4B resin (Pharmacia) following the manufacturer's protocol. The protein was then applied to a 10 ml Sephadex G-25 column equilibrated and eluted with 30 mM HEPES–KOH, pH 7.6, 0.1 mM EDTA, 0.5 mM DTT, 5% (v/v) glycerol, 0.01% Triton X-100 and 10 µM

PMSF. It has proven impossible to express the cyclase as a 6×His-tagged protein using either pQE-60 (Qiagen) or pET11d (Novagen) as vectors. The protein concentration was measured by the method of Bradford using the reagent obtained from Bio-Rad and BSA as a standard.

***E. coli* RNA cyclase.** *Nco*I and *Bam*HI sites were introduced 5' and 3', respectively, of the cyclase coding sequence by PCR-based site-directed mutagenesis. The *Nco*I–*Bam*HI fragment was cloned into the pET-11d vector to produce a recombinant protein containing a sequence GSHHHHHH at the C-terminus. Overexpression was performed in *E. coli* strain BL21(DE3)pLysS. The protein remained soluble and could be purified in the native form under non-denaturing conditions using the Ni–NTA resin according to the protocol provided by Qiagen. The purified cyclase was applied to a Sephadex G-25 column as described above.

Assays of cyclase and chromatography

Preparation of substrates. Two oligoribonucleotides CCCCACCCCG and AAAAUAAAAG were synthesized and purified by HPLC by MWG-Biotech (Munich). They were 3'-terminally labelled using [5'- 32 P]pCp (*pCp) and T4 RNA ligase to produce (Np)₃Gp*pCp. Reaction mixtures (15 µl) contained 70 mM HEPES–KOH, pH 8.3, 10 mM MgCl₂, 3 mM dithiothreitol, 10% glycerol, 10% dimethylsulphoxide, 40 µM ATP, 3.3 µM p*pCp (specific activity 1000 Ci/mmol), 0.7 µg of oligoribonucleotides and 6 U T4 RNA ligase (New England Biolabs). After 14 h at 4°C, samples were diluted with 60 µl of 50 mM HEPES–KOH, pH 7.6, containing 3 mM EDTA, and incubated at 37°C for 1 h with 5 U of RNase T1. SDS was then added to a final concentration of 0.1% and the samples extracted with phenol/chloroform/isoamyl alcohol. The substrates, corresponding to (Np)₃Gp*, were purified on a Sephadex G-25 (fine) spin column pre-equilibrated with 20 mM ammonium acetate, pH 5.5. Aliquots of radioactive substrates were analysed by digestion with RNase T2, nuclease P1 or CIP, followed by TLC on cellulose plates in solvent A. Over 90% of the label was always present as the terminal G³p*. (Np)₃Gp*pCp were prepared as described above, except that the RNase T1 digestion was omitted.

Preparation of the competitors

Preparation of unlabelled oligoribonucleotides AAAAUAAAAG³p (referred to as RNA³p) and CCCCACCCCG³p from the 3'-OH-terminated counterparts was similar to that for the substrates labelled at the 3'-terminal phosphate (see above), except that ligation reactions contained 3 mM pCp. A small amount of [5'- 32 P]pCp (final specific activity 0.33 Ci/mmol) was included to calculate the reaction yields. Approximately 60% of the input oligonucleotide was found in the 3'-phosphorylated form following the ligation and RNase T1 digestion reactions. AAAAUAAAAG³OH (see above) used for competition experiments was additionally purified by electrophoresis in an 8 M urea/20% polyacrylamide gel.

The 3'-phosphorylated oligodeoxynucleotide AAAATAAAAAG³p (referred to as DNA³p) was a gift from Dr J.Hall (Central Research Laboratories, Ciba, Basel). This oligonucleotide was prepared by standard phosphoramidite chemistry using a modified solid support in an analogous method to that reported by Efimov *et al.* (1983). The oligonucleotide was purified by HPLC and analysed by mass spectroscopy and capillary gel electrophoresis. The 3'-OH-containing AAAATAAAAAG (DNA³OH) was either directly synthesized on an Applied Biosystems synthesizer, followed by purification by PAGE, or obtained by dephosphorylation of AAAATAAAAAG³p with CIP. The (dN)_npN³p, representing a mixture of 3'-phosphorylated oligodeoxyribonucleotides ($n = 8$ –14), was obtained by limited digestion of the synthetic 80-mer oligodeoxynucleotide (CCGCTAACTGGTACCCGATGTT GGCCTAGCTTCACTAGTA-ACCGGTGCTCATCGTAGCTTGCTAACGATGGATTGTC) with micrococcal nuclease (Pharmacia), followed by fractionation on an 8 M urea–20% polyacrylamide gel. Calculation of the molar concentration is based on an average length of 11 nt.

