The Composition, Function and Evolution of the tRNA Splicing Endonuclease

Thesis by

Christopher R. Trotta

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1999

(Defended June 17, 1998)

Acknowledgements

Ultimate thanks are due Dr. John Abelson for creating an incredible atmosphere within his laboratory which has allowed me to learn how great science is done. John has been a constant source of academic inspiration and enlightenment and I am very honored to have been able to complete my degree under his guidance.

Thanks also to members of the Abelson lab without which I would have surely faultered. Specifically I owe a debt of gratitude to Drs. Peggy Saks, Jaime Arenas, John Wagner, Scott Stevens and Jeff Sampson who provided me with sage scientific advice as well as incredible personal support. Thanks also to Drs., Dan Ryan, Randy Story, Tracy Johnson, Imre Barta, Andrey Balakin, and Tau-Mu Yi who have all contributed to my success in one way or another over the years.

Huge thanks to Dr. Hong Li who has provided an extremely fruitful collaboration.

And of course thanks are due to my fellow Entruvidorian counterparts, Chang Hee Kim for being there during my darkest hour and Dr. Eric Arn for showing me the path to good science and good "rock 'n roll."

Thanks to DJ for being there during all of the best times and making the bad ones seem bearable.

Special thanks to Jon, Chris, Steve, Alan, Bill, Rick, Patrick, Trevor, Billy, Peter, Geoff and Tony, for providing the ultimate inspiration, the "keys" to my soul.

Finally thanks to my family for their constant and unyielding support from afar.

Abstract

The splicing of tRNA precursors is essential for the production of mature tRNA in organisms from all major phyla. In Bacteria, intron removal is through autocatalysis of group I introns. In Archaea and Eukaryotes, intron removal is dependent on protein-based enzymes. Recognition and cleavage of the splice sites is accomplished by the tRNA splicing endonuclease. In order to understand how the eukaryotic endonuclease accomplishes this task, the genes encoding all four subunits of the *S. cerevisiae* enzyme have been cloned. All four genes are essential. Two subunits, Sen2 and Sen34, contain a homologous domain of approximately 130 amino acids. Surprisingly this domain is found in the gene encoding the archaeal tRNA splicing endonuclease of *H.volcanii* and in other Archaea. Utilizing the alignment as a guide, a mutation was made in the yeast Sen34 subunit which demonstrated that the eukaryotic endonuclease contains two functionally independent active sites for cleavage of the 5' and 3' splice sites, encoded by the SEN2 and SEN34 genes, respectively. The homology to the archaeal enzymes suggests an ancient origin for the tRNA splicing reaction. Exploiting the smaller and simpler archaeal version of the endonuclease, a crystal structure of the *Methanoccocus jannaschii* enzyme was determined to a resolution of 2.3 angstroms. The structure indicates that the cleavage reaction is similar to that of ribonucleaseA. Insight gained from the architecture of this homotetrameric enzyme has allowed for a clearer understanding of the heterotetrameric splicing endonuclease of yeast. In particular, the eukaryotic enzyme has preserved the important structural features found in the archaeal enzyme which allow for precise spatial positioning of the two active sites for cleavage.

Table of Contents

Chapter I - Review: tRNA Splicing - An RNA World Add-On or an Ancient Reaction? 1

- Introduction 1
- Intron Containing tRNA 1
- The Enzymatic Gymnastics of tRNA Splicing 8
- Characterization of the tRNA Splicing Enzymes 12
- Recognition of pre-tRNA by the tRNA Splicing Endonuclease 17
- The Genes for the Endonuclease: A Bridge Between Kingdoms 21
- tRNA Splicing at the Atomic Level 21
- On the Evolution of the Endonucleases 30
- Origin of tRNA Introns 36
- Concluding Remarks 38
- References 38

Chapter II - The Yeast tRNA Splicing Endonuclease: A Two Active Site Tetrameric Enzyme Homologous to the Archaeal tRNA Splicing Endonucleases 44

- Abstract 44
- Introduction 44
- Results 50
- Discussion 77
- Experimental Procedures 81
- Acknowledgments 88
- References 89

Chapter III – Three-Dimensional Structure and Evolution of a tRNA Splicing Enzyme 94

- Abstract 94
- Introduction 94
- Structure Determination 99
- A Two Domain Monomer 99
- The Homotetramer: A Dimer of Two Dimers 105
- Subunit Interfaces 108
- Active Sites 110
- Implications for the Structure of Other Splicing Endonucleases 124
- References 133

Chapter IV – Investigations into the Endonuclease Active Site and Architecture 136

- Introduction 136
- Dissection of the Interaction Domains of the Archaeal Endonuclease 136

The Endonuclease Active Site 140
The Molecular Ruler 142
Materials and Methods 145

Appendix I 147
Appendix II 152
Appendix III 163

Chapter I – Review: tRNA Splicing: An RNA World Add-On or an Ancient Reaction? (to be published in RNA World II eds. Tom Cech, Ray Gesteland and John Atkins. Cold Spring Harbor Laboratory Press, 1998; a shorter version of this chapter written as a minireview was published in JBC 273:12685-12688, 1998; Appendix I.)

INTRODUCTION

Introns are encoded in the genes for tRNA in organisms from all three kingdoms of life. Their removal is an essential step in the maturation of tRNA precursors. In Bacteria, introns are self-splicing and are removed by a Group I splicing mechanism (Biniszkiewicz et al., 1994; Kuhsel et al., 1990; Reinhold-Hurek and Shub, 1992). In Eukaryotes and Archaea, intron removal is mediated enzymatically by proteins. Recent progress in understanding both eukaryotic and archaeal tRNA splicing has revealed that the two processes, previously thought to be unrelated, are in fact similar. Insight gained from the comparison has provided a clearer understanding of intron recognition, the catalysis of intron removal and has given new insight into the evolution of the tRNA splicing process.

INTRON-CONTAINING tRNA

tRNA Introns in Eukaryotes

Intervening sequences (introns) were discovered 20 years ago in the yeast genes for the tyrosine-inserting nonsense suppressor tRNA (Goodman et al.,1977) and for $tRNA^{Phe}$ (Valenzuela et al., 1978). With the completion of the yeast genome, it is known that of the 274 yeast tRNA genes 61 or 20% contain introns. Table I-1 lists the tRNAs which contain introns. PCR cloning of tRNAs from higher eukaryotes has revealed a similar distribution of intron-containing tRNA (Green et al., 1990; Schneider et al., 1993) Stange and Beier, 1986). The introns in all of the genes are small (14-61 bases) and they

Yeast tRNA Precursors Containing Introns

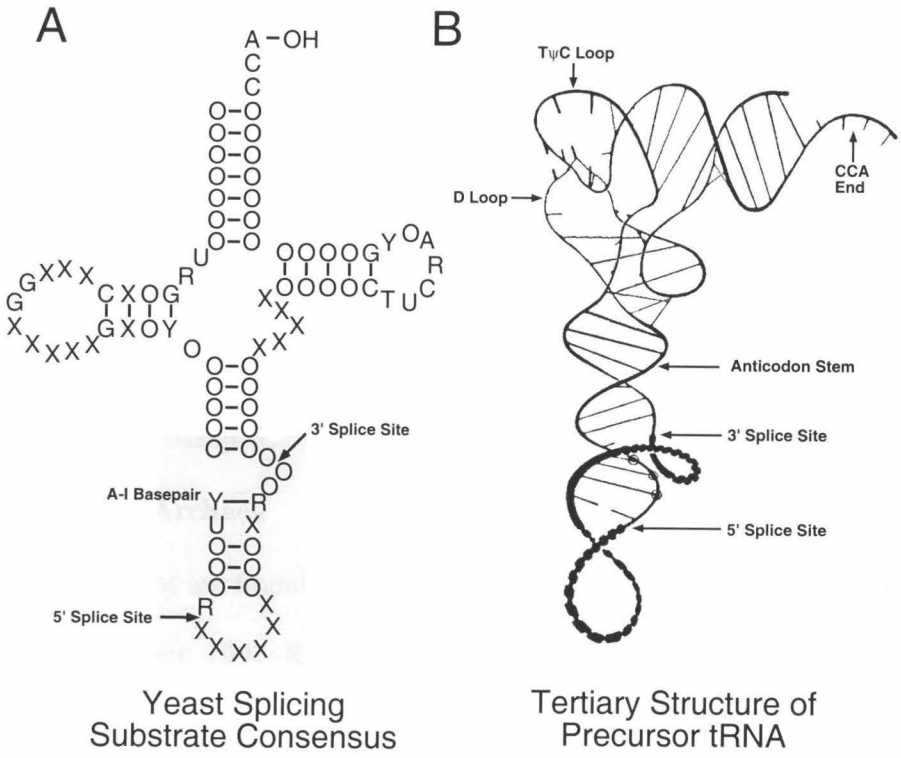
tRNA	Intron Length (nucleotides)	No. of Genes
tRNA $^{Ser}_{CGA}$	19	1
tRNA $^{Ser}_{GCU}$	19	4
tRNA $^{Lys}_{UUU}$	23	7
tRNA $^{Pro}_{UGG}$	31,30,33	7,2,1
tRNA $^{Trp}_{CCA}$	34	6
tRNA $^{Phe}_{GAA}$	18,19	3,7
tRNA $^{Leu}_{CAA}$	32,33	8,2
tRNA $^{Ile}_{UAU}$	60	2
tRNA $^{Leu}_{UAG}$	19	3
tRNA $^{Tyr}_{GTA}$	14	8
		Total = 61
	Total tRNA Genes = 274	

Table I-1

FIGURE I-1. Eukaryotic precursor tRNA

A. Consensus of yeast pre-tRNA. O = nonconserved base in a region of conserved length or secondary structure; X= nonconserved base in a region of variable length or secondary structure. G,C,A,U = conserved base among the pre-tRNAs. Y and R = conserved pyrimidines and purines, respectively. The conserved anticodon-intron (A-I) base-pair is depicted.

