The EMBO Journal Vol.16 No.20 pp.6290–6300, 1997

RNA–protein interactions of an archaeal homotetrameric splicing endoribonuclease with an exceptional evolutionary history

Jens Lykke-Andersen and Roger A.Garrett[1]

RNA Regulation Centre, Institute of Molecular Biology, Copenhagen University, Sølvgade 83H, DK-1307 Copenhagen K, Denmark

[1]Corresponding author
email: garrett@mermaid.molbio.ku.dk

The splicing endoribonuclease from *Methanococcus jannaschii*, a member of a recently defined family of enzymes involved in splicing of archaeal introns and eukaryotic nuclear tRNA introns, was isolated and shown by cross-linking studies to form a homotetramer in solution. A non-cleavable substrate analogue was synthesized by incorporating 2′-deoxyuridines at the two cleavage sites and complexed to the splicing enzyme. The complex was subjected to protein footprinting and the results implicated an RNP1-like sequence and a sequence region immediately N-terminal to a putative leucine zipper in substrate binding. In addition, a histidine residue (His125), positioned within a third RNA binding region, was shown to be involved in catalysis by mutagenesis. The splicing enzyme was localized on the central helix and the two 3 nt bulges of the conserved archaeal 'bulge–helix–bulge' substrate motif by RNA footprinting. Sequence comparison with the dimeric splicing enzyme from *Halobacterium volcanii* demonstrates that the latter is a tandemly repeated duplication of the former, where alternating segments within each protein half degenerated after the duplication event. Another duplication event, in the eukaryotic domain, produced two different homologues of the *M.jannaschii*-type enzyme structure. The data provide strong evidence that the tetrameric *M.jannaschii* enzyme consists of two isologously associated dimers, each similar to one *H.volcanii* monomer and each consisting of two monomers, where one face of monomer 1 and the opposite face of monomer 2 are involved in RNA binding.

Keywords: archaeal introns/*Methanococcus jannaschii*/protein footprinting/RNA–protein interactions/splicing endoribonuclease

Introduction

RNA maturation involves RNA splicing in all three domains of life. Three splicing mechanisms, which differ both mechanistically and, probably, in their evolutionary origins, are involved in removal of (i) group I introns, (ii) group II, III and spliceosomal introns and (iii) archaeal introns and nuclear tRNA introns (Cech, 1990; Moore *et al.*, 1993; Phizicky and Greer, 1993; Michel and Ferat, 1995; Lykke-Andersen *et al.*, 1997). The first two mechanisms both involve two consecutive transesterifications and are essentially RNA catalysed, although factor dependence has evolved which for the group I and II introns usually consists of a single protein and for the spliceosomal introns includes five RNA species and >100 proteins. The third mechanism of pre-RNA splicing, for archaeal tRNA and rRNA introns and eukaryotic nuclear tRNA introns, involves cleavage by a splicing endoribonuclease at each exon–intron junction, which generates a 2′,3′-cyclic phosphate and a 5′-OH (Peebles *et al.*, 1983; Kjems and Garrett, 1988; Thompson and Daniels, 1988). The exons are subsequently ligated by one of two alternative mechanisms which involve either direct attack of the 5′-OH on the 2′,3′-cyclic phosphate or a complex series of reactions requiring ATP and GTP (Belford *et al.*, 1993; Phizicky and Greer, 1993; Westaway *et al.*, 1993). In archaea the large rRNA introns contain open reading frames (ORFs), some of which encode homing endonucleases, and they circularize, after excision, to generate highly structured and stable RNA species (Kjems and Garrett, 1988; Dalgaard and Garrett, 1992; Dalgaard *et al.*, 1993; Lykke-Andersen and Garrett, 1994; Lykke-Andersen *et al.*, 1994).

The genes encoding the splicing endoribonucleases of yeast and the euryarchaeote *Halobacterium volcanii* have recently been sequenced and the results reveal that two subunits of the heterotetrameric yeast enzyme are homologues of the homodimer from *H.volcanii* (Kleman-Leyer *et al.*, 1997; Trotta *et al.*, 1997). Despite their possible common evolutionary origin, important differences exist between the nuclear tRNA introns and archaeal introns. For example, in contrast to eukaryotic tRNA introns, which are always positioned 1 nt 3′ of the anticodon, archaeal tRNA introns also exist at other tRNA positions (Wich *et al.*, 1987; Kjems *et al.*, 1989; Lykke-Andersen *et al.*, 1997). Moreover, archaeal introns also occur in 16S and 23S rRNAs of crenarchaeotes (Kjems and Garrett, 1985; Dalgaard and Garrett, 1992; Burggraf *et al.*, 1993). Another difference is that a 'bulge–helix–bulge' motif forms at the archaeal exon–intron junctions, which is necessary and sufficient for cleavage (Thompson and Daniels, 1988, 1990; Kjems *et al.*, 1989; Kjems and Garrett, 1991), whereas for the eukaryotic pre-tRNA the splicing enzyme complex primarily recognizes the tRNA structure (Phizicky and Greer, 1993; Trotta *et al.*, 1997).

In the present work the protein sequences of the yeast and *H.volcanii* splicing enzymes were exploited to isolate the enzyme of *Methanococcus jannaschii* using the recently sequenced genome of the archaeon (Bult *et al.*, 1996). An artificial RNA substrate lacking 2′-OH groups at the cleavage sites was used to generate a stable, uncleaved RNA–protein complex, which was subjected to both RNA and protein footprinting studies. Furthermore, to investigate the catalytic reaction possible hydrogen donor and acceptor residues were subjected to mutagenesis

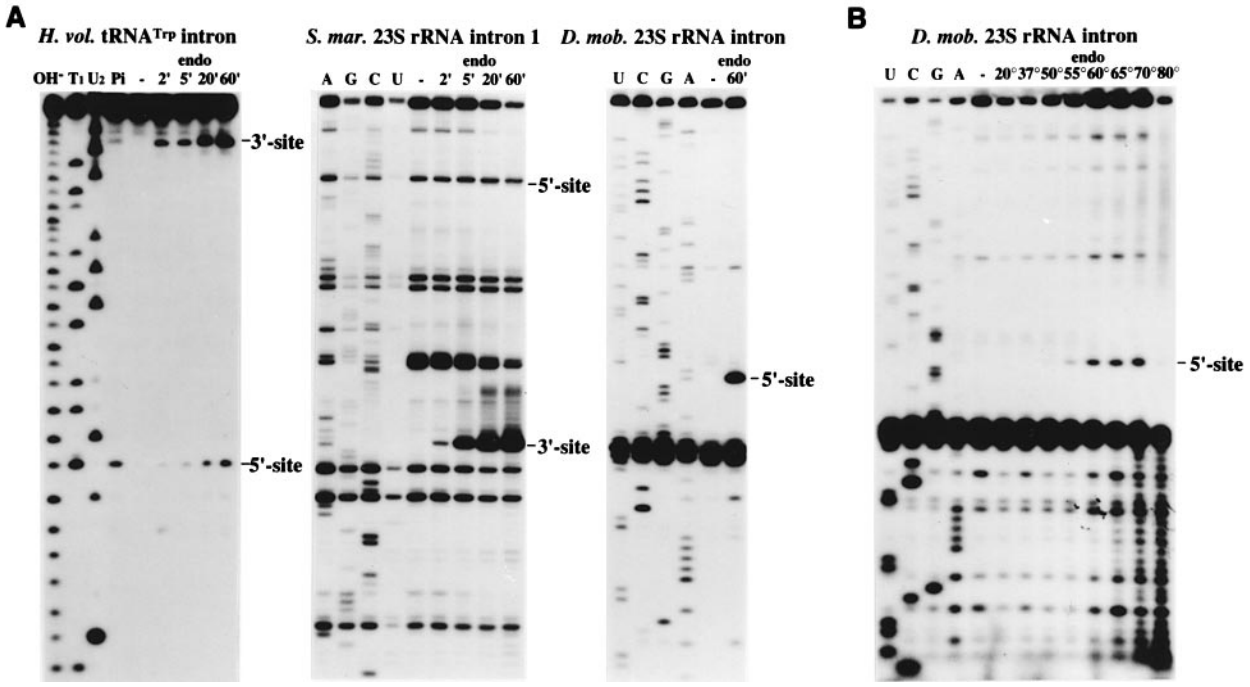

Fig. 1. Cleavage activity and temperature optimum for the *M.jannaschii* splicing enzyme. (**A**) The enzyme was incubated with three different intron-containing RNA substrates. The substrate in the left panel is a 36 nt 5′-^{32}P-end-labelled transcript corresponding to the exon–intron junctions of *H.volcanii* pre-tRNA^{Trp} (J.Diener and P.Moore, unpublished results). The RNA (~1 pmol) was incubated at 65°C for 60 min, alone (–) or in the presence of a *Pyrobaculum islandicum* cell extract (Pi), or for 2, 5, 20 or 60 min in the presence of the splicing enzyme (endo, ~0.05 pmol) as indicated. Basic hydrolysis (OH^-) and RNase T_1 (guanosine-specific) and U_2 (adenosine-specific) reactions with the substrate are run alongside as sequence markers. The substrates in the centre and right panels are T7 RNA polymerase transcripts containing intron 1 from *S.mariunus* pre-23S rRNA and the intron from *D.mobilis* pre-23S rRNA respectively, with flanking exons. The substrates (~2 pmol) were incubated alone (–) or in the presence of splicing enzyme (endo, ~0.05 pmol) for 2, 5, 20 or 60 min as indicated. Cleavage sites were determined by primer extension from oligonucleotide primers downstream of the 3′- and 5′-cleavage sites of the *S.marinus* and *D.mobilis* pre-23S rRNA transcripts respectively. Dideoxynucleotide sequencing reactions were co-electrophoresed (A, G, C and U). The cleavage sites, which correspond to the exon–intron junctions, are indicated on the right of each autoradiogram. The 5′-cleavage site of the *S.marinus* pre-23S rRNA, which is immediately below a control band, is barely visible, probably due to concurrent cleavage of 3′- and 5′-sites on each molecule. (**B**) Temperature optimum of the *M.jannaschii* splicing enzyme. The *D.mobilis* pre-23S rRNA transcript (~2 pmol) was incubated for 15 min, alone (–) or in the presence of ~0.05 pmol enzyme (endo), at different temperatures ranging from 20 to 80°C as indicated. Endonucleolytic cleavage was visualized by primer extension and dideoxynucleotide sequencing reactions were included (U, C, G and A).

and tested for cleavage activity and substrate binding. These data provide important insights into the endoribonuclease–RNA interaction and subunit structure of the recently identified family of splicing enzymes and, together with amino acid sequence comparisons, they provide a detailed new insight into the evolution of the proteins.