Cyclase assays and TLC analysis

Cyclase activity was assayed by the Norit method as described elsewhere (Filipowicz and Vicente, 1990). Unless indicated otherwise, the 10 µl assays contained 40 fmol of the substrate and incubations were for 20 min at 25°C. Other details are indicated in the figure legends. Digestions with nuclease P1 (Gibco-BRL), RNases T1 and T2 (Calbiochem) and the 2',3'-cyclic nucleotide, 3'-phosphodiesterase (CNPase) from bovine brain (Sigma) were performed as described previously (Konarska *et al.*, 1981). The TLC was in solvent A [saturated (NH₄)₂SO₄, 3 M Na acetate, isopropyl alcohol (80:6:2)] or B [isobutyric acid/conc. NH₃/H₂O (66:1:33)].

Labelling of the cyclase with [α - 32 P]ATP and AMP release assays

Reactions (15 μ l) containing 20 ng of cyclase, 22 μ M [α - 32 P]ATP (specific activity 89 Ci/mmol) were incubated at 25°C under standard cyclase assay conditions, except that the (Np)_nGp* substrate was omitted. In the release assays, 2.5 ng of the cyclase was first adenylated under standard conditions in the presence of 2.1 mM [α - 32 P]ATP (specific activity 800 Ci/mmol), with no substrate added, for 3 h at 25°C. The reactions were then diluted 7-fold with the cyclase assay buffer, different amounts of substrates were added and incubations continued for 15 min at 25°C. The reactions were analysed by SDS-PAGE and autoradiography. Immediately before application to the gel, samples were supplemented with unlabelled ATP (final concentration 10 mM).

Cell transfections and indirect immunofluorescence

To construct pBact-CYC-myc, the *Eco*RI sites were introduced 5' and 3' of the cyclase coding sequence by PCR-based site-directed mutagenesis using oligonucleotides 7 (GAGGAGAAAGAATTCATGGCGGGGCC-GTGG) and 8 (AGTGGTGATGGTGGAA TTCCTATAGATTG) as the upstream and downstream primers, respectively. The *Eco*RI-*Eco*RI fragment was cloned into the pBact-myc vector (Cravchik and Matus, 1993) for expression of the myc-epitope-tagged cyclase in transfected cells. The plasmid expressing the myc-epitope-tagged ASF/SF2 splicing factor (pMyc-ASF) was kindly supplied by Drs P.Kreivi and A.Lamond, University of Dundee, UK.

Cells were cultured in DME medium (Sambrook *et al.*, 1989) with the addition of 10% fetal calf serum in a 5% CO₂ humidified atmosphere at 37°C. Cells were transfected with pBact-CYC-myc (2–10 μ g) and pMyc-ASF (10 μ g) using the calcium phosphate DNA precipitate transformation method (Chen and Okayama, 1987), cultured for 24 or 48 h in a 5% CO₂ humidified atmosphere and fixed with 3% paraformaldehyde in PBS (Sambrook *et al.*, 1989) at 20°C for 20 min. Cells were treated for 10 min with 0.5% Triton X-100 solution in PBS, and with 5% non-immune goat serum in PBS for 20 min. The slides were then treated with either mouse monoclonal antibody GE10 raised against the human *c-myc* epitope (Evan *et al.*, 1985) and obtained from the European Collection of Cell Cultures (Porton Down, UK) or the mouse monoclonal anti-U2B' antibody (Organon Teknica-Cappel, Belgium). A FITC-conjugated goat anti-mouse antibody [AffiniPure F(ab')₂ fragment; Jackson ImmunoResearch Laboratories] was used as a secondary antibody. Samples were examined with a Zeiss Axiophot microscope and a Leica confocal scanning laser microscope using a 63 $\times$ objective. Images were recorded using Leica software (SCANware 4.2) provided with the system, and analysed with Imaris software on a Silicon Graphics work station. Quantification of the FITC signals was performed on a PowerMac 9500 computer using the public domain image analysis program V-1.57 developed at the US National Institutes of Health and available from the Internet by anonymous FTP from zippy.nimh.nih.gov.