B. Model of the tertiary structure of pre-tRNAs based on the crystal structure of yeast $tRNA^{Phe}$ (from Lee and Knapp, 1985). Intron is indicated in bold, dashed lines.



Yeast Splicing Substrate Consensus

Tertiary Structure of Precursor tRNA

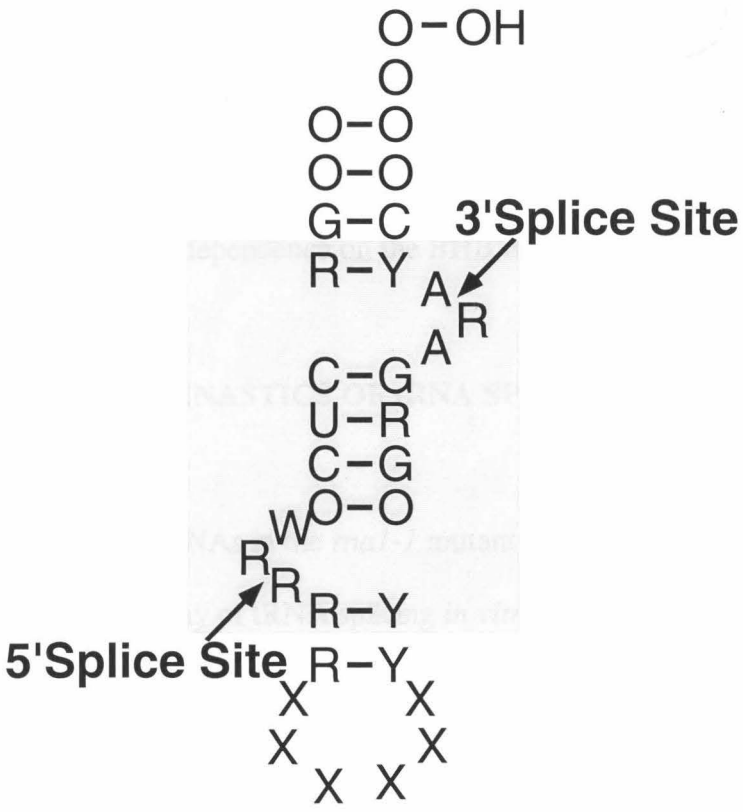
Figure I-1

are all located in the same position, one base to the 3' side of the anticodon (Figure I-1A). Structure probing revealed that the common "cloverleaf" tertiary structure seen in the crystal structure of $tRNA^{Phe}$ is maintained in intron-containing tRNAs and that the intron, including the splice sites, is the most exposed region of the molecule (Lee and Knapp, 1985; Swerdlow and Guthrie, 1984). This led to a proposed model for the tertiary structure of pre-tRNA shown in Figure I-1B. More recent experiments by Tocchini-Valentini and co-workers (Baldi et al., 1992) have demonstrated that a conserved base pair (the A-I base pair) between a base of the 5'-exon (position 32) immediately following the anticodon stem and a base in the single-stranded loop of the intron (position -3) is required for correct excision at the 3' splice site (Figure I-1A). Thus the distinctive feature in eukaryotic intron-containing pre-tRNAs is the position of the intron and the presence of the A-I basepair.

tRNA Introns in Archaea

tRNA introns are found in every archaeal species which has been studied (Lykke-Anderson and Garrett, 1997; Klenk et al., 1998; Kleman-Leyer et al., 1997; Thompson et al., 1989). Though they are often similar, archaeal introns are different from their eukaryotic counterparts. Comparative sequence analysis of a host of archaeal introns has revealed that the 5' and 3' splice sites are always located in a three nucleotide bulge separated by a four base-pair helix: the Bulge-Helix-Bulge (BHB) motif (Figure I-2) (Thompson et al., 1989). Intron length is variable from a short intron of 33 nucleotides found in the $tRNA^{Met}$ gene of *Methanococcus jannaschii*, to 105 nucleotides of the $tRNA^{Trp}$ gene of *Haloferax volcanii* (Daniels et al., 1985), but the intron is required to

FIGURE I-2. Consensus archaeal substrate. Positions depicted as in Figure I-1A. The bulge-helix-bulge (BHB) motif is shaded in grey.



Archaeal Splicing
Substrate Consensus

Figure I-2

base-pair with the 5'exon to allow formation of the BHB motif. In general, the introns reside one base 3' to the anticodon, but introns have been found elsewhere in the tRNA molecule. One such example is that of the $tRNA^{Pro}$ gene of *Methanobacterium thermoautotrophicum* where there are two introns in the anticodon stem that must be removed in an obligate order to produce the mature tRNA (Trotta, unpublished). Another repository for introns in Archaea is the rRNA genes. In fact, one of the earliest discovered archaeal introns, the 622 nucleotide intron of the 23S rRNA of *Desulfurococcus mobilis*, was shown through extensive biochemical characterization to be removed by the archaeal tRNA splicing system with dependence on the BHB motif (Kjems and Garrett, 1985; Kjems et al., 1989).

THE ENZYMATIC GYMNASTICS OF tRNA SPLICING

The Eukaryotic Pathway

The accumulation of pre-tRNAs in the *rna1-1* mutant (Hopper et al. 1978) provided a source of pre-tRNA for assay of tRNA splicing *in vitro* (Knapp et al. 1978). The system was employed to work out the pathway of tRNA splicing (Figure I-3). The first step of the reaction is recognition and cleavage of the pre-tRNA substrate at the 5' and 3' splice sites by a tRNA splicing endonuclease. The cleavage by the endonuclease at the splice sites results in 2',3' cyclic phosphate and 5'-hydroxyl termini (Peebles et al., 1983). The tRNA 5'- and 3'-exons are then the substrate for a series of reactions catalyzed by the tRNA splicing ligase (for a detailed review of the ligation mechanism see Arn and Abelson, 1998). Ligase is a multi-functional enzyme possessing phosphodiesterase, polynucleotide kinase, and RNA ligase activities. In the ligase reaction, phosphodiesterase opens the 2',3' cyclic phosphate to give a 2'-phosphate. Next the 5'-hydroxyl is phosphorylated by transfer of the γ-phosphate from an NTP cofactor.

FIGURE I-3. The tRNA splicing pathway of yeast. Gene names are given in parentheses for the proteins of the pathway. CPDase, cyclic phosphdiesterase; ASTase, adenylyl synthetase. See text for details (Taken with permission from Abelson et al., 1998).