Results

Selection, cloning and purification of the splicing endoribonuclease from M.jannaschii

The ORF of the splicing enzyme was located in the genome of the archaeon *M.jannaschii* (Bult *et al.*, 1996) by sequence comparison with the Sen2p subunit of the tRNA splicing endoribonuclease from *Saccharomyces cerevisiae* (Rauhut *et al.*, 1990; Trotta *et al.*, 1997) and the enzyme from *H.volcanii* (Kleman-Leyer *et al.*, 1997). A protein of 179 amino acids, and unknown function (MJ1412) showed 21% identity to the yeast enzyme and 30% identity to the C-terminal half of the 341 amino acid *H.volcanii* enzyme respectively. This ORF was inserted into the glutathione *S*-transferase (GST) fusion protein expression vectors pGEX-2TK and pGEX-GTH (Jensen *et al.*, 1995a), allowing one-step purification on a glutathione–Sepharose matrix. The GST tag was removed by cleavage with the endoproteinase thrombin and the proteins were, in some experiments, site specifically ^{32}P-labelled at either the N-terminus (pGEX-2TK) or the C-terminus (pGEX-GTH) using heart muscle kinase and [γ-^{32}P]ATP. Both derivatives were used in the protein footprinting experiments in order to distinguish primary from secondary proteinase cuts, but only the pGEX-GTH derivative was used in the following experiments designed to characterize the enzyme.

These included testing the purified protein for its capacity to cleave different intron-containing substrates (Figure 1A). All substrates were cut specifically at both intron–exon junctions, confirming that the *M.jannaschii* protein is a splicing enzyme. The temperature optimum of the enzyme was determined by treating the pre-23S rRNA of *Desulfurococcus mobilis* over the range 20–80°C; maximum cleavage occurred at 65–70°C (Figure 1B). At 80°C cleavage was obscured by extensive RNA hydrolysis, probably induced at 10 mM Mg^{2+}.

The splicing endoribonuclease from M.jannaschii is a homotetramer

Protein cross-linking experiments were performed to investigate the quarternary structure of the endoribonuclease. C-terminally ^{32}P-labelled enzyme was incubated at

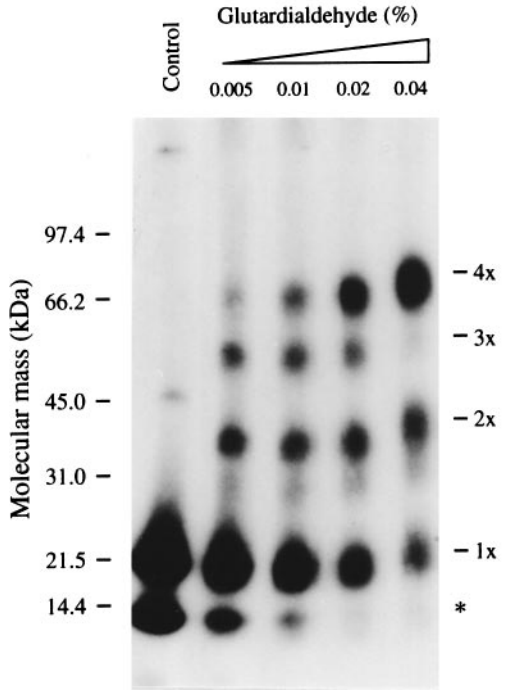

Fig. 2. Quarternary structure of the splicing enzyme determined by protein cross-linking. ^{32}P-labelled enzyme (~20 kDa) was incubated at 1 ng/µl for 2 h at 65°C alone (control) or in the presence of 0.005, 0.01, 0.02 or 0.04% glutardialdehyde as indicated. Products were run on an 8% polyacrylamide–SDS–tricin gel which was subsequently subjected to autoradiography. Migration and molecular mass of unlabelled marker proteins (in kDa) are indicated on the left and the migration of cross-linked protein species containing one, two, three and four subunits are indicated on the right. *, a degradation product always present with the full-length purified splicing enzyme (see also Figure 4A).

a low concentration (1 ng/µl) at 65°C before adding glutardialdehyde at different concentrations. After 2 h incubation the products were separated on SDS–tricin–polyacrylamide gels and subjected to autoradiography (Figure 2). Four bands appeared in the cross-linking reactions, at ~20, 40, 60 and 80 kDa, which correspond in size to monomeric, dimeric, trimeric and tetrameric enzyme respectively, with no bands larger than the 80 kDa band at the highest glutardialdehyde concentration. This strongly suggests that the endoribonuclease is a tetramer in solution. This inference received support from the more rapid disappearance of the trimeric than the dimeric product on increasing the glutardialdehyde concentration, indicating that a 2:2 symmetry exists in the tetramer (Figure 2). Addition to the reaction mixture of bovine serum albumin at 100 ng/µl or DmU2 RNA substrate (see below) at 1 µM produced unaltered cross-linking patterns (data not shown).

Synthesis of a non-cleavable substrate analogue

A stable enzyme–substrate complex was produced by constructing 36 and 54 nt substrates, termed DmU1 and DmU2 respectively. They were generated on the basis of the exon–intron sequence of the pre-23S rRNA of *D.mobilis* (Figure 3A). Both substrates contain only two uridines, which are positioned on the 5′-side of each of the two cleavage sites, corresponding to the central positions of the 3 nt bulges (Figure 3A). 2′-Deoxyribonucleotides were

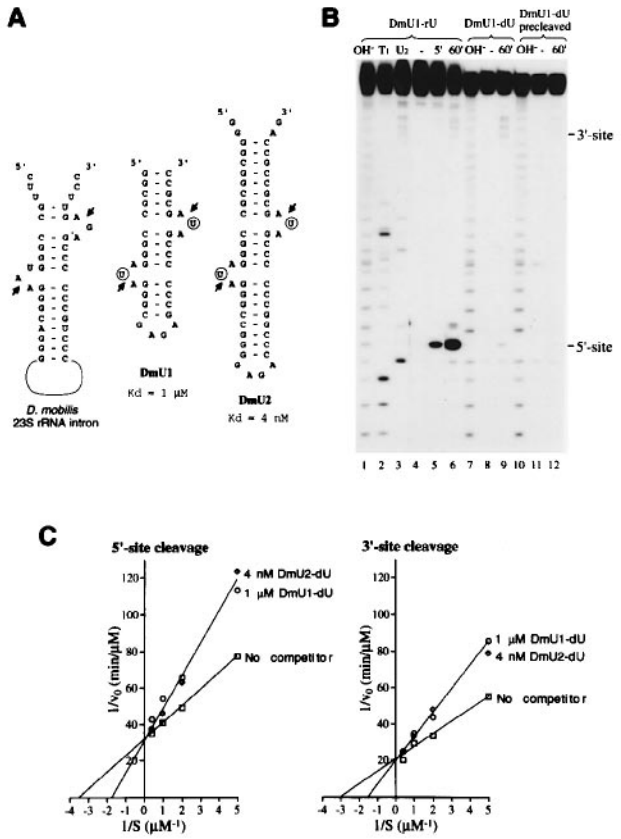

Fig. 3. Formation of a stable endoribonuclease–substrate complex. (**A**) The cleavage motif of the *D.mobilis* pre-23S rRNA (left) was used as the basis for constructing the substrates DmU1 (middle) and DmU2 (right). DmU1 and DmU2 contain only two uridines (circled Us), one in the middle position of each 3 nt bulge. This enables the insertion of 2′-deoxynucleotides at each cleavage site, indicated by arrows. The dissociation constants (K_d) for binding of the splicing enzyme to DmU1 and DmU2, as determined in (C), are given below the sequences. (**B**) Cleavage assays on 5′-^{32}P-end-labelled DmU1 (~1 pmol) containing uridines (DmU1-rU) or deoxyuridines (DmU1-dU) at the cleavage junctions or deoxyuridine-containing DmU1 gel purified after preincubation with splicing enzyme (DmU1-dU precleaved) were incubated at 65°C alone (–) or in the presence of the enzyme (~0.05 pmol) for 5 or 60 min as indicated. Products were separated on 20% polyacrylamide–7 M urea gels and autoradiographed. Basic hydrolysis reactions (OH^-) were performed on all three substrates and RNase T_1 and U_2 reactions were performed on DmU1-rU only. The 5′- and 3′-cleavage sites are indicated on the right; notice the gaps in the basic hydrolysis reactions on the DmU1-dU substrates at these positions. The 3′-cleavage is barely visible, probably due to concurrent cleavage at the 5′- and 3′-sites of the molecules. Cleavage at the 3′-site was confirmed on a 3′-^{32}P-end-labelled substrate (data not shown). (**C**) Cleavage competition experiments. The initial cleavage rate of the *S.marinus* 23S rRNA intron with flanking exons was measured at four different RNA concentrations (0.2, 0.5, 1.0 and 3.0 µM) in the absence or presence of competitor RNA and the results for the 5′- and 3′-cleavages are given in a double reciprocal plot. In the experiments shown the competitors were DmU1-dU (open circle) and DmU2-dU (open diamond) at 1 µM and 4 nM respectively, as indicated. The cleavage rates were measured using internally ^{32}P-labelled RNA substrate. The RNA substrate, with or without competitor, was incubated under mineral oil in 60 µl cleavage buffer (see Materials and methods) at 65°C for 5 min and the enzyme was added to ~4 nM. Samples (10 µl) were stopped after 2, 4, 8, 14 and 20 min incubation and products were separated on 8% polyacrylamide–7 M urea gels. Band intensities were measured in an Instant Imager (Packard, Meriden) and the extents of cleavage at the two cleavage sites were calculated by comparison with uncleaved and fully cleaved substrate.

specifically inserted at these positions by replacing rUTP with dUTP in *in vitro* transcription reactions (Conrad *et al.*, 1995). The DmU1 substrate was prepared with both uridines (DmU1-rU) and deoxyuridines (DmU1-dU) at the cleavage site positions and 5′-^{32}P-end-labelled before testing for cleavage (Figure 3B). While the uridine derivative of DmU1 was strongly cleaved, the deoxyuridine derivative was cleaved at a very low level (Figure 3B, lanes 1–9). The residual cleavage of the latter is due to low levels of contaminating ribonucleotides at the deoxyuridine positions, since purified, pre-cleaved DmU1-dU was not cut by the splicing endoribonuclease (Figure 3B, lanes 10–12). Cleavage experiments with DmU2 gave similar results (not shown).