Acknowledgements

We gratefully acknowledge the gifts of plasmids or phages from P.Kreivi and A.Lamond (pMyc-ASF construct), J.Takeda (Hs2 EST clone), G.Plunkett and F.Blattner (λ phage DD765), the UK HGMP Resource Center, Cambridge (CpG island clone). We also thank P.Matthias for providing the Northern blot of cell culture RNA, J.Hall for synthesis of AAAATAAAAG³p, T.Pohl for providing sequence of the yeast gene, K.Drabikowski for the sequence alignments, S.Kaech for advice on immunofluorescence during the localization experiments and G.Thomas, H.Rothnie and F.Dragon for critical reading of the manuscript. P.G. was supported by a fellowship from the European Community.

References

Arn,E.A. and Abelson,J.N. (1996) The 2'-5' RNA ligase of *Escherichia coli*. Purification, cloning and genomic disruption. *J. Biol. Chem.*, **269**, 31145–31153.

Ausubel,F., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1990) *Current Protocols in Molecular Biology*. Greene Publishing, New York.

Bhagwat,M. and Gerlt,J.A. (1996) 3'- and 5'-strand cleavage reactions catalyzed by the Fpg protein from *Escherichia coli* occur via successive β - and δ -elimination mechanisms, respectively. *Biochemistry*, **35**, 659–665.

Branch,A.D., Robertson,H.D., Greer,C., Gegenheimer,P., Peebles,C. and Abelson,J. (1982) Cell-free circularization of viroid progeny RNA by an RNA ligase from wheat germ. *Science*, **217**, 1147–1149.

Bult,C.J. *et al.* (1996) Complete genome sequence of the metanogenic archaeon, *Methanococcus jannaschii*. *Science*, **273**, 1058–1073.

Chen,C. and Okayama,H. (1987) High efficiency transformation of mammalian cells by plasmid DNA. *Mol. Cell. Biol.*, **7**, 2745–2752.

Christiansen,K., Knudsen,B.R. and Westergaard,O. (1994) The covalent eukaryotic topoisomerase I-DNA intermediate catalyzes pH-dependent hydrolysis and alcoholysis. *J. Biol. Chem.*, **269**, 11367–11373.

Cravchik,A. and Matus,A. (1993) A novel strategy for the immunological tagging of cDNA constructs. *Gene*, **137**, 139–145.

Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P. (1994) Purification of CpG islands using a methylated DNA binding column. *Nature Genet.*, **6**, 236–244.

Efimov,V.A., Buryakova,A.A., Reverdatto,S.V., Chakhmakhcheva,O.G. and Ovchinnikov,Y.A. (1983) Rapid synthesis of long-chain deoxyribonucleotides by the *N*-methylimidazole phosphotriester method. *Nucleic Acids Res.*, **11**, 8369.

Evan,G.I., Lewis,G.K., Ramsay,G. and Bishop,J.M. (1985) Isolation of monoclonal antibodies specific for human *c-myc* proto-oncogene product. *Mol. Cell. Biol.*, **5**, 3610–3616.

Feinberg,A.P. and Vogelstein,B. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.*, **132**, 6–13.

Filipowicz,W. and Gross,H.J. (1984) RNA ligation in eukaryotes. *Trends Biochem. Sci.*, **9**, 68–71.

Filipowicz,W. and Shatkin,A.J. (1983) Origin of splice junction phosphate in tRNAs processed by HeLa cell extract. *Cell*, **32**, 547–557.

Filipowicz,W. and Vicente,O. (1990) RNA 3'-terminal phosphate cyclase from HeLa cells. *Methods Enzymol.*, **181**, 499–510.

Filipowicz,W., Konarska,M., Gross,H.J. and Shatkin,A.J. (1983) RNA 3'-terminal phosphate cyclase activity and RNA ligation in HeLa cell extract. *Nucleic Acids Res.*, **11**, 1405–1418.

Filipowicz,W., Strugala,K., Konarska,M. and Shatkin,A.J. (1985) Cyclization of RNA 3'-terminal phosphate by cyclase from HeLa cells proceeds via formation of N(3')pp(5')A activated intermediate. *Proc. Natl Acad. Sci. USA*, **82**, 1316–1320.