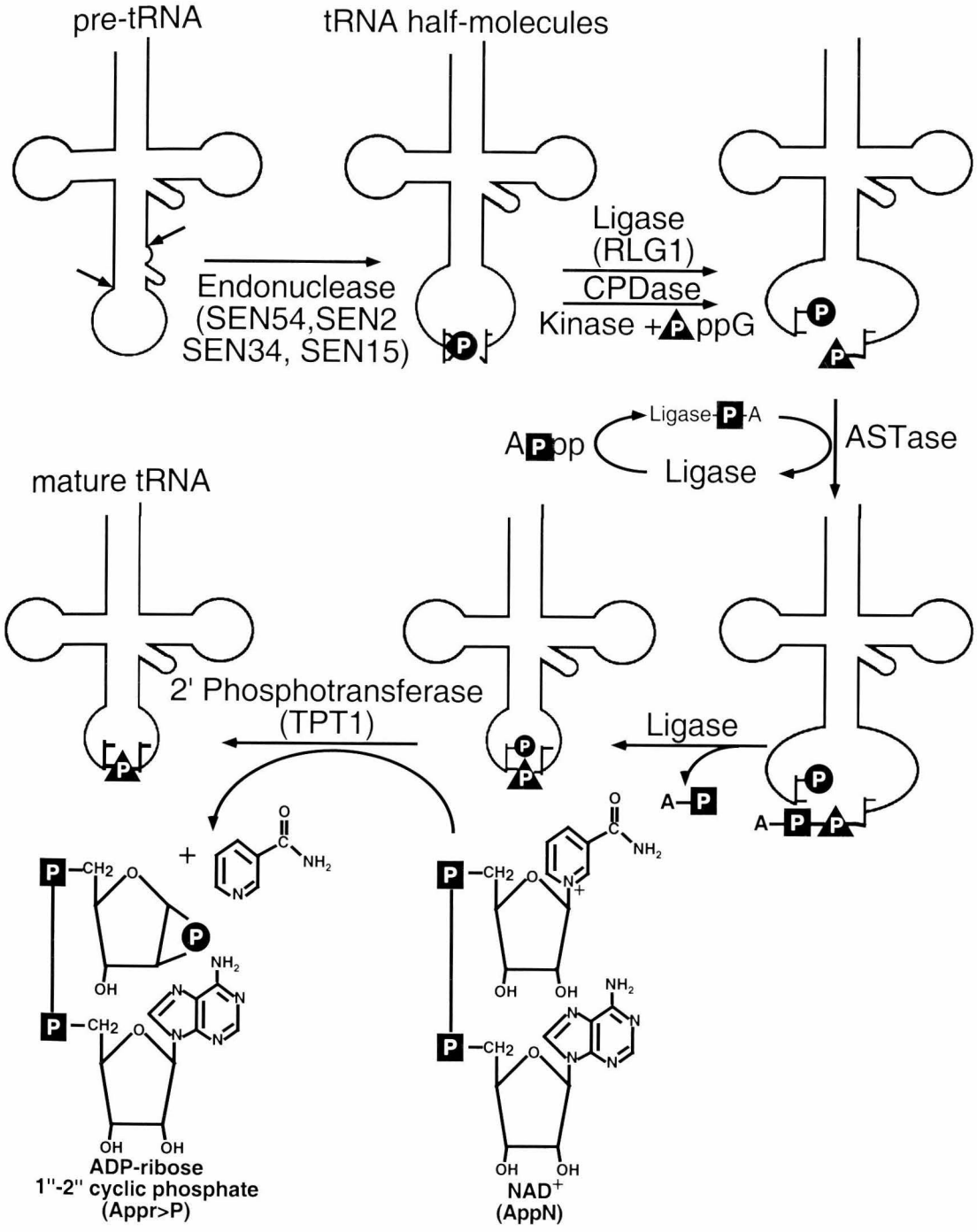

pre-tRNA
tRNA half-molecules
Endonuclease
(SEN54,SEN2
SEN34, SEN15)
Ligase
(RLG1)
CPDase
Kinase +pppG
ASTase
Ligase
Ligase
mature tRNA
2' Phosphotransferase
(TPT1)
Ligase
ADP-ribose
1''-2'' cyclic phosphate
(Appr>P)
NAD+
(AppN)

Figure I-3

Ligase is then adenylylated and the AMP moiety is transferred to the 5'-phosphate on the 3'-exon forming an activated phosphoanhydride. Finally ligase joins the two exons by catalyzing an attack of the 3'-hydroxyl on the activated donor phosphoanhydride to form a 3',5' phosphodiester, 2'-monophosphoester bond with the release of AMP. This results in a pre-tRNA molecule with a 2'-phosphate at the splice junction. Removal of the 2'-phosphate is accomplished by the activity of a 2'-phosphotransferase that catalyzes the transfer of the phosphate to NAD, releasing nicotinamide and yielding ADP-ribose 1"-2" cyclic phosphate and mature tRNA (Appr>P; Figure I-3).

The Archaeal Pathway

Much less is known about the enzymes involved in the processing of archaeal pre-tRNA due to the difficulty in working with extracts from these organisms. As in the eukaryotic pathway, removal of introns requires the function of both an endonuclease and a ligase. The archaeal endonuclease recognizes substrates containing the consensus BHB motif and cleaves at the 5' and 3' splice sites to yield 2',3' cyclic phosphate and 5'-hydroxyl termini (Thompson et al., 1989). Thus the cleavage mechanism is identical to the eukaryotic mechanism providing the earliest clue to the relatedness of the two enzymes (see below). Upon cleavage by the endonuclease, the tRNA half molecules must be ligated to produce a mature tRNA, but the ligase activity has yet to be defined. It is clear that the yeast mechanism does not pertain since the γ-phosphate of ATP is not incorporated into spliced tRNA as seen in the yeast mechanism (Kjems and Garrett, 1985; Gomes and Gupta, 1997).

CHARACTERIZATION OF THE tRNA SPLICING ENZYMES

The Eukaryotic tRNA Endonuclease (described in detail in Chapter II)

The first step of tRNA splicing is the recognition and cleavage of the splice sites by the tRNA splicing endonuclease. Utilizing the *in vitro* system for tRNA splicing, the endonuclease was shown to be nuclear membrane-associated (Peebles et al., 1983). The detergent, Triton X-100, at 0.9% is required to release the activity from the membrane fraction. In a particularly difficult purification, the endonuclease was purified one-million-fold from a yeast nuclear membrane preparation to homogeneity and appeared to be a $\alpha\beta\gamma$ trimer of 54, 44, and 34 kD subunits (Rauhut et al., 1990). Later a fourth subunit of 15kD was discovered (Trotta et al., 1997). Thus the enzyme is a $\alpha\beta\gamma\delta$ heterotetramer. Cloning of the genes encoding each of these protein subunits was accomplished by both genetic and protein chemistry. Ho et al. (1990) screened a bank of temperature sensitive mutants and isolated a yeast mutant that accumulated 2/3 tRNA precursor molecules that were not cleaved at the 5' splice site. The gene, SEN2 (splicing endonuclease (Winey and Culbertson, 1988), was cloned by complementation and antibodies to the product were shown to recognize the 44kD subunit of the purified splicing endonuclease. Development of a simpler affinity purification utilizing a tagged SEN2 gene allowed preparation of sufficient material to identify the other three gene products, SEN54, SEN34 and SEN15 (Trotta et al., 1997).

Two of the subunits, SEN2 and SEN34, contain a homologous domain of approximately 130 amino acids. SEN2 contains the only transmembrane sequence and presumably mediates membrane association of the enzyme. All four subunits contain a nuclear localization sequence. Two-hybrid interactions were detected between Sen54-Sen2 and Sen34-Sen15. Since the enzyme is a stable tetramer, interaction must occur

between these two presumed heterodimers although they are not detected in the two-hybrid assay (Trotta et al., 1997).

The Eukaryotic tRNA Ligase

The purification of the soluble tRNA ligase from yeast was simpler and quickly led to the isolation of a 95kD polypeptide and the cloning of its essential gene, RLG1 (Phizicky et al., 1986; Phizicky et al., 1992). This single polypeptide carries out all three of the steps in the ligase mechanism (Figure I-3). Interestingly, partial proteolysis of the protein resulted in three fragments that were shown to contain distinct activities supporting a domain-like structure of the enzyme (Xu et al., 1990). The amino-terminal fragment is adenylylated at lysine 114 and sequence comparisons have shown this lysine residue to be equivalent to the active lysine of the T4 RNA ligase (Apostol et al., 1991; Xu et al., 1990). The carboxy-terminal fragment contains the cyclic phosphodiesterase activity. The central domain then would be the likely locus of the kinase activity, however, an enzyme deleted in this domain paradoxically retains ligase activity. The enzyme has two nucleotide-binding sites, for GTP and for ATP (Belford et al., 1993). GTP is employed in the kinase step and ATP is used for the formation of the activated adenyly-RNA intermediate. The *in vivo* function of such a complex NTP requirement is unclear, but it has been suggested that such a requirement could couple the splicing reaction to transcription and/or translation (Belford et al., 1993).

The complexity of the domain structure of ligase raises interesting questions as to the origin of this protein. The yeast tRNA ligase is mechanistically and structurally related to phage T4 RNA ligase and polynucleotide kinase (Apostol et al., 1991). The phage enzymes together contain the three activities present in the tRNA ligase but in addition T4 polynucleotide kinase contains a phosphatase which removes the 2'-phosphate (Walker et

al., 1975). Thus T4 RNA ligase and T4 polynucleotide kinase can ligate pre-tRNA half molecules produced by endonuclease yielding a product that does not contain a 2'-phosphate.

The 2'-Phosphotransferase

Early investigation into the activity responsible for 2'-phosphate removal at the splice junction demonstrated the requirement for two separate components present in yeast extracts (McCraith and Phizicky, 1990). The first component was purified and determined to be the cellular co-factor NAD^+ (McCraith and Phizicky, 1991). As described earlier, NAD^+ is the receptor in a reaction which involves the transfer of the 2'-phosphate at the splice junction to the 2" position of the ribose of NAD^+ yielding the high energy compound ADP-ribose 1"-2" cyclic phosphate (Appr>p) (Culver et al., 1993). The second component was purified and shown to be a 26.2kDa polypeptide encoded by the TPT1 gene (Culver et al., 1997). Expressed in *E. coli*, the single polypeptide could catalyze the transfer of the 2'-phosphate to NAD^+. The gene was shown to be essential in yeast and it was suggested that its essential role could be either removal of the 2'-phosphate from all intron-containing tRNA molecules or generation of the novel molecule Appr>p. To further understand the role of each of these potential functions, a conditional lethal phosphotransferase mutant was generated (Spinelli et al., 1997). Yeast expressing this mutant accumulated tRNA molecules with a 2'-phosphate at the splice junction. Interestingly, these tRNA molecules were undermodified at positions near the splice junction residue while other modifications appeared to be normal. These results suggest that the removal of the 2'-phosphate is essential for correct modification of the residues near the splice junction and that tRNA containing the 2'-phosphate is not a substrate for

the modification enzyme and likely inactive in carrying out its function in translation. The results also suggest a temporal order for the maturation of pre-tRNA with splicing and 2'-phosphate removal occurring before modification of certain positions in the tRNA molecule. The TPT1 gene was also found to have homologs in other eukaryotes and surprisingly in *E. coli* and some Archaea (Culver et al., 1997; Trotta, 1997). The function of this protein in Bacteria and Archaea is unclear since a source of 2'-phosphates is not apparent (see the discussion on the archaeal ligase mechanism below). *E. coli* does not possess intron-containing tRNA or a eukaryotic-like RNA ligase. It will therefore be interesting to elucidate its role in bacteria hopefully providing a clue as to the origin of this unique processing event.