Next, the deoxyuridine-containing DmU1 and DmU2 substrates were tested for specific binding to the splicing enzyme in cleavage competition experiments at 65°C (Figure 3C). Intron 1 from the pre-23S rRNA of *Staphylothermus marinus* was used as substrate and the uncleavable DmU1-dU and DmU2-dU RNAs were added as competitors (Figure 3C). Initial rate kinetics were measured and approximate K_m values were calculated for the *S.marinus* intron. Cleavage at the 5′- and 3′-bulges gave K_m values of 0.29 ± 0.05 and 0.34 ± 0.05 μM respectively. While no competition was observed with non-specific RNA (total RNA from *Escherichia coli*) at a concentration corresponding to DmU1 at 1 μM (12.5 ng/μl) (data not shown), the splicing endoribonuclease specifically bound to the uncleavable DmU1-dU and DmU2-dU RNAs with dissociation constants (K_d) of 1.0 ± 0.2 μM and 4 ± 1 nM respectively (Figure 3C). The relatively high K_d value for the DmU1-dU RNA probably reflects its small size and relative instability at 65°C. This interpretation was reinforced by competition experiments at 50°C that yielded a K_d of ~0.1 μM for DmU1-dU (data not shown).

Mapping the RNA substrate on the splicing endoribonuclease

Formation of the stable enzyme–substrate complex enabled us to map RNA binding regions on the enzyme using a protein footprinting approach (Jensen *et al.*, 1995b; Lykke-Andersen *et al.*, 1996; Tange *et al.*, 1996). The endoribonuclease was ^{32}P-labelled at either the N- or C-terminus and complexed with DmU2-dU RNA at 100 nM and 65°C. This concentration was found to be optimal from titration experiments (see Figure 4D). Samples were cooled to 50°C and subjected to limited proteolysis with eight different proteinases (Figure 4A). Endoribonuclease incubated with total *E.coli* tRNA served as a control for each reaction. Several proteinase cleavage sites were detected in the endoribonuclease and all of the major sites were present in both the N- and C-terminally labelled protein, although the sites near the unlabelled ends could not be separated from the full-length protein. This shows that the cleavages are primary events, as opposed to secondary cuts that occur as a result of cleavage elsewhere in the protein. The positions of the cuts were determined by analysing the cleavage patterns generated by the amino acid-specific proteinases and plotting the logarithm of the molecular weights of the assigned fragments against the migration distances in the gels (not shown). For the C-terminally labelled endoribonuclease a fragment of ~125 amino acids was invariably detected in relatively high yield (~10–15%). This fragment probably derives from a 'nicked' endoribonuclease molecule where the N- and C-terminal regions are maintained together in the structure, since the C-terminal fragment co-purified with the full-length endoribonuclease (the GST tag is at the N-terminus) on the glutathione–Sepharose matrix. Furthermore, after removal of the GST tag the fragment did not separate from the full-length protein during gel filtration (data not shown). A subset of the proteinase cuts was specifically protected in the presence of the DmU2 substrate (summarized in Figure 4C). They were located at, or near, L30, Y35, Y108, F112, Y115, E116, Y128, F139, E143 and F147, which were all protected in both N- and C-terminally labelled endoribonuclease, and at Y176, which was protected for one end-label while cleavage could not be resolved from full-length protein with the other (Figure 4A–C). Protections at K131 and R149 and enhanced cleavages at E72/73, R79 and R155 were seen exclusively for one end-label and might result from secondary cuts.

In contrast to protein footprinting studies on other RNA binding proteins (Jensen *et al.*, 1995b; B.T.Porse and R.A.Garrett, unpublished results), none of the proteinase cleavage sites were protected by >50% on RNA substrate binding (Figure 4A and B). For example, band intensities are given for two strong chymotrypsin cleavages (Y100 and Y128) on titrating DmU2-dU RNA with the C-terminally labelled splicing endoribonuclease in Figure 4D. These band intensities were quantified and the curves show the extent of cleavage at each site compared with the uncomplexed protein. Residual cleavage at the protected site (Y128) was >50% at the highest substrate concentrations and half-protection was obtained at ~4 nM, in accordance with the previously calculated K_d. To eliminate the possibility that the residual 50% proteinase cleavage of the complexed endoribonuclease was due to the presence of a large fraction of inactive enzyme, the protein footprinting experiments were performed on three independent protein preparations, with similar results.

Mapping the endoribonuclease on the RNA substrate

The stable endoribonuclease–substrate complex was also subjected to RNA footprinting experiments using Fe^{2+}/EDTA/H_2O_2 (RNA backbone) and diethyl pyrocarbonate (DEP, N-7 positions of adenines) as probes (Figure 5). The DmU2-dU substrate, labelled at either the 3′- (Figure 5A) or 5′-end (data not shown), was complexed with the splicing enzyme at 65°C and the complex was probed with the chemicals (Figure 5A, lanes 4 and 7). RNA probed alone (lanes 3 and 6) and the RNA–protein complex incubated in the absence of chemicals (lanes 5 and 8) served as controls. A minor part of the DmU2-dU species (~0.5%) was cleaved by the enzyme (see Figure 3B), but this did not affect interpretation of the data (Figure 5A, lanes 4, 5, 7 and 8).

The RNA footprinting data obtained for the 3′- and 5′-end-labelled DmU2-dU RNA are summarized in Figure 5B. Similar protections were seen with both labellings, indicating that all the events are primary. A symmetrical protection pattern was observed on the RNA substrate; only the middle 4 bp helix and the two 3 nt bulges were protected against chemical probes on binding the tetrameric enzyme.