Fleischmann,R.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.

Fraser,C.M. *et al.* (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397–403.

Furneaux,H., Pick,L. and Hurwitz,J. (1983) Isolation and characterization of RNA ligase from wheat germ. *Proc. Natl Acad. Sci. USA*, **80**, 3933–3937.

Gandini-Attardi,D., Margarit,I. and Tocchini-Valentini,G.P. (1985) Structural alterations in mutant precursors of the yeast tRNA₃^{Leu} gene which behave as defective substrates for a highly purified splicing endoribonuclease. *EMBO J.*, **4**, 3289–3297.

Gegenheimer,P., Gabius,H.J., Peebles,C.L. and Abelson,J. (1983) An RNA ligase from wheat germ which participates in transfer RNA splicing *in vitro*. *J. Biol. Chem.*, **258**, 8365–8373.

Greer,C.L., Peebles,C.L., Gegenheimer,P. and Abelson,J. (1983a) Mechanism of action of a yeast RNA ligase in tRNA splicing. *Cell*, **32**, 537–546.

Greer,C.L., Javor,B. and Abelson,J. (1983b) RNA ligase in Bacteria: formation of a 2',5' linkage by an *E. coli* extract. *Cell*, **33**, 899–906.

Hinton,D.M., Brennan,C.A. and Gumport,R.I. (1982) The preparative synthesis of oligodeoxyribonucleotides using RNA ligase. *Nucleic Acids Res.*, **10**, 1877–1894.

Kiberstis,P.A., Haseloff,J. and Zimmern,D. (1985) 2'-phosphomonoester, 3'-5' phosphodiester bond at a unique site in a circular viral RNA. *EMBO J.*, **4**, 817–827.

Kikuchi,Y., Tyc,K., Filipowicz,W., Sanger,H.L. and Gross,H.J. (1982) Circularization of linear viroid RNA via 2'-phosphomonoester, 3',5'-phosphodiester bonds by a novel type of RNA ligase from wheat germ and *Chlamydomonas*. *Nucleic Acids Res.*, **10**, 7521–7529.

Kimball,A.S., Lee,J., Jayaram,M. and Tullius,T.D. (1993) Sequence-specific cleavage of DNA via nucleophilic attack of hydrogen peroxide, assisted by Flp recombinase. *Biochemistry*, **32**, 4698–4701.

Kjems,J. and Garrett,R.A. (1988) Novel splicing mechanism for the ribosomal RNA intron in the Archaeobacterium *Desulfurococcus mobilis*. *Cell*, **54**, 693–703.

Konarska,M., Filipowicz,W., Domdey,H. and Gross,H.J. (1981) Formation of a 2'-phosphomonoester, 3',5'-phosphodiester linkage by a novel RNA ligase in wheat germ. *Nature*, **293**, 112–116.