The Archaeal tRNA Endonuclease

The endonuclease of members of the Archaea proved to be just as difficult to purify as those of Eukaryotes. Daniels and co-workers chose the halophilic archaeon *Haloferax volcanii* to carry out the purification. The enzyme is present in low abundance and is extremely difficult to stabilize during the purification protocol. When finally purified the endonuclease of *H. volcanii* proved to be composed of a single 37kDa protein encoded by the EndA gene (Kleman-Leyer et al., 1997). Unlike the heterotetrameric yeast enzyme, the active enzyme is composed of a single subunit shown by gel filtration to behave as a homodimer in solution. The protein was shown to contain a 130 amino acid domain homologous to the two yeast subunits SEN2 and SEN34. This was the observation that unified the two lines of research and led to the progress described below.

The Archaeal tRNA Ligase

The identity of the archaeal tRNA ligase is at present unknown; however, with the complete genome sequence of four members of the Archaea, it is clear that there is no

homolog of the yeast tRNA splicing ligase (a remarkable statement to be able to make, now common in the genomic sequence era). Thus it appears that ligation of tRNA half molecules occurs via a different reaction mechanism to that of the eukaryotic ligation reaction.

All sequenced archaeal genomes do, however, contain an open reading frame encoding a protein homologous to the 2'-5' ligase of *E. coli* characterized by E. Arn (Arn and Abelson, 1996). This protein functions *in vitro* as an RNA ligase joining tRNA half molecules to form a 2'-5' phosphodiester bond at the splice junction. The exact mechanism of this reaction is unknown but it does not require the addition of exogenous ATP. Thus it is tempting to speculate that the homolog found in Archaea may function to ligate the tRNA half-molecules of the splicing reaction. 2'-5' linkages have not been detected in tRNA but it is possible that the archaeal ligase has an altered specificity and catalyzes the formation of a 3'-5' linkage. Clearly, it will be interesting to further investigate the archaeal homologues of the 2'-5' ligase.

tRNA Splicing in Higher Eukaryotes

Like the yeast tRNA introns, the position of the introns in higher eukaryotes is conserved and the cleavage reaction operates in a manner identical to the yeast tRNA splicing endonuclease as exemplified by the *Xenopus* endonuclease. Ligation of the half molecules to yield mature tRNA has been shown to proceed through a mechanism similar to the yeast ligase (Konarska et al., 1981; Schwartz et al., 1983; Zillmann et al., 1991). Using HeLa cell extracts, Zilmann et al. (1991) have demonstrated the incorporation of exogenous γ-phosphate from ATP upon ligation of yeast $tRNA^{Phe}$, a hallmark of the yeast tRNA ligase reaction mechanism. Furthermore, they demonstrated the presence of a 2'-phosphate at the splice junction. A 2'-phosphotransferase activity was detected in both

HeLa extracts (Zillmann et al., 1992) and is implicated in the dephosphorylation of ligated tRNA in microinjected *Xenopus* oocytes (Culver et al., 1993). However, early studies implicated a different ligase in tRNA splicing in HeLa extracts and *Xenopus* oocytes. It was demonstrated that the product of the tRNA splicing reactions catalyzed by these extracts did not contain a 2'-phosphate at the splice junction. This novel ligase uses the cyclic phosphate at the end of the 5' half of tRNA to generate the 3',5'-phosphodiester bond in the mature tRNA as there is no incorporation of exogenous phosphate at the splice junction (Filipowicz et al., 1983; Laski et al., 1983; Nishikura and De Robertis, 1981). Thus there appears to be redundant pathways for ligation of tRNA molecules in metazoans. To date, no homolog of the bacterial 2'-5' RNA ligase has been sequenced in higher eukaryotes. The further characterization of the alternate eukaryotic ligase activity could conceivably shed light on the nature of the archaeal ligase mechanism.

RECOGNITION OF PRE-tRNA BY THE tRNA SPLICING ENDONUCLEASE

Eukaryotic Endonuclease: the Ruler Mechanism

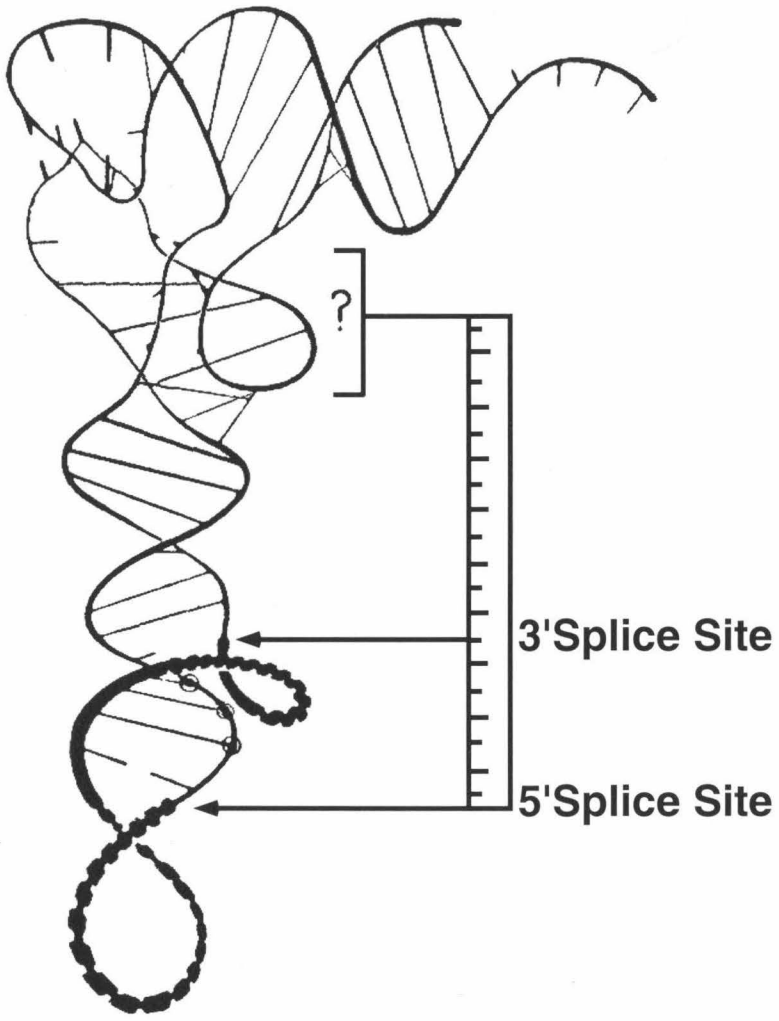
Early speculation concerning the ability of the tRNA endonuclease to recognize and cleave tRNA substrates relied on comparison of the primary and tertiary sequences of the precursor molecules. As previously mentioned there is no sequence conservation around the splice sites and the precursors all fold into the same tertiary structure. The introns are of variable sequence and can be altered by both addition and deletion of nucleotides as well as changes in the sequence with little effect on removal by endonuclease (Johnson et al., 1980; Reyes and Abelson, 1988; Strobel and Abelson, 1986). The exception to this generalization is the A-I base pair necessary for accurate cleavage at the 3' splice site. Mutations which abolish this base pair are not substrates for the splicing endonuclease (Baldi et al., 1992; Willis et al., 1984). This led to the

hypothesis that recognition involved specific interactions with the conserved tertiary structure present in the mature domain of the pre-tRNA. Experimental support for this hypothesis involved the engineering of mutant pre-tRNA which altered distinct features of the primary, secondary or tertiary structures of the molecule (reviewed in Culbertson and Winey, 1989). This suggested a ruler model for recognition and cleavage by the endonuclease in which the enzyme recognizes the mature domain and measures the conserved distance to the splice sites (Greer et al., 1987; Reyes and Abelson, 1988) (Figure I-4). To test this model, tRNA molecules were constructed in which the conserved distance was altered by insertion and deletion of base pairs in the anticodon stem. Addition of a single base pair increased the length of the intron by 2 nucleotides, one at either end (Reyes and Abelson, 1988) as predicted by the model.

Archaeal Endonuclease: the Bulge-Helix-Bulge Motif

In contrast to the yeast results, it was demonstrated that for the archaeal endonuclease, the pre-tRNA substrate could be drastically altered and cleavage would still occur. Synthetic tRNA molecules where almost the entire tRNA mature domain was removed were accurately cleaved by the endonuclease (Thompson and Daniels, 1988). As in the case of the eukaryotic endonuclease, the intron plays a mostly passive role as large deletions can be tolerated by the enzyme. There is, however, a strict requirement for sequence and structure present at the exon-intron boundaries (Thompson and Daniels, 1988). This helix-bulge-helix (BHB) motif is the only requirement for intron recognition and removal by the archaeal tRNA splicing endonuclease.

FIGURE I-4. Ruler model for interaction of the yeast endonuclease with pre-tRNA (see text; taken from Reyes and Abelson, 1988).



Yeast Endonuclease
Ruler Model

Figure I-4

THE GENES FOR THE ENDONUCLEASE: A BRIDGE BETWEEN KINGDOMS (Discussed in detail in Chapter II)

Once the genes for the eukaryotic and archaeal endonuclease were in hand, a comparison between the yeast subunits and the single *H. volcanii* endonuclease revealed that these proteins were related. A domain of 130 amino acid is found in the C-terminus of the SEN2 and SEN34 genes of the yeast endonuclease and in the *H. volcanii* endonuclease. Furthermore, homologs of this domain were discovered in the Archaea *M. jannaschii* and *P. aerophilum* where the domain represented the entirety of the protein (Figure I-5). This homology led to the suggestion that the two yeast subunits each contain an active site for the cleavage reaction. It had been demonstrated that a mutant in the SEN2 gene, *sen2-3*, was defective in 5' splice site cleavage (Ho et al., 1990). This mutation, glycine292 to glutamate, lies within the conserved domain suggesting that SEN2 contains the active site for 5' splice site cleavage and by extension that SEN34 carries the active site for 3' splice site cleavage. A mutant in SEN34 changing a conserved histidine at position 242 to alanine resulted in a marked decrease in cleavage at the 3' splice site, while cleavage at the 5' splice site was normal, strongly supporting the two active site model (Trotta et al., 1997).