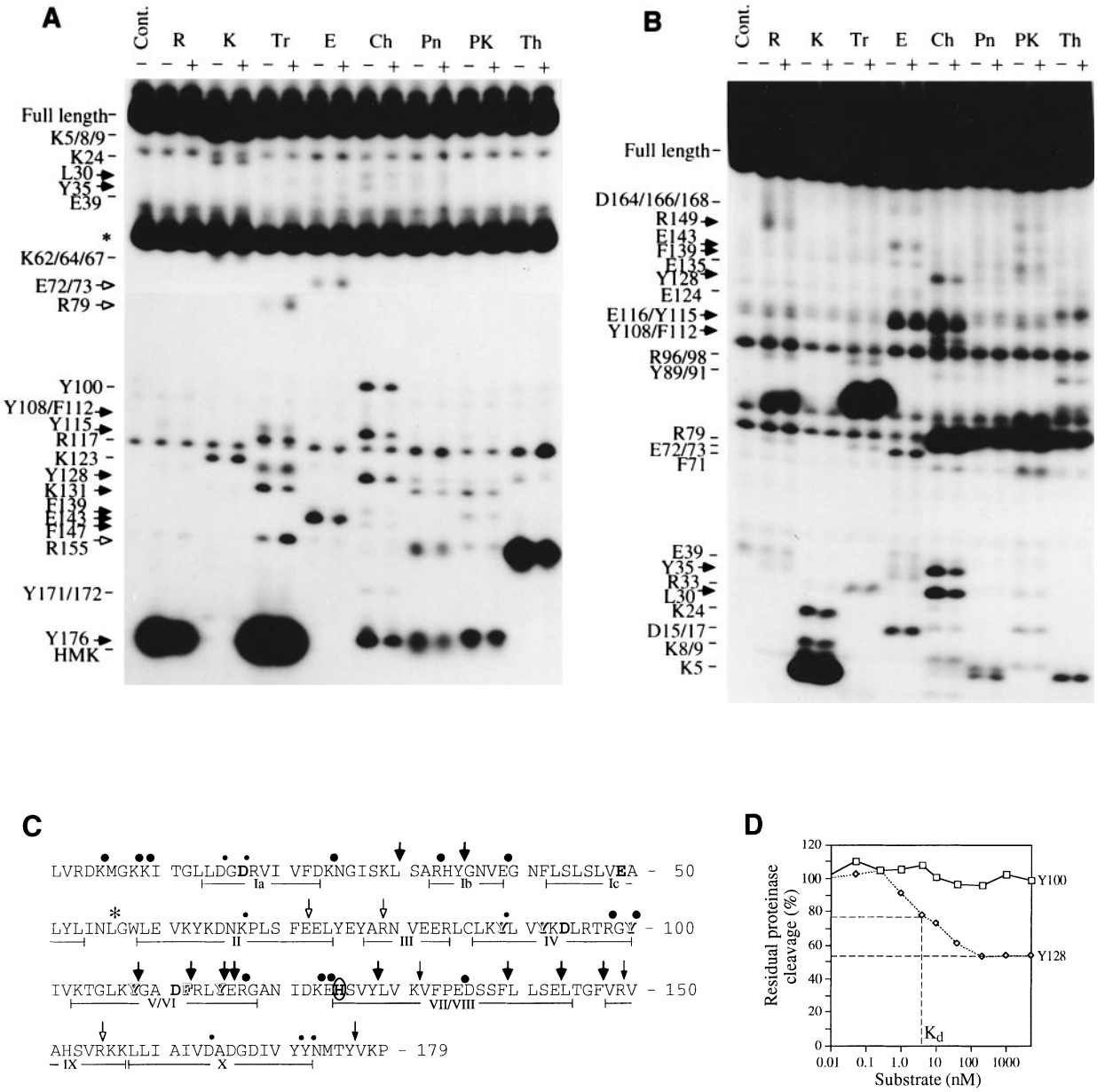

Fig. 4. Protein footprinting of the enzyme in the presence of DmU2-dU. (**A**) C-terminally and (**B**) N-terminally ^{32}P-labelled enzyme was incubated with DmU2-dU (+) or, as a negative control, an equal amount (w/v) of *E.coli* tRNA (–) and then probed with different proteinases. Products were separated on large 20% polyacrylamide–SDS–tricin gels and subjected to autoradiography. A control sample of endoribonuclease incubated in the absence of proteinase was included (Cont.). Assignments of the cleaved positions are given on the left. Those positions that were protected against cleavage on substrate binding are indicated by filled arrows and enhanced cleavages are denoted by open arrows. In (A) the top and bottom halves are short and long exposures of the same gel and a major control band is indicated with an asterisk (see text). HMK, a major short fragment generated by cleavage at an arginine in the heart muscle kinase site. Proteinases, with amino acid specificities given in parentheses, are: R, endoproteinase Arg-C (R); K, endoproteinase Lys-C (K); Tr, trypsin (R, K); E, endoproteinase Glu-C (E > D); Ch, chymotrypsin (Y, F, W > L, A, M); Pn, pronase (non-specific); PK, proteinase K (hydrophobic); Th, thermolysin (L, F, I, V, M, A). (**C**) Proteinase footprinting data superimposed on the amino acid sequence of the *M.jannaschii* enzyme. The data summarizes at least four independent sets of experiments, of which only one is shown in (A) and (B). Cleaved sites that were not affected by substrate binding are denoted by dots. Positions of reduced and enhanced proteinase cleavage on substrate binding are given by filled and open arrows respectively. Large symbols (dots or arrows) indicate effects seen with both C- and N-terminally labelled proteins. Residues mutated to alanines in this study are given in bold and His125, which participates in catalysis, is circled (see Figure 6). Aromatic residues conserved between the seven known splicing enzyme sequences are given as outline letters. Lines below the sequence, denoted by roman numerals, indicate segments defined in Figure 7A. The approximate site of strand breakage generating the major control band in (A) is indicated by an asterisk. The translation start codon could correspond to either amino acid L1, V2 or M6 shown here. (**D**) Proteinase protections do not exceed 50%. Proteinase footprinting experiments were performed on C-terminally ^{32}P-labelled endoribonuclease, using chymotrypsin in the presence of DmU2-dU substrate at concentrations of from 0.05 to 4000 nM. Band intensities corresponding to strong cleavages after an unaffected residue (Y100) and a residue protected on substrate binding (Y128) were measured in an Instant Imager (Packard, Meriden). Band intensites, normalized to intensities obtained on chymotrypsin cleavage of the protein in the absence of substrate, are shown as a function of substrate concentration (nM, logarithmic scale). The maximum protection on Y128 is ~45% and the K_d value is ~4 nM (indicated by dashed lines).

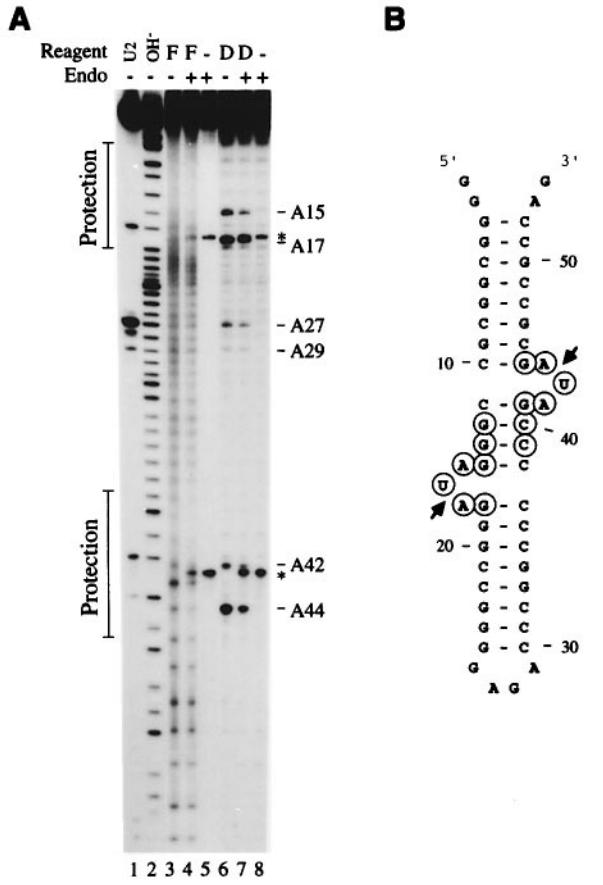

Fig. 5. Footprinting the splicing enzyme on DmU2-dU RNA. (**A**) 3′-^{32}P-end-labelled DmU2-dU substrate was incubated in the absence and presence of enzyme (Endo), as indicated, and subsequently probed with Fe^{2+}/EDTA (F) or DEP (D). Reaction products were separated on 20% polyacrylamide–7 M urea gels and autoradiographed. Basic hydrolysis (OH^-) and RNase U_2 reactions were run as sequence markers. Regions protected by the enzyme are indicated by vertical lines on the left. Adenine positions in the 3 nt bulges and in the terminal tetra-loop are indicated on the right. A minor part (<0.5%) of the substrates contained a ribonucleotide, instead of deoxyuridine, at one of the cleavage site junctions (see Figure 3B) and this leads to cleavage (indicated by asterisks) of the otherwise uncleavable substrate. The fragment generated by the DEP/aniline reaction at A17 migrates just below the 5′-cleavage fragment and protection at this site is difficult to see in the reproduction. (**B**) RNA footprinting data superimposed on the secondary structure of DmU2. Nucleotides protected against chemical reagents on endoribonuclease binding in both 3′- (A) and 5′-end-labelled DmU2-dU RNA (not shown) are circled.

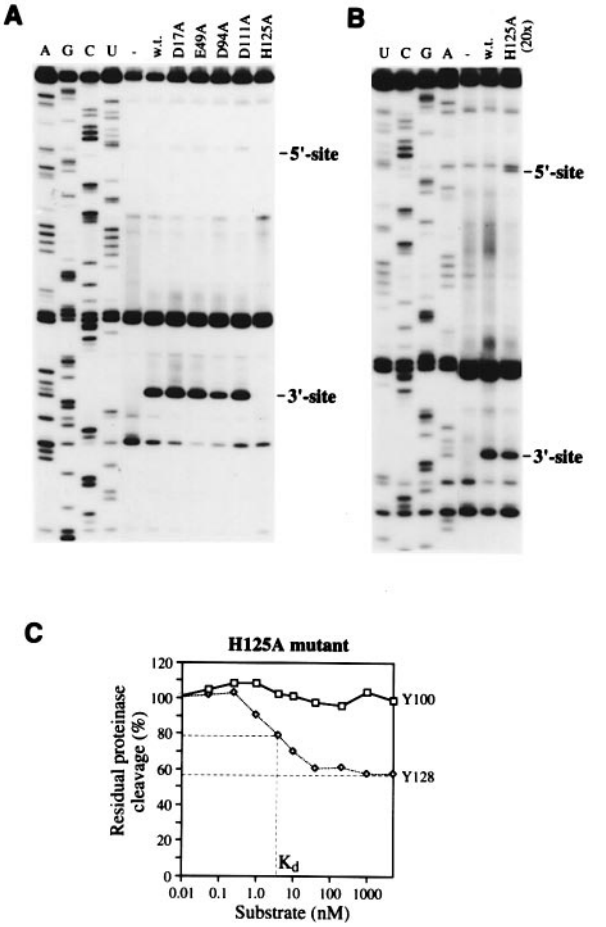

Fig. 6. Mutagenic approach to determining residues involved in the catalytic site. (**A**) T7 RNA polymerase transcript containing intron 1 from *S.marinus* pre-23S rRNA, with flanking exons, was incubated at 100 nM alone (–) or in the presence of ~2.5 nM splicing enzyme containing no mutation (w.t.) or single residue alanine substitutions as indicated (see also Figure 4C). RNA cleavage was visualized by primer extension and dideoxynucleotide sequencing reactions are included (A, G, C and U). Position of the 3′- and 5′-cleavage sites are indicated on the right. As in Figure 1A, 5′-cleavage is barely visible. (**B**) The His125→Ala mutant exhibits low catalytic activity. The *S.marinus* intron transcript (~100 nM) was incubated alone (–) or in the presence of ~2.5 nM unmutated (w.t.) or ~50 nM splicing enzyme mutated at His125 (His125→Ala). The cleavage sites are indicated to the right. Note that, as expected for a slow catalyst with two separate active sites, the 5′-site is visible for the catalytically impaired His125→Ala mutant. This reflects slow catalysis compared with the dissociation rate and results in non-concurrent cleavage at the 5′- and 3′-sites. (**C**) The His125→Ala mutant binds to the substrate with the same affinity as the unmutated enzyme. Proteinase footprinting experiments were performed on C-terminally ^{32}P-labelled His125→Ala mutant endoribonuclease, using chymotrypsin, in the presence of DmU2-dU substrate at concentrations of from 0.05 to 4000 nM, as described in Figure 4D. As for the unmutated enzyme, the His125→Ala mutant showed a maximum protection on Y128 of ~45% and a K_d value of ~4 nM (indicated by dashed lines).