- Konarska, M., Filipowicz, W. and Gross, H.J. (1982) RNA ligation via 2'-phosphomonoester, 3',5'-phosphodiester linkage: requirement of 2',3'-cyclic phosphate termini and involvement of a 5'-hydroxyl polynucleotide kinase. *Proc. Natl Acad. Sci. USA*, **79**, 1474–1478.
- Kornberg, A. and Baker, T. (1991) *DNA Replication*, 2nd edn. W.H. Freeman and Company, New York.
- Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.*, **15**, 8125–8132.
- Laski, F.A., Fire, A.Z., RajBhandary, U.L. and Sharp, P.A. (1983) Characterization of tRNA precursor splicing in mammalian extracts. *J. Biol. Chem.*, **258**, 11974–11980.
- Laskowski, M., Sr (1971) Venom exonuclease. In Boyer, P.D. (ed.), *The Enzymes*, vol. IV, 3rd edn. Academic Press, pp. 313–328.
- Latham, K.A. and Lloyd, R.S. (1995) δ -Elimination by T4 endonuclease V at a thymine dimer site requires a secondary binding event and amino acid Glu-23. *Biochemistry*, **34**, 8796–8803.
- Lund, E. and Dahlberg, J.E. (1992) Cyclic 2',3'-phosphates and nontemplated nucleotides at the 3' end of spliceosomal U6 small nuclear RNAs. *Science*, **255**, 327–330.
- Manley, J.L. and Tacke, R. (1996) SR proteins and splicing control. *Genes Dev.*, **10**, 1569–1579.
- Peebles, C.L., Gegenheimer, P. and Abelson, J. (1983). Precise excision of intervening sequences from precursor tRNAs by a membrane-associated yeast endonuclease. *Cell*, **32**, 525–536.
- Perkins, K.K., Furneaux, H. and Hurwitz, J. (1985) Isolation and characterization of an RNA ligase from HeLa cells. *Proc. Natl Acad. Sci. USA*, **82**, 684–688.
- Phizicky, E.M. and Greer, C. (1993) Pre-tRNA splicing: variation on a theme or exception to the rule? *Trends Biochem. Sci.*, **18**, 31–34.
- Phizicky, E.M., Schwartz, R.C. and Abelson, J. (1986) *Saccharomyces cerevisiae* tRNA ligase. *J. Biol. Chem.*, **261**, 2978–2986.
- Phizicky, E.M., Consaul, S.A., Nehrke, K.W. and Abelson, J. (1992) Yeast tRNA ligase mutants are nonviable and accumulate tRNA splicing intermediates. *J. Biol. Chem.*, **267**, 4577–4582.
- Pick, L., Furneaux, H. and Hurwitz, J. (1986) Purification of wheat germ RNA ligase. II. Mechanism of action of wheat germ RNA ligase. *J. Biol. Chem.*, **261**, 6694–6704.
- Rand, K.N. and Gait, M.J. (1984) Sequence and cloning of bacteriophage T4 gene 63 encoding RNA ligase and tail fibre attachment activities. *EMBO J.*, **3**, 397–402.
- Rauhut, R., Green, P.R. and Abelson, J. (1990) Yeast tRNA-splicing endonuclease is a heterotrimeric enzyme. *J. Biol. Chem.*, **265**, 18180–18184.
- Reinberg, D., Arenas, J. and Hurwitz, J. (1985) The enzymatic conversion of 3'-phosphate terminated RNA chains to 2',3'-cyclic phosphate derivatives. *J. Biol. Chem.*, **260**, 6088–6097.
- Richards, G.M. and Laskowski, M., Sr (1969) Negative charge at the 3' terminus of oligonucleotides and resistance to venom exonuclease. *Biochemistry*, **8**, 1786–1795.
- Sadowski, P.D. (1995) The F1p recombinase of the 2-microns plasmid of *Saccharomyces cerevisiae*. *Prog. Nucleic Acid Res. Mol. Biol.*, **51**, 53–91.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Schwartz, R.C., Greer, C.L., Gegenheimer, P. and Abelson, J. (1983) Enzymatic mechanism of an RNA ligase from wheat germ. *J. Biol. Chem.*, **258**, 8374–8383.
- Shuman, S. and Schwer, B. (1995) RNA capping enzyme and DNA ligase: a superfamily of covalent nucleotidyl transferases. *Mol. Microbiol.*, **17**, 405–410.
- Stange, N. and Beier, H. (1987) A cell-free plant extract for accurate pre-tRNA processing, splicing and modification. *EMBO J.*, **6**, 2811–2818.
- Tazi, J., Forne, T., Jeanteur, P., Cathala, G. and Brunel, C. (1993) Mammalian U6 small nuclear RNA undergoes 3' end modifications within the spliceosome. *Mol. Cell. Biol.*, **13**, 1641–1650.
- Thompson, L.D. and Daniels, C.J. (1988) A tRNA^{Tp} intron endonuclease from *Halobacterium volcanii*. *J. Biol. Chem.*, **263**, 17951–17959.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
- Vicente, O. and Filipowicz, W. (1988) Purification of RNA 3'-terminal phosphate cyclase from HeLa cells. *Eur. J. Biochem.*, **176**, 431–439.
- Westaway, S.K. and Abelson, J. (1995) Splicing of tRNA precursors. In Soell, D. and Uxxx, R. (eds), *tRNA: Structure, Biosynthesis and Function*. American Society for Microbiology, Washington, DC, pp. 79–92.
- Zillmann, M., Gorovsky, M.A. and Phizicky, E.M. (1991) Conserved mechanism of tRNA splicing in eukaryotes. *Mol. Cell. Biol.*, **11**, 5410–5416.

Received on October 31, 1996; revised on January 29, 1997