In the dimeric *H. volcanii* endonuclease, identical subunits cleave the symmetrically disposed splice sites in the BHB substrate. Thus the arrangement of active sites in eukaryotic and archaeal endonucleases is similar (as depicted in Figure I-6).

tRNA SPLICING AT THE ATOMIC LEVEL (Discussed in detail in Chapter III)

Given the relatedness of the tRNA splicing endonuclease from the two kingdoms, it seemed expedient to choose the archaeal system for structural work. The endonuclease of the archaeon *M. jannnaschii* was chosen as this enzyme consists of only the

FIGURE I-5. Graphic representation of archaeal and eukaryotic endonuclease proteins. The sizes of the various proteins (open rectangle) and location of the conserved sequence regions (black box) are indicated. Light grey blocks represent a degenerative domain repeat present in several archaeal endonucleases (see text). Proteins are placed on the Woesnian evolutionary tree. Endonuclease configuration is noted (see text for details): α4=tetramer; α2=dimer; $\alpha\beta\gamma\delta$=heterotetramer; ?=configuration unknown. Protein designations: H.volcanii, *H. volcanii* EndA (AF001578), A.fulgidus, *Archaeoglobus fulgidus* (AE001041); M. jann, *M. jannaschii* (MJ1424); M. thermo, *Methanobacterium thermoautotrophicum* (AF001577); P. horik, *Pyrococcus horikoshii* (AB009474); Sc Sen2, *Saccharomyces cerevisiae* Sen2 (P16658); Sc Sen34, *S. cerevisiae* Sen34 (P39707); Sp Sen34, *Schizosaccharomyces pombe* Sen34; Z. mays, *Zea mays* open reading frame contained in the intron of HMG (X72692); Mouse, Drosophila, human, represent homologs of the endonuclease domain which have been detected in partially sequenced genomic database submissions (Trotta, unpub.).

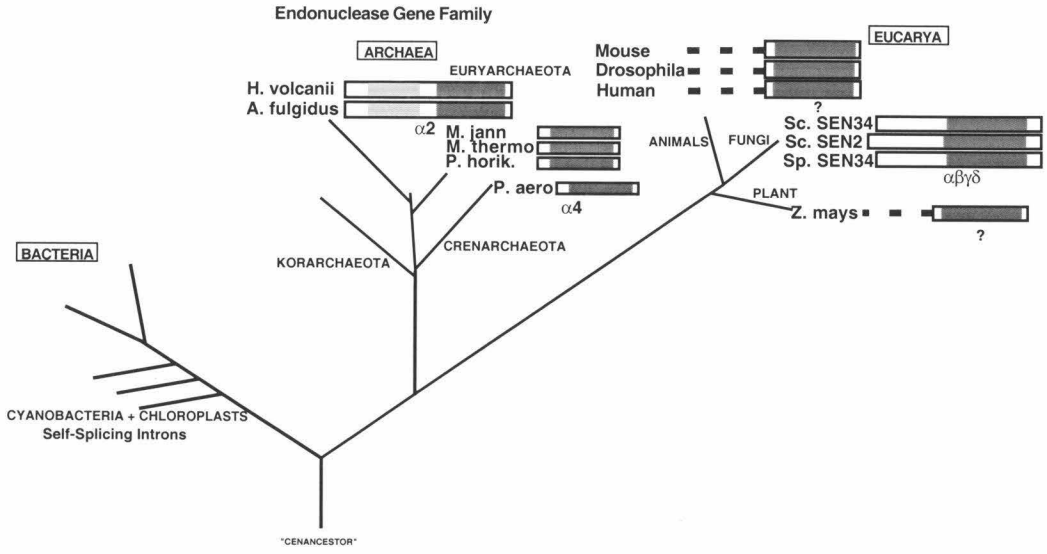

Endonuclease Gene Family
ARCHAEA
EURYARCHAEOTA
H. volcanii
A. fulgidus
α2
M. jann
M. thermo
P. horik.
P. aero
α4
KORARCHAEOTA
CRENARCHAEOTA
BACTERIA
CYANOBACTERIA + CHLOROPLASTS
Self-Splicing Introns
"CENANCESTOR"
EUCARYA
Mouse
Drosophila
Human
?
ANIMALS
FUNGI
Sc. SEN34
Sc. SEN2
Sp. SEN34
αβγδ
PLANT
Z. mays
?

Figure I-5

FIGURE I-6. Comparison of the models proposed for the eukaryotic (yeast) and archaeal (*H.volcanii*) endonucleases (see text for details) (Taken with permission from Li et al., 1998).

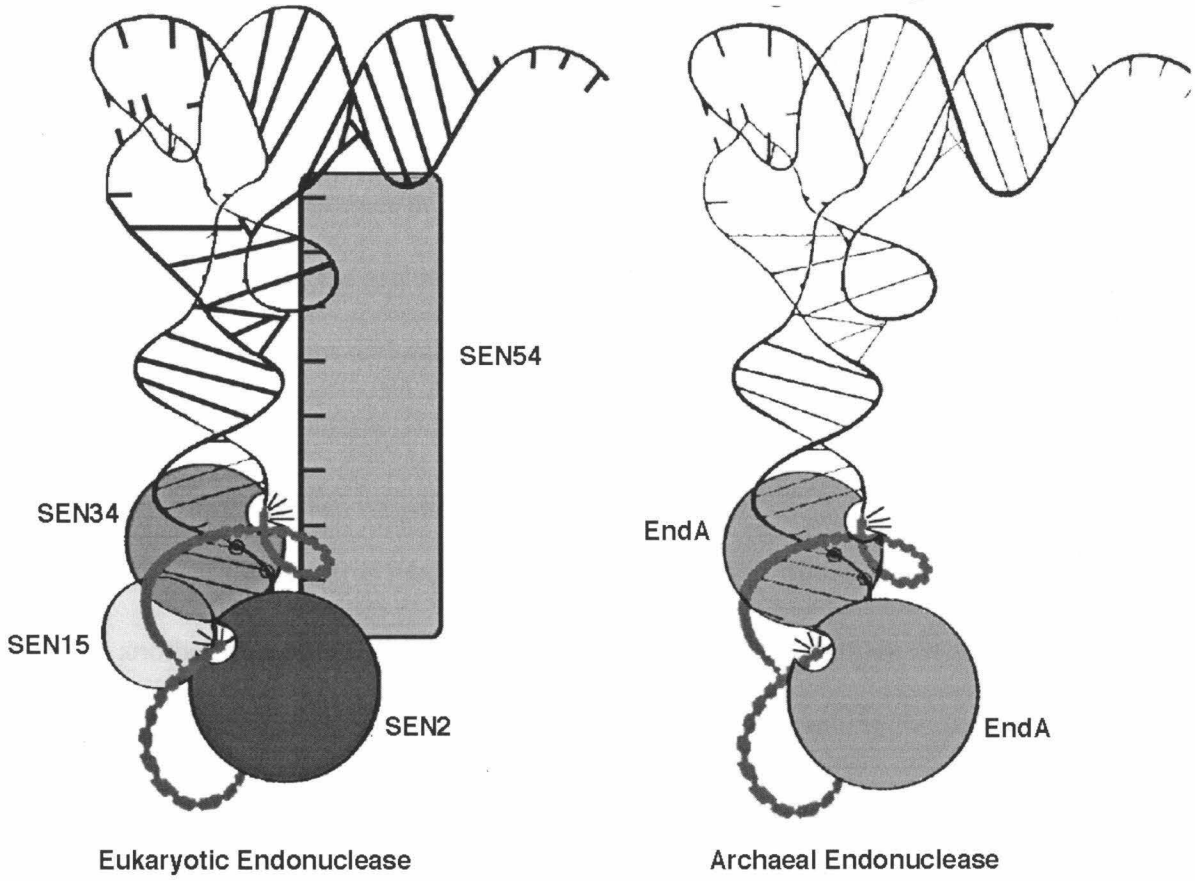

SEN54
SEN34
SEN15
SEN2
Eukaryotic Endonuclease
EndA
EndA
Archaeal Endonuclease

Figure I-6

homologous domain implicated by the endonuclease sequence alignment. It is small and proved easy to express and purify.

Extensive biochemical characterization of this enzyme by Lykke-Andersen and Garrett (1997) showed it to be a homotetramer in solution, an observation confirmed by the crystal structure (Li et al., 1998)(Figure I-7B). Each monomer consists of two domains: the N-terminal domain (residues 9-84) which is composed of three alpha helices and a mixed antiparallel/parallel β-pleated sheet of four strands and the C-terminal domain (residues 85-179) which contains two alpha helices flanking a five-stranded mixed B-sheet (Figure I-7A). The manner in which four of these monomers are brought together provides insight into the architecture and evolution of other members of the endonuclease family of both Archaea and eukaryotes.

Two sets of interactions are crucial to formation of the tetramer. The first interaction involves the formation of an isologous dimer between two monomers. Formation of this dimer is mediated by the interaction of β9 from one monomer and β9' from another monomer. This tail-to-tail interaction results in main chain hydrogen bonds between monomers and leads to a two-stranded β-sheet spanning the subunit boundary. Loop L8 also interacts to form more hydrogen bonds and together these interactions enclose a hydrophobic core at the subunit interface. This leads to an extremely stable dimeric unit that Li et al. (1998) have suggested is conserved in endonucleases from other Archaea and eukaryotes (see below).