A conserved histidine residue is involved in catalysis

The *M.jannaschii* splicing enzyme contains three histidines and 22 acidic residues (Figure 4C), any of which could act as proton acceptors or donors in a cleavage reaction. Aligning the sequence of the *M.jannaschii* enzyme with those of other organisms reveals that one histidine (His125, *M.jannaschii* numbering, Figure 4C) is conserved in all proteins, while the other two are exclusive to *M.jannaschii*. This histidine and four acidic residues (Asp17, Glu49, Asp94 and Asp111), which at an early stage appeared to be conserved but of which only Glu49 is now conserved in the seven sequences available, were changed into alanines. Mutant proteins were purified and tested for activity on the intron substrate of *S.marinus* (Figure 6A). While all four acidic residue mutants produced activities similar to that of the unmutated enzyme, the histidine mutant (His125→Ala) showed greatly reduced reactivity. This result was confirmed for two separate protein purifications. Raising the enzyme concentration 20 times showed that the His125→Ala mutant exhibited some activity (Figure 6B), but this was too low to permit reliable studies of initial rate kinetics. Therefore, the His125→Ala mutant was tested for substrate binding in protein foot-

printing experiments using DmU2-dU RNA. The same pattern of altered proteinase reactivities was obtained as for the unmutated protein, with a K_d of ~4 nM, establishing that only the catalytic activity was affected by the mutation (Figure 6C; compare with Figure 4D).

Discussion

The 37 kDa splicing endoribonuclease from H. volcanii arose by duplication of the 20 kDa M.jannaschii-type enzyme

Amino acid sequence comparisons showed that the small enzyme from *M.jannaschii* exhibited 30% identity with the C-terminal half of the *H.volcanii* enzyme (Kleman-Leyer *et al.*, 1997; Trotta *et al.*, 1997). However, a more careful examination of the sequences reveals that the *M.jannaschii* enzyme is also homologous to the N-terminal half of the *H.volcanii* enzyme, showing that the latter is a duplication of the former, with the N- and C-terminal halves of the 339 amino acid *H.volcanii* enzyme showing 26 and 30% identity over 159 and 180 amino acids respectively, to the full-length 179 amino acid *M.jannaschii* enzyme (Figure 7A). This correlates with the finding that the *H.volcanii* enzyme is a dimer (Kleman-Leyer *et al.*, 1997) while the *M.jannaschii* enzyme is a tetramer (this work). Therefore, where the former enzyme can be denoted as α_2, the latter will have the structure $(\alpha'_2)_2$, where α is similar in size and structure to α'_2 (see Figure 8A). However, the N- and C-terminal halves of the *H.volcanii* enzyme show only 17% identity to each other and a closer analysis of the sequence reveals that the relative locations of the conserved regions in the N- and C-terminal halves, when compared with the *M.jannaschii* enzyme, are different (Figure 7A). The conserved regions constitute 10 segments (segments I–X, Figure 7A) and, whereas they occur in the *M.jannaschii* enzyme in the order I to X (where segment I is split into Ia, Ib and Ic), in the duplicated protein of *H.volcanii* every second segment is missing from each protein half, so that all even numbered segments are present in the N-terminal half and all odd numbered segments occur in the C-terminal half (Figure 7A and B). Some overlaps are found in segments Ib, IV, VIII and X that further strengthen the duplication hypothesis. A 'mosaic' protein, generated by combining the conserved segments from the N- and C-terminal halves of the *H.volcanii* enzyme, shows 43% identity in 165 amino acids to the *M.jannaschii* enzyme, ~1.5 times higher than for each separate enzyme half and comparable with the identity found between other archaeal enzymes. This has implications both for the structure and evolution of the splicing enzyme. First, the original endoribonuclease must have been structured like the *M.jannaschii* enzyme, since the eukaryotic homologues have a similar primary structure. Thus, during evolution, somewhere after the split between the ancestors to *M.jannaschii* and *H.volcanii*, an in-frame gene duplication must have taken place in the *H.volcanii* branch and, since that time, alternate segments degenerated to produce the present splicing enzyme of *H.volcanii* (Figure 7B). Location of the duplication event was facilitated by the observation that the splicing enzyme from *Methanosarcina barkeri* has approximately the same size as the enzyme from *H.volcanii* (Garrett *et al.*, 1994) while the enzyme from *Methanobacterium thermoauto-*

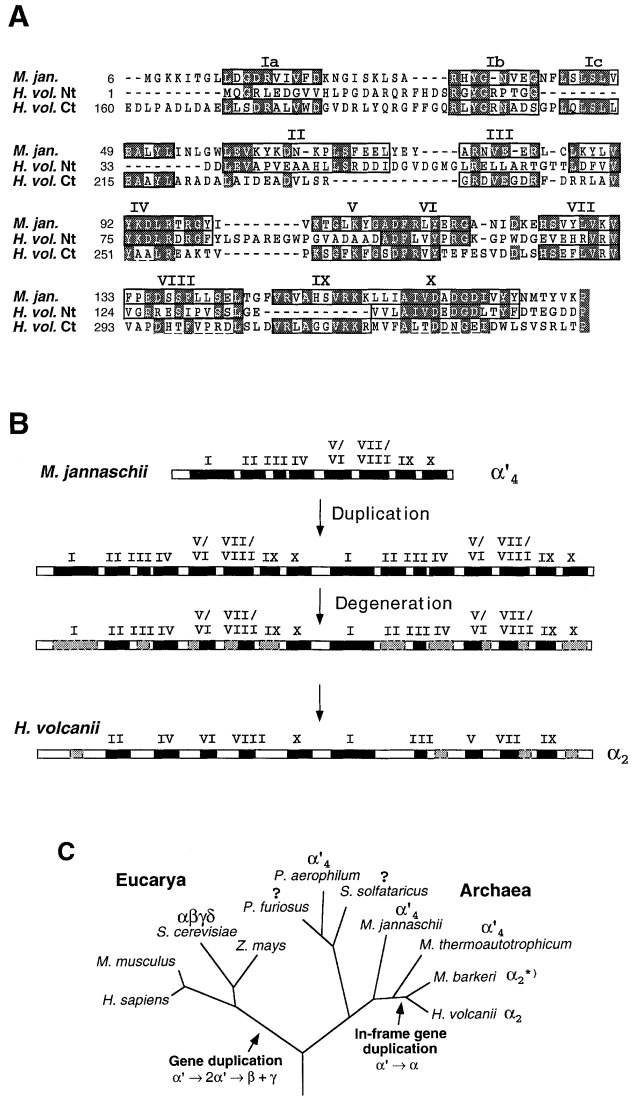

Fig. 7. Evolution of the splicing enzyme. (**A**) Amino acid alignment of the enzyme from *M.jannaschii* (top line) with the N- and C-terminal halves (middle and bottom lines) of the *H.volcanii* enzyme. Residues in the *M.jannaschii* enzyme found in either the N- or C-terminal half of the *H.volcanii* enzyme are shaded and similar sequence segments are boxed. These segments are numbered from I to X (see Discussion). (**B**) Possible evolutionary steps leading to generation of the duplicated and partially degenerate endoribonuclease found in *H.volcanii* from the earlier endoribonuclease represented by the *M.jannaschii* enzyme. Protein sequences are indicated as open boxes and the conserved segments (see above) are shown as black boxes. Partly conserved segments are shaded. (**C**) Evolutionary tree with organisms where one or more subunits of the splicing endoribonuclease has been sequenced or will be sequenced in genome sequencing projects (?). In the archaeal domain α'_4 and α_2 indicate proteins with primary structures like the tetrameric *M.jannaschii* and duplicated dimeric *H.volcanii* enzymes respectively (see Discussion). In the eukaryotic domain $\alpha\beta\gamma\delta$ represents the yeast heterotetrameric enzyme, of which the β (Sen2p) and γ (Sen34p) subunits are homologues of the archaeal enzymes (α'). Eukaryotes where sequences of one or more homologues of the yeast subunits are available are shown. The possible point of in-frame gene duplication is indicated by an arrow within the euryarchaeotal branch and a putative gene duplication event that produced two proteins homologous to the archaeal enzyme is shown in the eukaryotic domain. The primary structure of the *M.barkeri* enzyme (*) was deduced from the molecular weight, which is ~40 kDa (Garrett *et al.*, 1994). The *Homo sapiens* and *Mus musculus* sequences are from expressed sequence tags. The *Zea mays* sequence is an ORF in the intron of HMG (X72692). The *P.aerophilum* sequence is from the published genome (Fitz-Gibbon *et al.*, 1997).

