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

Jens Lykke-Andersen and Roger A. Garrett

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

1Corresponding 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 degenerate 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 transesterific-
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 32P-labelled at either the N-terminus (pGEX-2TK) or the C-terminus (pGEX-GTH) using heart muscle kinase and [γ-32P]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 M.jannaschii splicing 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 Mg2+.

The splicing endoribonuclease from M.jannaschii is a homotetramer

Protein cross-linking experiments were performed to investigate the quaternary structure of the endoribonuclease. C-terminally 32P-labelled enzyme was incubated at
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–tricine 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 always present with the full-length purified splicing enzyme (see also Figure 4A).

**Fig. 2.** Quaternary structure of the splicing enzyme determined by protein cross-linking. 32P-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–tricine 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 visible, probably due to concurrent cleavage at the 5′-site was confirmed on a 20% polyacrylamide–7 M urea gel and autoradiographed. Basic hydrolysis reactions (OH –) were performed on products from 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 products from the enzyme (~0.05 pmol) for 5 or 60 min as indicated. Products were separated on 20% polyacrylamide–7 M urea gels and autoradiographed.

**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′-deoxyribonucleotides at each cleavage site, indicated by arrows. The dissociation constants (Kd) for binding of the splicing enzyme to DmU1 and DmU2, as determined in (C), are given below the sequences. (B) Cleavage assays on 5′-32P-end-labelled DmU1 (1 pmol) containing uridines (DmU1-rU) or deoxyuridines (DmU1-dU) at the cleavage junctions or deoxyuridine-containing DmU1 gel purified after precinubation 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′-32P-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 32P-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.
Archaeal 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 32P-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 Kd. 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 Fe2+/EDTA/H2O2 (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 intensities, 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 Kd value is ~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 protease reactivity 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 (Klemann-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 (Klemann-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 (α`_2)_2, where α is similar in size and structure to α` (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 endonuclease 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 endonuclease has been sequenced or will be sequenced in genome sequencing projects (*). In the eukaryotic domain α_4 and ω_2 indicate proteins with primary structures like the tetrameric M.jannaschii and duplicated dimeric H.volcanii enzymes respectively (see Discussion). In the eukaryotic domain ω_2 represents the yeast heterotetrameric enzyme, of which the α_4 (Sen34p) 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 archaean 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–IX), 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 footprinting. The M.jannaschii tetramer (α老年人2), which corresponds to the H.volcanii dimer (α老年人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 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老年人2) and the M.jannaschii enzyme is a tetramer [(α老年人2)老年人4], each dimeric subunit of the M.jannaschii tetramer (α老年人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老年人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老年人α和α老年人α老年人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老年人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 RNPI 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 RNPI 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 RNPI-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 T1 and RNase T2/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 T1 and T2/Rh families, and they extract the proton from the 2'-hydroxyl group at the cleavage site. This leads to nucleophilic attack by the 2'-oxygen on the cleavage site phosphate and the resulting 5'-oxygen 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 T1 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'-GAGGATCCAGGCTGACATTTAAAGGATCCGCTGTCGGTGGCCGGTGTT (upstream primer) and 5'-GAGGATCCCTCCTGTTATACAGTGCTGATTT (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 MgSO4, 10 mM (NH4)2SO4, 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 BamHI and EcoRI restriction sites at the upstream and downstream ends respectively. These products were ligated into BamHI/EcoRI-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 X1L-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 endoproteasome 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 promoter-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'-GGCTATGCTGGGGCTTCCGGCATCT-CACTGCGGCGGGGTCGATGCTGGATTAAT and 5'-CTGCGGCGGGTATCGGGGGCCCGGCCGTCCATAG, were used 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 9.0, 2 mM MnCl2, 5 mM DTT, 1 mM spermidine for 2 h at 37°C. Mn2+ was added to the reaction mixture, 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 DmU1 and the 54 nt DmU2 substrates were used in separate PCRs, each together with an internal mutagenic primer encompassing the mutation site on the nonsense and unmutated protein. 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.arkeri 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


Archaeal splicing endoribonuclease

A stable splicing enzyme–RNA complex was generated by mixing 5'- or 3'-32P-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(NH4)2SO4, 50 mM EDTA, 100 mM ascorbate and 2 μl 10% H2O2 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 thiouria, 0.3 M sodium acetate, pH 6.0, 5 mM EDTA, 50 ng/μl heavily methylated E.coli rRNA (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 CHCl3, 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, 42°C 45 s, 72°C 1.5 min. The PCR product was cleaved with restriction endoribonucleases BamHI and EcoRI, gel purified and ligated into BamHI/EcoRI-cleaved pGEX-GTH vector (Jensen et al., 1995a). Mutant proteins were expressed and purified as described above for the unmutated protein.