The second interaction involves the heterologous interaction between the two dimers which compose the tetramer. The main interaction is an insertion of acidic residues in Loop L10 into a polar groove formed between the N and C terminal domains of the

FIGURE I-7. Crystal structure of the *M. jannaschii* tRNA splicing endonuclease.

A. Ribbon representation of the endonuclease monomer. The proposed catalytic triad residues are within 7 Å of one another and are shown in red ball-and-stick. Also shown is electron density detected near the putative catalytic triad, where the scissile phosphate is proposed to be located for cleavage.

B. Endonuclease tetramer. Each subunit is represented by a distinct label and color. The main chain hydrogen bonds formed between β9 and β8' and between loops L8 and L8' for the formation of isologous dimers are shown as thin lines. Side chains of the hydrophobic residues enclosed at the dimer interface are shown as blue ball-and-stick models. The heterologous interaction which forms the tetramer is mediated by interaction between subunits A1 and B2 (or B1 and A2) through the acidic loops L10 and L8 and are highlighted by dotted surfaces (Taken with permission from Li et al., 1998).

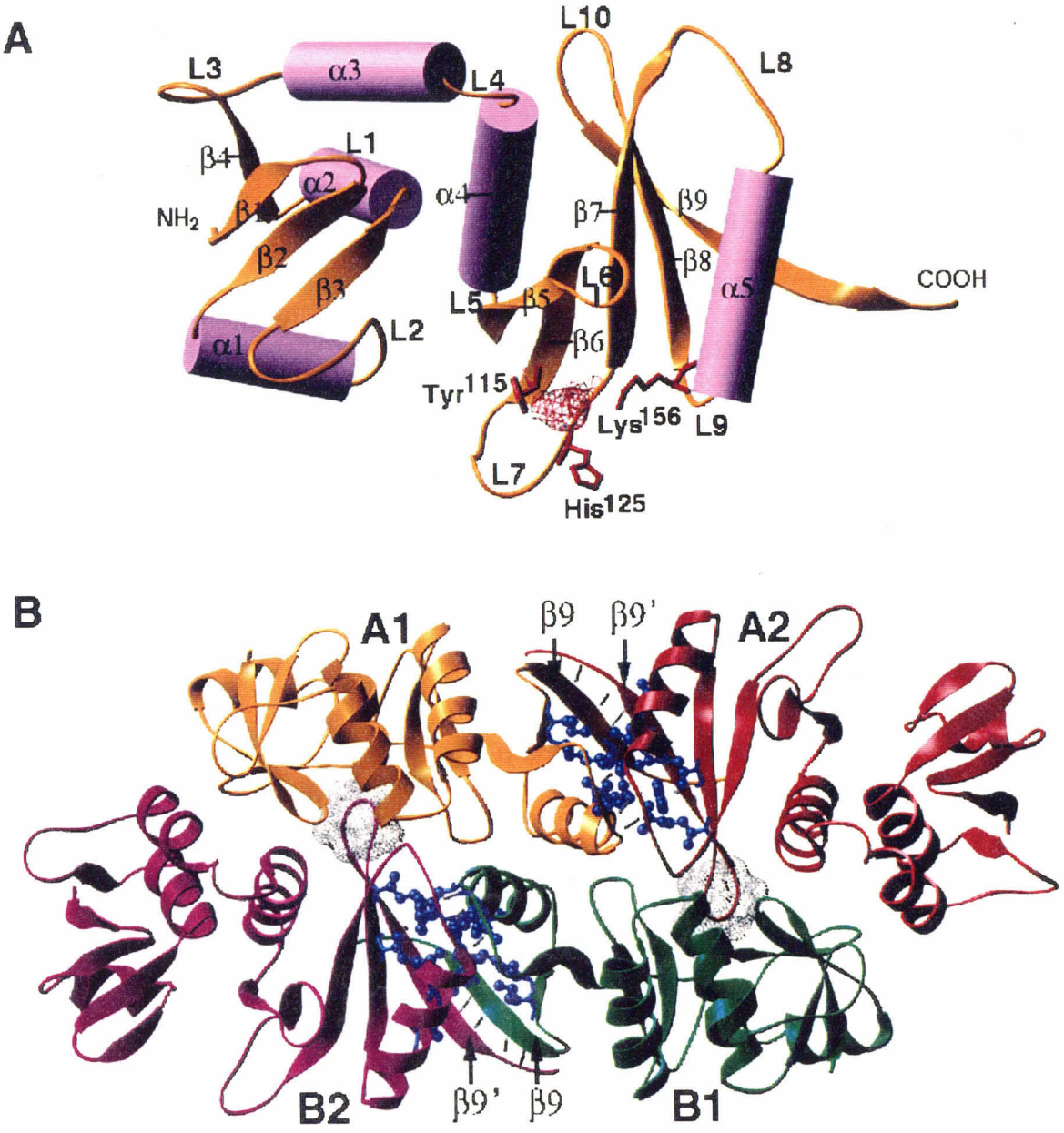

A
L3
α3
L4
L10
L8
β4
L1
α2
α4
β7
β9
NH2
β1
β2
β8
COOH
β3
L6
α5
L5
β5
L2
α1
β6
Tyr115
Lys156
L9
L7
His125
B
A1
β9
β9'
A2
B2
β9'
β9
B1

Figure I-7

endonuclease monomer (Figure I-7B). This causes the two dimers to be translated relative to each other by about 20 Å and brings two subunits, A1 and B1, much closer together than the other two subunits A2 and B2. This interaction is essential to the configuration of the two symmetrically disposed active sites and is also proposed to be conserved in the endonucleases from other organisms (see below).

Active Site of the tRNA Splicing Endonuclease

As discussed earlier, the product of the endonuclease cleavage reaction is a 2',3' cyclic phosphate and a 5-hydroxyl, a product identical to that of other ribonucleases such as RNaseA. In a well characterized acid-base catalyzed reaction RNaseA utilizes a histidine, His12, to abstract a proton from the 2'-hydroxyl of the ribose leading to an in-line attack on the adjacent phosphodiester bond and the formation of a pentacovalent reaction intermediate stabilized by lysine41. The general acid, His119, then protonates the 5'-leaving group leading to the 2',3' cyclic phosphate and 5'-hydroxyl product. In a second step, a proton is abstracted from a water molecule, OH- attacks, and the 2',3' cyclic phosphate is hydrolyzed to the 3'-phosphate (Thompson and Raines, 1994; Walsh, 1979). Inspection of the sequence alignment between the conserved members of the endonuclease family showed a single histidine residue that is absolutely conserved. Recall that the histidine to alanine mutation in Sen34 causes a marked reduction in the catalytic efficiency of 3' splice site cleavage. Similar mutants have been created in this conserved histidine of *M. jannaschii* and *H. volcanii* with a similar reduction in the catalytic efficiency of the cleavage reaction (Daniels, pers. comm.; Lykke-Andersen and Garrett, 1997). Thus it would appear that the tRNA splicing endonucleases catalyze the cleavage of RNA by a general acid-base catalysis.

With the crystal structure for the *M. jannaschii* endonuclease, this histidine was localized at the molecular level. Figure I-7A shows the environment surrounding histidine125 which is found in a cluster with the absolutely conserved residues Tyr115 (from Loop L7) and Lys156 (from α5), forming a pocket into which the scissile phosphate is proposed to fit. The three residues can be spatially superimposed with the catalytic triad of RNaseA (Li et al., 1998) leading to the prediction that His125 of the *M. jannaschii* enzyme is equivalent to the general base, Tyr115 is equivalent to the general acid and Lys156 stabilizes the transition state. Preliminary experiments indicate a role for both His125 and Lys156, but the role of Tyr115 has yet to be established (Trotta, unpub).

Since the *M. jannaschii* enzyme is a homotetramer, each monomer is expected to contain a separate active site for cleavage. However, it is expected that only two of these active sites function in cleavage of the symmetric substrate. Recently, Diener and Moore (1998) have solved the NMR solution structure of a true archaeal substrate molecule. This molecule and a similar modeled substrate based on the TAR RNA structure (Li et. al. 1998) dock in plausible fashion with the enzyme (Figure I-8). The scissile phosphates are shown to fit nicely into the A1 and B1 active sites of the tetramer. The A2 and B2 active sites are too far apart to allow for interaction with the substrate.

ON THE EVOLUTION OF THE ENDONUCLEASES

The Archaeal Endonucleases

The archaeal endonuclease of *H. volcanii* appears to be configured in a different manner than the *M. jannaschii* endonuclease. The protein behaves as a homodimer, not a homotetramer, but must be arrayed in a similar manner due to the identical substrate specificity. Based on the observation that the *H. volcanii* endonuclease is actually an in-frame duplication of the endonuclease domain, Li et al. (1998) proposed a model to

FIGURE I-8. Endonuclease docked with a hypothetical pre-tRNA substrate. The hypothetical tRNA substrate was constructed from joining the known crystal structure of yeast $tRNA^{Phe}$ with the archaeal substrate solved by Diener and Moore (1998). This represents the $pre\text{-}tRNA^{Arch\text{-}Euk}$ demonstrated to be a substrate for both the eukaryotic and archaeal endonucleases (Fabbri et al., 1998) (see text). The scissile phosphates are shown to dock directly in the active sites present on subunits A1 and B1 of the endonuclease.