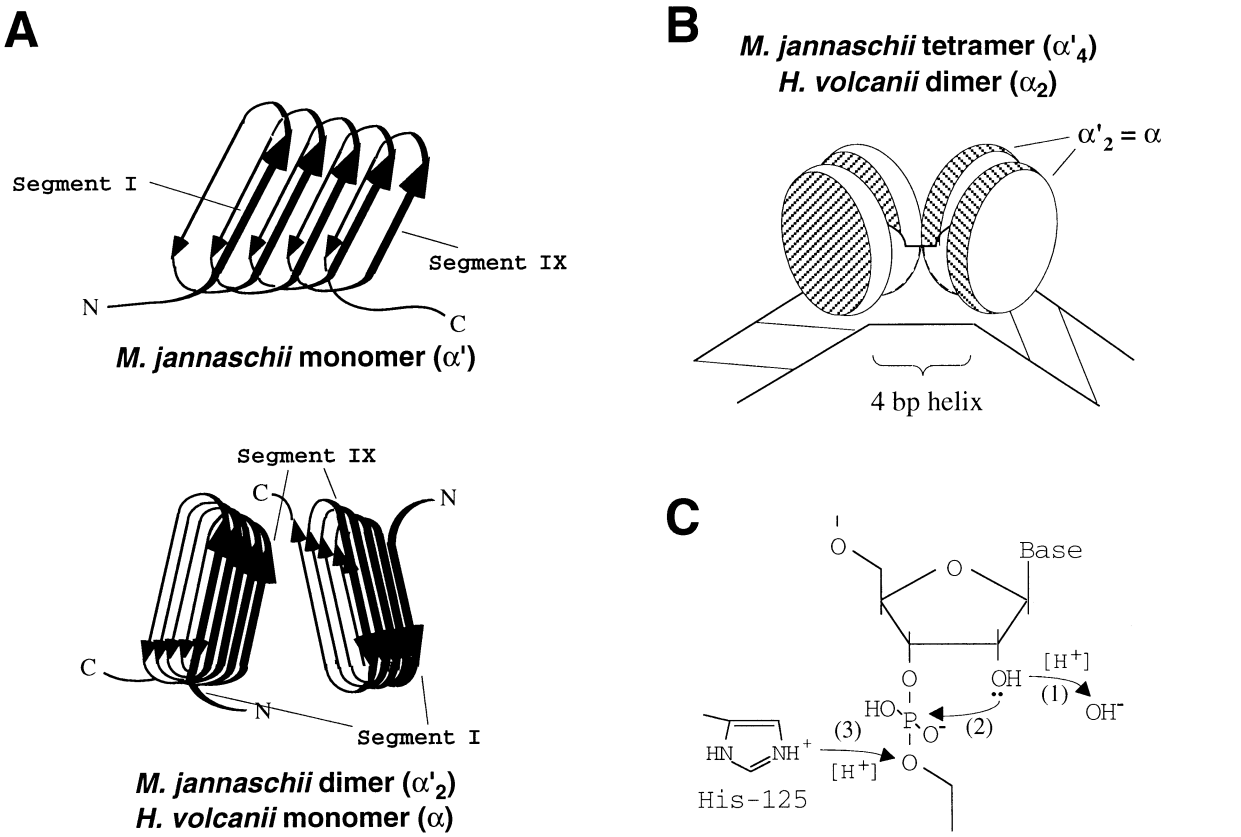

Fig. 8. Structural and catalytic model of the splicing endonuclease. (**A**) A structural model deduced from the sequence alignments and protein footprinting data. Each segment (I–X), shown in Figure 7A and B is represented by an arrow. Every second segment lies on the same face of the *M.jannaschii* monomer (upper panel). In the lower panel the heterologously associated *M.jannaschii* monomers, which constitute the dimer, which is similar to each *H.volcanii* monomer, are shown (see Discussion). The segment positions are arbitrary and the RNA substrate can be accommodated between the two subunits. (**B**) Model of the splicing enzyme on the RNA substrate deduced from protein and RNA fooprinting. The *M.jannaschii* tetramer (α'_4), which corresponds to the *H.volcanii* dimer (α_2), binds to the two bulges and the middle helix (4 bp helix) of the 'bulge–helix–bulge' motif (see Figure 5). The RNA substrate is drawn as a bent molecule, based on unpublished gel shift results (J.Z.Dalgaard and R.A.Garrett). The two faces of each monomer are shown in white and hatched. Notice that according to this model a proteinase site that is protected on substrate binding within the hatched region in monomer 1 will not be protected in monomer 2. (**C**) Model of the catalytic reaction leading to cleavage at each of the exon–intron junctions. (1) A hydrogen acceptor, which can be either a hydroxyl ion, as depicted, or an acidic or histidine residue of the endoribonuclease (see Discussion), extracts the proton from the 2′-OH group of the 3′-terminal exon or intron nucleotide. (2) The resulting 2′-oxoanion performs a nucleophilic attack on the 3′-phosphate group, releasing the 5′-oxoanion, which is subsequently (3) protonated by His125 of the enzyme. It cannot be ruled out from our data that His125 is involved in the first and not the third step of the reaction.

trophicum is similar to that of *M.jannaschii* (Kleman-Leyer *et al.*, 1997; Trotta *et al.*, 1997). A phylogenetic tree showing the possible point of in-frame gene duplication within the euryarchaeotal domain is given in Figure 7C.

Structural predictions

The following can be inferred about the overall structure of the splicing endoribonucleases. Since the *H.volcanii* enzyme is a dimer (α_2) and the *M.jannaschii* enzyme is a tetramer [$(\alpha'_2)_2$], each dimeric subunit of the *M.jannaschii* tetramer (α'_2) must be similar to each *H.volcanii* monomer (α). As a consequence of the 2-fold symmetry axis of the 'bulge–helix–bulge' substrate each cleavage site is likely to be cut by one *H.volcanii* monomer subunit (α), which corresponds to one *M.jannaschii* dimeric subunit (α'_2), and each subunit will show an isologous association corresponding to a 180° turn around an axis perpendicular to the 'bulge–helix–bulge' motif (α:α and α'_2:α'_2) (Figure 8A and B). Since every second segment has degenerated in the duplicated *H.volcanii* monomer, the segments that are important in monomer 1 of each *M.jannaschii* dimeric subunit (α'_2) are unimportant in monomer 2 and vice versa. It is likely, therefore, that every second segment lies on one face of each monomer such that one face of monomer 1 is involved in cleavage at one site, together with the opposite face of monomer 2 (Figure 8A and B). This inference is further supported by the protein footprinting results, where substrate binding produced only partial protection (<50%) against proteinase cleavage (Figure 4A, B and D). Assuming that proteinase sites on each monomer are cleaved equally, a theoretical maximum of 50% protection can be obtained for this model (Figure 8A and B). A similar result was obtained when the α subunit of the *E.coli* RNA polymerase was subjected to protein footprinting; 50% protection was seen on assembly of the polymerase, due to the presence of two asymmetrically associated α subunits (Heyduk *et al.*, 1996).

Possible subunit interactions

A leucine zipper-like region identified in the proteins (Kleman-Leyer *et al.*, 1997) corresponds to segment Ic

and such motifs are often involved in protein dimerization of DNA or RNA binding proteins. In the *M.jannaschii* splicing enzyme this region is characterized by high resistance to proteinase cleavage (see the gap between E39 and F71 in Figure 4B) and, although we cannot rule out that the region is buried in each monomer, the data are consistent with the region being involved in inter-subunit contacts. Another region that might be involved in subunit interactions is segment X, which constitutes a region rich in hydrophobic residues with some acidic residues in the centre. This region is conserved between all of the splicing enzymes, including the homologous yeast enzyme subunits Sen2p and Sen32p. Moreover, analysis of the two non-homologous yeast enzyme subunits Sen54p and Sen15p shows that both contain sequences with strong similarity to segment X at their C-termini (424-SFIIAIMDNGLISFV and 102-ILLALLNDDGTI-VYY respectively; compare with Figure 7A). Sen54p and Sen15p are known to interact with Sen2p and Sen32p respectively, and therefore, by analogy, the segment X region may contribute to the heterologous subunit–subunit contact proposed between the monomers of each dimer (α'_2) in the *M.jannaschii* enzyme.

Splicing endoribonuclease–RNA substrate interactions

A schematic model for association of the tetrameric *M.jannaschii* enzyme with the RNA substrate is given in Figure 8B. It is evident from the RNA footprinting experiments that the splicing enzyme only protects the RNA backbone at the two 3 nt bulges and the central 4 bp helix of the the 'bulge–helix–bulge' motif and not at the two adjoining double helical segments (Figure 5). This correlates well with the phylogenetic data on archaeal introns, which show that while the the 'bulge–helix–bulge' motif is always conserved, it is often flanked by only one stable double-helical segment in either the intron or exon part of the structure (Kjems and Garrett, 1991; Lykke-Andersen *et al.*, 1997). Therefore, the presence of the flanking helix, or helices, probably only serves to stabilize the three-dimensional structure of the the 'bulge–helix–bulge' motif. Gel retardation experiments indicated that the 'bulge–helix–bulge' RNA structure generates a bent molecule (J.Z.Dalgaard and R.A.Garrett, unpublished results) and one can envision that the two 3 nt bulges, which are located on one side of the RNA substrate, are accessible to the tetrameric enzyme, which can contact the bulges and the central helix (Figure 8B).

The proteinase cleavage sites protected by the RNA substrate fall in segments Ib, V and VII, which would be protected in monomer 1, and in VI and VIII in monomer 2, according to the structural model described above. Segments V and VI together contain a sequence similar to the RNP1 motif of the RNP consensus RNA binding sequence (Kleman-Leyer *et al.*, 1997). Four proteinase cuts observed in this region were all protected in the protein–RNA complex, demonstrating that this region is involved in RNA binding (Figure 4A–C). The co-crystal structure of the U1A spliceosomal protein with an RNA hairpin ligand demonstrated that the RNP1 motif generates a β-strand that can interact with an RNA substrate via stacking of an aromatic residue on a nucleotide base (Oubridge *et al.*, 1994). The RNP1-like segment V/VI region contains three fully conserved aromatic residues, all protected on substrate binding, and it is likely that one or more of these residues stack on bases in the 3 nt bulges at the cleavage site junctions, which may be partly unstacked in solution (Lykke-Andersen and Garrett, 1994). Segment Ib, which is also protected on substrate binding, is positioned just N-terminal of the putative leucine zipper (see above). A third region protected by the RNA substrate lies C-terminal of a conserved histidine involved in catalysis (see below).

The cleavage reaction pathway

Ribonucleases generating 2′-3′-cyclic phosphates or 3′-phosphates on cleavage, like the three families of ribonucleases homologous to RNase A, RNase T_1 and RNase T_2/Rh, use a hydrogen acceptor and a hydrogen donor residue to promote catalysis (Deavin *et al.*, 1966; Wodak *et al.*, 1977; Gohda *et al.*, 1994; Kurihara *et al.*, 1996). The hydrogen acceptor residue is a histidine in the RNase A family and a glutamate and/or histidine in the RNase T_1 and T_2/Rh families, and they extract the proton from the 2′-hydroxyl group at the cleavage site. This leads to nucleophilic attack by the 2′-oxoanion on the cleavage site phosphate and the resulting 5′-oxoanion leaving group is subsequently protonated by the hydrogen donor residue, which is a histidine in all three RNase families. The observation that the eukaryotic and archaeal splicing endoribonucleases, like the above-mentioned RNase families, generate 2′,3′-cyclic phosphates on cleavage (Peebles *et al.*, 1983; Kjems and Garrett, 1988; Thompson and Daniels, 1988) and that a substrate lacking a 2′-OH at the cleavage junctions was bound to but not cleaved by the enzyme (Figure 3B) points to a similar acid–base catalysed cleavage mechanism for the splicing reaction (Figure 8C). For the *M.jannaschii* enzyme a histidine residue (His125) is involved in catalysis and mutation at the same histidine in the Sen34p subunit of the yeast enzyme also showed impaired cleavage activity (Trotta *et al.*, 1997). Since histidines can both donate and accept protons, we cannot definitely decide what the function of this residue is, although it is most likely to be a proton donating residue, since the two other histidine residues found in the *M.jannaschii* enzyme are not conserved. Mutation of four possible candidates for proton acceptor activity, one of which (Glu49) is a fully conserved acidic residue of the enzymes, did not seriously impair cleavage (Figure 6A). Possible explanations include that, as may be true for RNase T_1 and RNase Rh (Gohda *et al.*, 1994; Kurihara *et al.*, 1996), more than one residue can act as proton acceptor or that the enzyme uses a solvent hydroxyl ion, which eventually could be positioned and activated by a divalent metal ion, as occurs for self-cleaving hammerhead ribozymes, which also generate 2′,3′-cyclic phosphates on cleavage (Scott *et al.*, 1996). In support of the latter explanation, human and plant tRNA introns were found to be cleaved at, and near, each of the exon–intron junctions in the absence of protein (van Tol *et al.*, 1989; Weber *et al.*, 1996). Resolution of this problem must probably await the co-crystallization of the enzyme–RNA substrate complex, which could be achieved using the uncleavable substrate analogue described here.

Materials and methods

Preparation of GST splicing endoribonuclease fusion proteins

PCR was performed on *M.jannaschii* total genomic DNA, using oligodeoxynucleotide primers 5′-GAGGATCCATGGTGAGAGATAAAATGGGGCA (upstream primer) and 5′-GAGAATTCCTGGTTTACATAGGTCATGTT (downstream primer). All PCRs were performed with the Pfu DNA polymerase (Stratagene) at 0.02 U/µl in reactions containing 20 mM Tris–HCl, pH 8.75, 2 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 0.1% Triton X-100, 10 µg/ml bovine serum albumin, 0.25 mM dNTPs and 0.2 µM oligodeoxynucleotide primers. Products generated with the upstream and downstream primers contained *Bam*HI and *Eco*RI restriction sites at the upstream and downstream ends respectively. These products were ligated into *Bam*HI/*Eco*RI-cleaved pGEX-2TK (Pharmacia) and pGEX-GTH vectors (Jensen *et al.*, 1995a) and the resulting recombinant plasmids were transformed into *E.coli* BL21/DE3 cells (containing the F-factor from *E.coli* Xl1-blue) and GST fusion proteins were expressed and purified by a general procedure (Lykke-Andersen *et al.*, 1996). After purification the GST tags of the fusion proteins were removed by cleavage with the endoproteinase thrombin.

Preparation of intron transcripts

In vitro transcripts of introns with flanking exons from the 23S rRNA of *S.marinus* and *D.mobilis* were prepared from the linearized T7 promotor-containing plasmids described earlier (Lykke-Andersen and Garrett, 1994). The 36 nt DmU1 and the 54 nt DmU2 substrates were prepared by the method described by Milligan and Uhlenbeck (1989). Oligodeoxynucleotides 5′-GGCGCTATCGGGGGGCTCTCGCCCTATCCCGGCGCCTATAGTGAGTCGTATTA and 5′-CTGGCGGCGCTATCGGGGGGCCGGGTCTCCCCGGCCCTATCCCGGCGCCGCCCCTATAGTGAGTCGTATTA, for DmU1 and DmU2 respectively, were mixed in water with the oligodeoxynucleotide 5′-TAATACGACTCACTATAG at 1 µM, heated at 95°C for 5 min and placed on ice. Transcription reactions were performed with 0.1 µM annealed oligonucleotides, 1 mM NTPs and 10 µg/ml T7 RNA polymerase in 40 mM Tris–HCl, pH 8.0, 2 mM $MnCl_2$, 5 mM DTT, 1 mM spermidine for 2 h at 37°C. Mn^{2+} was included instead of Mg^{2+} in the transcription reactions, which resulted in higher RNA yields. DmU1 and DmU2 transcripts containing deoxyuridines at the splice junctions were generated as described by Conrad *et al.* (1995), by replacing rUTP with dUTP at the same concentration. The 36 nt *H.volcanii* $tRNA^{Trp}$ substrate derivative was a gift from John Diener and Professor Peter Moore (Yale University). All RNAs were purified from denaturing polyacrylamide gels and concentrations were estimated from A_{260} measurements.

Intron cleavage assays

Assays with the splicing endoribonuclease were performed with either 5′- or 3′-^{32}P-end-labelled, ^{32}P-internally labelled or unlabelled RNA transcripts as specified in the figure legends. RNA substrates were incubated with the purified splicing enzyme in cleavage buffer (20 mM HEPES–KOH, pH 8.0, 100 mM KCl, 10 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine) at different temperatures and incubation times. Reactions were stopped by adding 4 vol. ice-cold 0.3 M sodium acetate, 5 mM EDTA and extracted with phenol:$CHCl_3$ (1:1, sodium acetate-saturated). RNA was ethanol precipitated, washed and dried. In reactions with labelled RNA the pellet was dissolved in 10 µl formamide, 5 mM EDTA, denatured at 95°C for 30 s and run on denaturing polyacrylamide gels before subjecting to autoradiography. Reactions with unlabelled RNAs were analysed by primer extension assays as described earlier (Lykke-Andersen and Garrett, 1994).

Protein cross-linking

About 100 000 d.p.m. (~50 ng) C-terminally labelled splicing enzyme was incubated at 65°C for 5 min in 50 µl cleavage buffer, followed by addition of glutardialdehyde at different concentrations. After 2 h incubation, 25 µl 0.1 M Tris base, 1.5% SDS, 100 mM DTT, 15% glycerol and 0.2% bromophenol blue was added. Samples were denatured at 95°C for 2 min and 15 µl aliquots were run on an 8% polyacrylamide–SDS–tricin gel before being subjected to autoradiography.

Protein footprinting

N- and C-terminally ^{32}P-labelled splicing enzymes were prepared as described earlier for other proteins (Lykke-Andersen *et al.*, 1996). In protein footprinting experiments ~20 000 d.p.m. ^{32}P-labelled enzyme (~0.1 pmol) were mixed with 1 pmol deoxyuridine-containing DmU2 substrate or *E.coli* tRNA at the same concentration (w/v) in 10 µl cleavage buffer containing 0.2 mg/ml bovine serum albumin. The mixture was incubated at 65°C for 10 min, followed by 50°C for 5 min and 2 µl proteinase were added. Proteinase concentrations were similar to those used earlier (Lykke-Andersen *et al.*, 1996). After incubating at 50°C for 15 min the reactions were stopped by adding 6 µl 0.1 M Tris base, 1.5% SDS, 100 mM DTT, 15% glycerol and 0.2% bromophenol blue. The samples were denatured at 95°C for 2 min before running on 35×43×0.04 cm 7% stacking/20% separation polyacrylamide–SDS–tricin gels and being subjected to autoradiography.

RNA footprinting

A stable splicing enzyme–RNA complex was generated by mixing 5′- or 3′-^{32}P-end-labelled DmU2-dU (~50 000 d.p.m., 1 pmol) with ~10 pmol freshly purified splicing enzyme in 50 µl cleavage buffer and incubating the mixture at 65°C for 10 min. The complex was subsequently probed at 65°C with 5 µl 20 mM $Fe(NH_4)_2SO_4$, 50 mM EDTA, 100 mM ascorbate and 2 µl 10% H_2O_2 for 1 min or with 1 µl DEP (>97% pure; Sigma) for 2 min. RNA incubated with BSA and RNA–enzyme complex incubated in the absence of chemical probe were included as control samples. Reactions were stopped by adding 200 µl ice-cold 100 mM thiourea, 0.3 M sodium acetate, pH 6.0, 5 mM EDTA, 50 ng/µl heavily methylated *E.coli* tRNA (Christiansen *et al.*, 1990) and 750 µl ethanol. The RNA was precipitated, redissolved in 200 µl 0.3 M sodium acetate, pH 6.0, 5 mM EDTA, extracted with phenol (sodium acetate-saturated) and $CHCl_3$, precipitated with 500 µl ethanol, washed in 80% ethanol and dried. For the DEP reactions strand cleavage was induced by dissolving the pellet in 20 µl 30% acetic acid, 10% distilled aniline and incubating for 20 min at 60°C in the dark, before addition of 180 µl 0.3 M sodium acetate, pH 6.0, 5 mM EDTA and 600 µl ethanol. RNA was precipitated, redissolved and phenol extracted as described above. RNA samples were dissolved in 10 µl deionized formamide, 5 mM EDTA, pH 8.0, and 5 µl of each reaction were applied to 20% polyacrylamide–7 M urea gels.

Preparation of splicing enzyme mutants

Splicing enzyme mutants were prepared by PCR using Pfu DNA polymerase (Stratagene). For each mutation the upstream and downstream oligonucleotide primers used for cloning of the splicing enzyme gene (see above) were used in separate PCRs, each together with an internal mutagenic primer encompassing the mutation site on the nonsense and the sense strand respectively. The PCR products were purified from an agarose gel, ~0.01 pmol each DNA fragment was mixed in a 100 µl PCR containing no primers and the mixture was subjected to 15 cycles of 94°C 30 s, 42°C 45 s, 72°C 1.5 min, before adding 20 pmol upstream and downstream primer and continuing incubation for 30 cycles of 94°C 30 s, 50°C 45 s, 72°C 1.5 min. The PCR product was cleaved with restriction endoribonucleases *Bam*HI and *Eco*RI, gel purified and ligated into *Bam*HI/*Eco*RI-cleaved pGEX-GTH vector (Jensen *et al.*, 1995a). Mutant proteins were expressed and purified as described above for the unmutated protein.