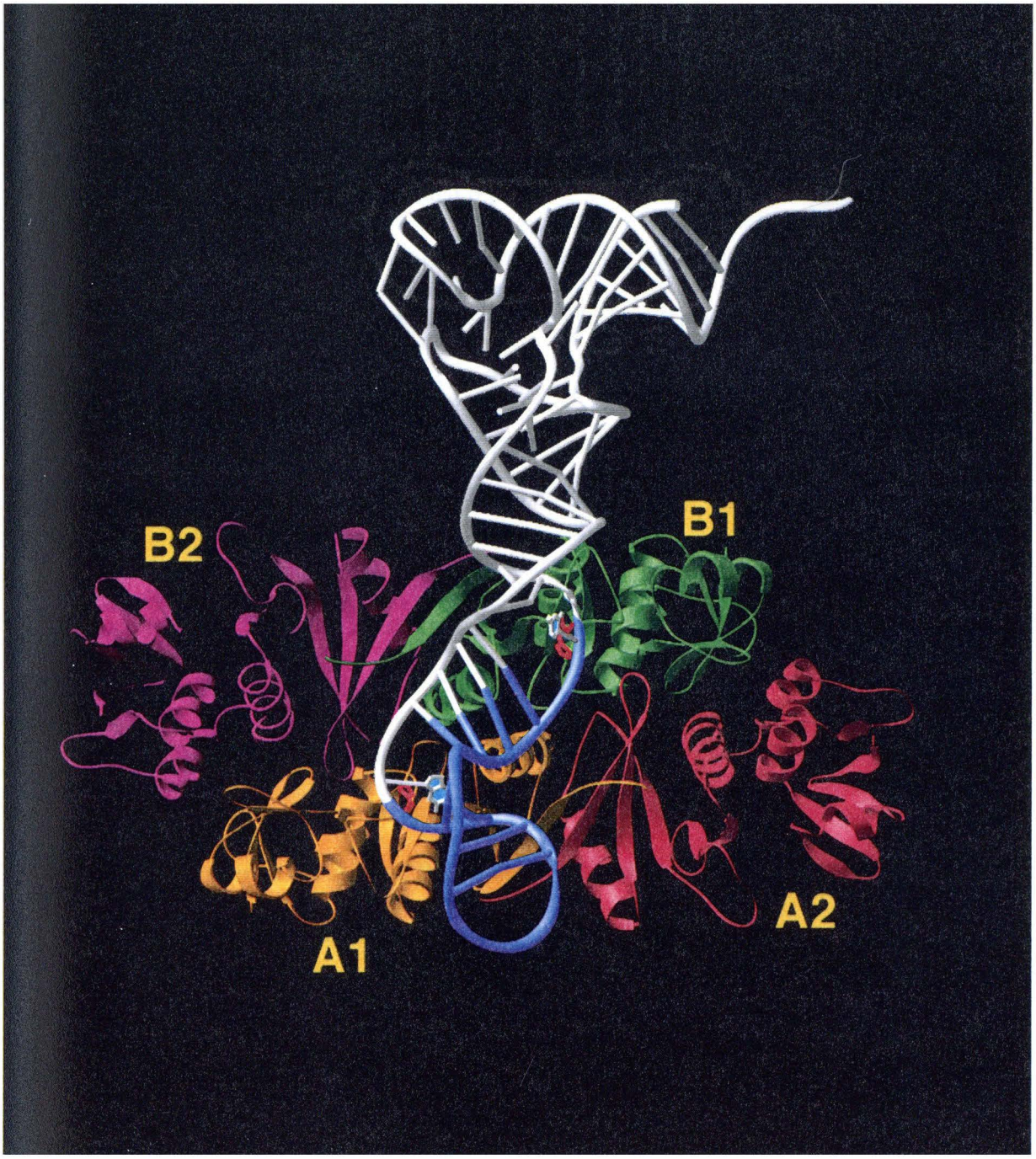


Figure I-8

describe how this enzyme is configured (Figure I-9B). The *H. volcanii* enzyme can be thought of as a pseudo-dimer with the N-terminal endonuclease domain repeat comprising one pseudo-monomer and the C-terminal repeat comprising the other. The repeats are connected by a stretch of polypeptide and serve to form the pseudo-dimer by the β9-β9' hydrophobic interaction which occurs between monomers in the *M. jannaschii* enzyme. Loop L10 present in the N-terminal repeat allows dimerization of the pseudo-dimers to form the active enzyme. Careful examination of the sequence of the N-terminal repeat shows the absence of the amino acids for the catalytic triad, thus the dimeric enzyme contains only two active sites present in the C-terminal repeat.

The occurrence of an in-frame gene duplication event explains the configuration of *H. volcanii* enzyme (Lykke-Andersen and Garrett, 1997). This event has been localized to a set of closely related organisms of the Euryarchaeota branch of the Archaea (Figure I-5) which includes the genus *Haloferax*, *Archaeoglobus* and *Methanosarcina*.

The Yeast Endonuclease

Like the *M. jannaschii* endonuclease, the yeast tRNA splicing endonuclease is a tetramer. The two active site containing subunits, Sen2 and Sen34, must be configured in a similar manner to the A1 and B1 subunits of *M. jannaschii*. This arrangement is proposed to be facilitated by interactions with Sen54 and Sen15. Sen54 and Sen15 both contain a stretch of amino acids at their C-terminus that is homologous to the *M. jannaschii* endonuclease C-terminus. This homology includes the crucial regions involved in the two sets of interactions seen in formation of the tetramer: the β9-β9' hydrophobic tail interaction and the Loop10 insertion sequence. Thus it is proposed (Figure I-9C) that the Sen54-Sen2 and Sen34-Sen15 interactions detected in the two hybrid assay are

FIGURE I-9. Model of the tRNA splicing endonucleases of *M. jannaschii*, *H. volcanii* and *S. cerevisiae*.

A. The *M. jannaschii* endonuclease is graphically depicted. Several important features are shown in the primary sequence and by a model: loop L10 for tetramerization, the COOH-terminal β9 strands (arrows) for dimerization and the conserved catalytic residue His125 (pentagon).

B. The *H. volcanii* endonuclease consists of a tandem repeat of the endonuclease domain. The N-terminal repeat has degenerated to possess only the loop L10 and COOH-terminal interaction domains present in the *M. jannaschii*. The carboxy-terminal repeat contains the groove for loop L10 insertion as well as the catalytic triad including the conserved histidine (pentagon). The dashed line represents the polypeptide chain connecting the amino-terminal repeat and the carboxy-terminal repeat.

C. Proposed structural model of the yeast endonuclease. Upper: the sequence of the carboxy terminus of the M. jann. *M. jannaschii*, H vol. Nt., *H. volcanii* amino-terminal repeat and Sc. Sen54 and Sc. Sen15 is shown to be homologous. Lower: These important interaction elements are modeled in the yeast endonuclease. The two interaction elements are modeled as Loop L10 conserved in Sen54 and Sen15 and the COOH-terminal interaction (circled arrows) (see text for details) (adapted from Li et al., 1998).