Acknowledgements

We thank Dr Charles Daniels for sharing his *H.volcanii* splicing enzyme sequence before publication and for an earlier 4 year collaboration (unpublished) on purifying the splicing enzyme from *M.barkeri* and John Diener and Peter Moore for sharing their RNA substrate. Hoa Phan Thi-Ngoc is thanked for excellent technical assistance. This research was supported by the Novo-Nordisk Fund and the Biotechnology program of the Danish Research Councils. J.L. was supported by Copenhagen University.

References

Belford,H.G., Westaway,S.K., Abelson,J. and Greer,C.L. (1993) Multiple nucleotide cofactor use by yeast ligase in tRNA splicing: evidence for independent ATP- and GTP-binding sites. *J. Biol. Chem.*, **268**, 2444–2450.

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

Burggraf,S., Larsen,N., Woese,C.R. and Stetter,K.O. (1993) An intron within the 16S ribosomal RNA gene of the archaeon *Pyrobaculum aerophilum*. *Proc. Natl Acad. Sci. USA*, **90**, 2547–2550.

Cech,T.R. (1990) Self-splicing of group I introns. *Annu. Rev. Biochem.*, **59**, 543–568.

Christiansen,J., Egebjerg,J., Larsen,N. and Garrett,R.A. (1990) Analysis of rRNA secondary structure: experimental and theoretical considerations. In Spedding,G. (ed.), *Ribosomes and Protein Synthesis: A Practical Approach*. IRL press, Oxford, UK, pp. 229–252.

Conrad,F., Hanne,A., Gaur,R.K. and Krupp,G. (1995) Enzymatic synthesis of 2′-modified nucleic acids: identification of important phosphate and ribose moieties in RNase P substrates. *Nucleic Acids Res*., **23**, 1845–1853.

Dalgaard,J.Z. and Garrett,R.A. (1992) Protein-coding introns from the 23S rRNA-encoding gene form stable circles in hyperthermophilic arhaeon *Pyrobaculum organotrophum*. *Gene*, **121**, 103–110.

Dalgaard,J.Z., Garrett,R.A. and Belfort,M. (1993) A site-specific endonuclease encoded by a typical archaeal intron. *Proc. Natl Acad. Sci. USA*, **90**, 5414–5417.

Deavin,A., Mathias,A.P. and Rabin,B.R. (1966) Mechanism of action of bovine pancreatic ribonuclease. *Nature*, **211**, 252–255.

Fitz-Gibbon,S., Choi,A.J., Miller,J.H., Stetter,K.O., Simon,M.I., Swanson,R. and Kim,U.-J. (1997) A fosmid-based genomic map and identification of 474 genes of the hyperthermophilic archaeon *Pyrobaculum aerophilum*. *Extremophiles*, **1**, 36–51.

Garrett,R.A. *et al*. (1994) Archaeal rRNA operons, intron splicing and homing endonucleases, RNA polymerase operons and phylogeny. *System. Appl. Microbiol*., **16**, 680–691.

Gohda,K., Oka,K., Tomita,K. and Hakoshima,T. (1994) Crystal structure of RNase T1 complexed with the product nucleotide 3′-GMP. *J. Biol. Chem*., **269**, 17531–17536.

Heyduk,T., Heyduk,E., Severinov,K., Tang,H. and Ebright,R.H. (1996) Determinants of RNA polymerase alpha subunit for interaction with beta, beta′, and sigma subunits: hydroxyl-radical protein footprinting. *Proc. Natl Acad. Sci. USA*, **93**, 10162–10166.

Jensen,T.H., Jensen,A. and Kjems,J. (1995a) Tools for the production and purification of full-length, N- or C-terminal ^{32}P-labeled protein, applied to HIV-1 Gag and Rev. *Gene*, **162**, 235–237.

Jensen,T.H., Leffers,H. and Kjems,J. (1995b) Intermolecular binding sites of human immunodeficiency virus type 1 Rev protein determined by protein footprinting. *J. Biol. Chem*., **270**, 13777–13784.

Kjems,J. and Garrett,R.A. (1985) An intron in the 23S ribosomal RNA gene of the archaebacterium *Desulfurococcus mobilis*. *Nature*, **318**, 675–677.

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

Kjems,J. and Garrett,R.A. (1991) Ribosomal RNA introns in archaea and evidence for RNA conformational changes associated with splicing. *Proc. Natl Acad. Sci. USA*, **88**, 439–443.

Kjems,J., Jensen,J., Olesen,T. and Garrett,R.A. (1989) Comparison of transfer RNA and ribosomal RNA splicing in the extreme thermophile and archaebacterium *Desulfurococcus mobilis*. *Can. J. Microbiol*., **35**, 210–214.

Kjems,J., Leffers,H., Olesen,T. and Garrett,R.A. (1989) A unique tRNA intron in the variable loop of the extreme thermophile *Thermofilum pendens* and its possible evolutionary implications. *J. Biol. Chem*., **264**, 17834–17837.

Kleman-Leyer,K., Armbruster,D.W. and Daniels,C.J. (1997) Properties of *H. volcanii* intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. *Cell*, **89**, 839–847.

Kurihara,H., Nonaka,T., Mitsui,Y., Ohgi,K., Irie,M. and Nakamura,K.T. (1996) The crystal structure of ribonuclease Rh from *Rhizopus niveus* at 2.0 Å resolution. *J. Mol. Biol*., **255**, 310–320.

Lykke-Andersen,J. and Garrett,R.A. (1994) Structural characteristics of the stable RNA introns of archaeal hyperthermophiles and their splicing junctions. *J. Mol. Biol*., **245**, 846–855.

Lykke-Andersen,J., Thi-Ngoc,H.P. and Garrett,R.A. (1994) DNA substrate specificity and cleavage kinetics of an archaeal homing-type endonuclease from *Pyrobaculum organotrophum*. *Nucleic Acids Res*., **22**, 4583–4590.

Lykke-Andersen,J., Garrett,R.A. and Kjems,J. (1996) Protein footprinting approach to mapping DNA binding sites of two archaeal homing endonucleases: evidence for a two-domain protein structure. *Nucleic Acids Res*., **24**, 3982–3989.

Lykke-Andersen,J., Aagaard,C., Semionenkov,M. and Garrett,R.A. (1997) Archaeal introns: splicing, intercellular mobility and evolution. *Trends Biochem. Sci*., **22**, 326–331.

Michel,F. and Ferat,J.L. (1995) Structure and activities of group II introns. *Annu. Rev. Biochem*., **64**, 435–461.

Milligan,J.F. and Uhlenbeck,O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol*., **180**, 51–62.

Moore,M.J., Query,C.C. and Sharp,P.A. (1993) Splicing of precursers to mRNA by the spliceosome. In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

Oubridge,C., Ito,N., Evans,P.R., Teo,C.H. and Nagai,K. (1994) Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.

Peebles,C.L., Gegenheimer,P. and Abelson,J. (1983) Precise excision of intervening sequences from precurser tRNAs by a membrane-associated yeast endonuclease. *Cell*, **32**, 525–536.

Phizicky,E.M. and Greer,C.L. (1993) Pre-tRNA splicing: variation on a theme or exception to the rule? *Trends Biochem. Sci*., **18**, 31–34.

Rauhut,R., Green,P.R. and Abelson,J. (1990) Yeast tRNA-splicing endonuclease is a heterotrimeric enzyme. *J. Biol. Chem*., **265**, 18180–18184.

Scott,W.G., Murray,J.B., Arnold,J., Stoddard,B.L. and Klug,A. (1996) Capturing the structure of a catalytic RNA intermediate: the hammerhead ribozyme. *Science*, **274**, 2065–2069.

Tange,T.O., Jensen,T.H. and Kjems,J. (1996) *In vitro* interaction between human immunodeficiency virus type 1 Rev protein and splicing factor ASF/SF2-associated protein, p32. *J. Biol. Chem*., **271**, 10066–10072.

Thompson,L.D. and Daniels,C.J. (1988) A $tRNA^{Trp}$ intron endonuclease from *Halobacterium volcanii*. *J. Biol. Chem*., **263**, 17951–17959.

Thompson,L.D. and Daniels,C.J. (1990) Recognition of exon–intron boundaries by the *Halobacterium volcanii* tRNA intron endonuclease. *J. Biol. Chem*., **265**, 18104–18111.

Trotta,C.R., Miao,F., Arn,E.A., Stevens,S.W., Ho,C.K., Rauhut,R. and Abelson,J.N. (1997) The yeast tRNA splicing endonuclease: a tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell*, **89**, 849–858.

van Tol,H., Gross,H.J. and Beier,H. (1989) Non-enzymatic excision of pre-tRNA introns? *EMBO J*., **8**, 293–300.

Weber,U., Beier,H. and Gross,H.J. (1996) Another heritage from the RNA world: self-excision of intron sequence from nuclear pre-tRNAs. *Nucleic Acids Res*., **24**, 2212–2219.

Westaway,S.K., Belford,H.G., Apostol,B.L., Abelson,J. and Greer,C.L. (1993) Novel activity of a yeast ligase deletion polypeptide: evidence for GTP-dependent tRNA splicing. *J. Biol. Chem*., **268**, 2435–2443.

Wich,G., Leinfelder,W. and Böck,A. (1987) Genes for stable RNA in the extreme thermophile *Thermoproteus tenax*: introns and transcription signals. *EMBO J*., **6**, 523–528.

Wodak,S.Y., Liu,M.Y. and Wyckoff,H.W. (1977) The structure of cytidyl (2′,5′) adenosine when bound to pancreatic ribonuclease S. *J. Mol. Biol*., **116**, 855–875.

Received on June 24, 1997; revised on July 25, 1997