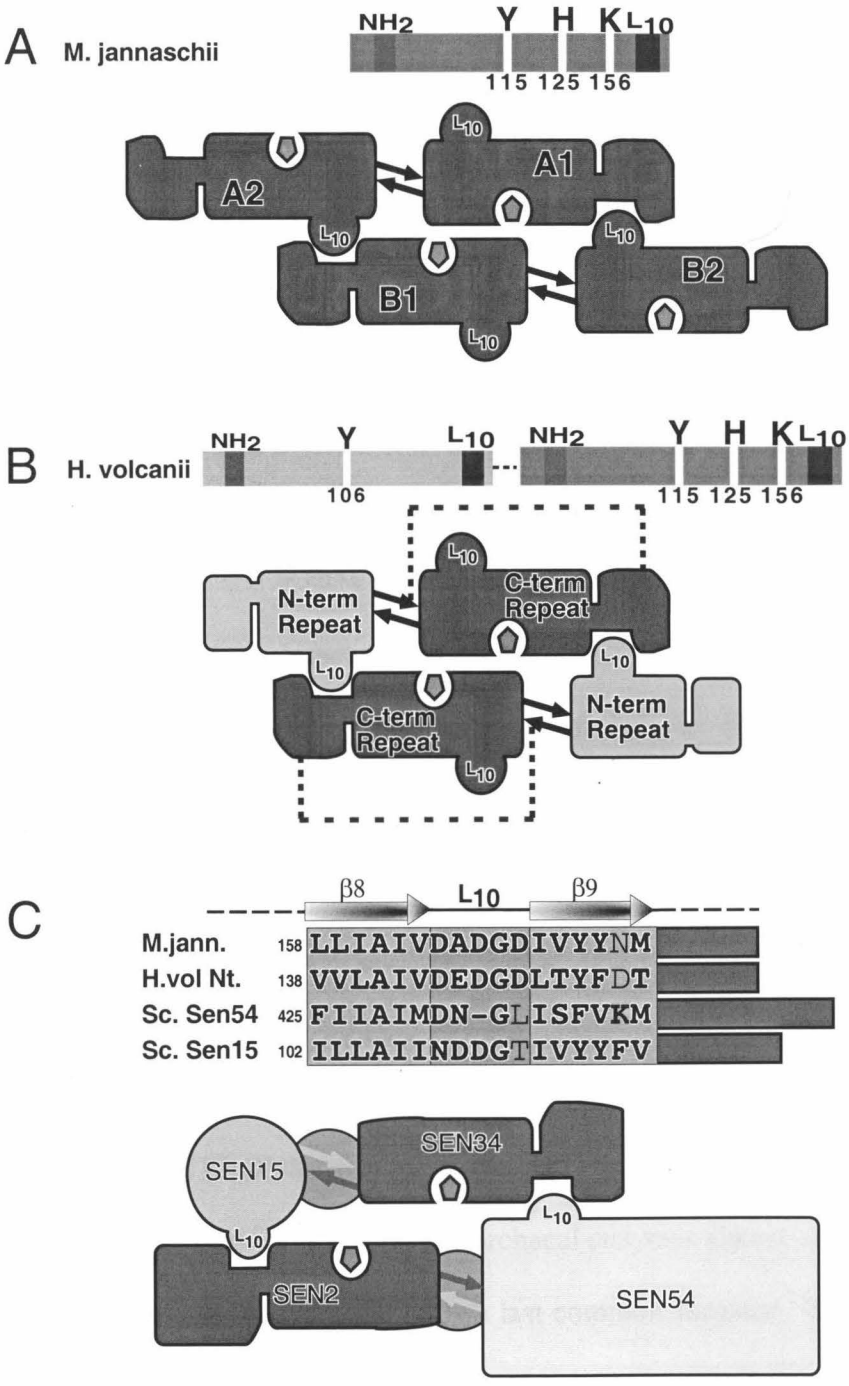

A
M. jannaschii
NH2
Y
H
K
L10
115 125 156
A2
A1
B1
B2
B
H. volcanii
106
N-term Repeat
C-term Repeat
C
β8
β9
M.jann. 158 LLIAIVDADGDIVYYNM
H.vol Nt. 138 VVLAIVDEDGDLTYFDT
Sc. Sen54 425 FIIAIMDN-GLISFVKM
Sc. Sen15 102 ILLAIINDDGTIVYYFV
SEN15
SEN34
SEN2
SEN54

Figure I-9

mediated by the β9-β9' interaction. The Loop L10 interactions are proposed as in the *M. jannaschii* structure to mediate dimer-dimer interactions.

Interestingly it has recently been shown that the eukaryotic endonuclease can recognize and cleave an archaeal bulge-helix-bulge containing tRNA substrate (Fabbri et al., 1998). This tRNA substrate was constructed by creating a hybrid tRNA containing an archaeal intron fused to yeast pre-tRNA^{Phe}. The eukaryotic enzyme can recognize and correctly cleave the 5' and 3' splice sites in this substrate. In doing so, the enzyme dispenses with the ruler mechanism. There is no change in the size or location of the intron released from the pre-tRNA upon addition or deletion of base-pairs in the anticodon stem as was seen for endonucleolytic cleavage of eukaryotic substrates. This result proves that the disposition of the active sites is identical in the archaeal and eukaryotic endonucleases and that this architecture has been conserved since the divergence from a common ancestor. It must be through specialization of the yeast Sen54 and Sen15 subunits that the ruler mechanism has evolved. The SEN2 gene has acquired a transmembrane sequence, which likely anchors the endonuclease in the inner nuclear membrane perhaps near nuclear pore structures. This has been shown to be a primary localization of the tRNA splicing ligase (Clark and Abelson, 1987) and the two enzymes likely act in concert at this site (Greer, 1986).

ORIGIN OF tRNA INTRONS

The relatedness of the eukaryotic and archaeal enzymes almost certainly proves that the endonuclease gene was present in their last common ancestor. The simplest hypothesis is that its function then was to splice tRNA precursors, but it could also be that it had a different function and has been independently recruited in both lines to a tRNA splicing function.

Certainly recruitment is a theme that is a feature of the enzymes in this system. It has recently been shown that tRNA ligase functions in the ligation of an unusually spliced yeast messenger RNA (Cox and Walter, 1996; Sidrauski et al., 1996). The mRNA encodes a transcription factor, HAC1, which upregulates transcription of genes involved in the maintenance of unfolded protein in the endoplasmic reticulum (ER). The presence of unfolded proteins appears to be sensed by a receptor tyrosine kinase-like protein, IRE1, which spans the membrane of the ER. The Ire1 protein contains a nuclease domain capable of cleaving the Hac1 mRNA and releasing an intron (Sidrauski and Walter, 1997). Ligation of the two Hac1 exons is dependent on the function of tRNA ligase both *in vivo* and *in vitro* (Sidrauski et al., 1996; Sidrauski and Walter, 1997).

Two hypothesis have been postulated for the origin of tRNA introns. The first (Cavalier-Smith, 1991) posits that existing proteins were recruited to splice introns resulting from the partial deletion of pre-existing group I or group II self-splicing introns in tRNA or rRNA genes. Support for such a hypothesis is derived from the presence of group I and group II introns in the tRNA genes of a handful of bacterial species. In particular a group I intron found in the $tRNA^{Leu}$ gene of cyanobacteria and bacteria interrupts the anticodon loop at the same position as a protein-dependent intron found in some Archaea (Wich et al., 1987). A second hypothesis proposes the expansion of loops of the tRNA molecule. In both instances the pre-existence of compatible splicing machinery in the cell would insure against lethality of the insertion (Belfort and Weiner, 1997). It is unlikely that agreement on a time for the origin of tRNA introns is possible; however, the new addition to the ongoing dialogue is that at least the splicing endonuclease is ancient.

CONCLUDING REMARKS

Whatever the origin of tRNA introns, it is clear that they are here to stay. Both Archaea and eukaryotes have failed to displace all introns from tRNA genes. This would suggest that the introns serve some selective advantage for the organism and that maintenance of the splicing system is necessary for survival. It is clear from studies in yeast that one function of introns is to aid in the modification of tRNAs (for review see Grosjean et al., 1997). Thus, the modification enzymes have evolved to depend on the presence of the intron for correct substrate recognition and modification.

By studying the enzymology of the removal process, we have begun to understand just how introns have evolved. It will be interesting to fully delineate the splicing pathway of the archaeal system through discovery of the ligase that must function to ligate the products of the endonuclease reaction. Perhaps we will again be afforded a glimpse into the RNA world that existed before the invention of protein enzymes.

REFERENCES

Abelson, J., Trotta, C. R., and Li, H. (1998). tRNA splicing. J. Biol. Chem. *273*, 12685-12688.

Apostol, B. L., Westaway, S. K., Abelson, J., and Greer, C. L. (1991). Deletion analysis of a multifunctional yeast tRNA ligase polypeptide: identification of essential and dispensable functional domains. J. Biol. Chem. *266*, 7445-7455.

Arn, E. A., and Abelson, J. (1998). RNA ligases: function, mechanism, and sequence conservation. In RNA structure and function (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 695-726.

Arn, E. A., and Abelson, J. N. (1996). The 2'-5' ligase from Escherichia coli: purification, cloning and genomic disruption. J. Biol. Chem. *271*, 31145-31153.

Baldi, M. I., Mattoccia, E., Bufardeci, E., Fabbri, S., and Tocchini-Valentini, G. P. (1992). Participation of the intron in the reaction catalyzed by the *Xenopus* tRNA splicing endonuclease. Science *255*, 1404-1408.

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.

Belfort, M., and Weiner, A. (1997). Another bridge between kingdoms:tRNA spling in Archaea and Eukaryotes. Cell *89*, 1003-1006.

Biniszkiewicz, D., Cesnaviciene, E., and Shub, D. A. (1994). Self-splicing group-I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. EMBO Journal *13*, 4629-4635.

Cavalier-Smith, T. (1991). Intron phylogeny: a new hypothesis. Trends Genet. *7*, 145-148.

Clark, M. W., and Abelson, J. (1987). The subnuclear localization of tRNA ligase in yeast. Journal Of Cell Biology *105*, 1515-1526.

Cox, J. S., and Walter, P. (1996). A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. Cell *86*, 391-404.

Culbertson, M. R., and Winey, M. (1989). Split tRNA genes and their products: a paradigm for the study of cell-function and evolution. Yeast *5*, 405-427.

Culver, G. M., McCraith, S. M., Consaul, S. A., Stanford, D. R., and Phizicky, E. M. (1997). A 2'-phosphotransferase implicated in tRNA splicing is essential in *Saccharomyces cerevisiae*. J. Biol. Chem. *272*, 13203-13210.

Culver, G. M., McCraith, S. M., Zillmann, M., Kierzek, R., Michaud, N., Lareau, R. D., Turner, D. H., and Phizicky, E. M. (1993). An NAD derivative produced during tRNA splicing: ADP-ribose 1"-2" cyclic phosphate. Science *261*, 206-208.

Daniels, C. J., Gupta, R., and Doolittle, W. F. (1985). Transcription and excision of a large intron in the $tRNA^{Trp}$ gene of an Archaebacterium, *Halobacterium volcanii*. J. Biol. Chem. *260*, 3132-3134.

Diener, J., and Moore, P. (1998). Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge-helix-bulge motif. Molecular Cell.

Fabbri, S., Fruscoloni, P., Bufardeci, E., Di Nicola Negri, E., Baldi, M. I., Gandini Attardi, D., Mattoccia, E., and Tocchini-Valentini, G. P. (1998). Conservation of substrate recognition mechanisms by tRNA splicing endonucleases. Science *280*, 284-286.

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

Gomes, I., and Gupta, R. (1997). RNA splicing ligase activity in the archaeon *Haloferax volcanii*. Biochemical and Biophysical Review *237*, 588-594.

Goodman, H. M., Olson, M. V., and Hall, B. D. (1977). Nucleotide sequences of a mutant eukaryotic gene: the yeast tyrosine inserting suppressor, SUP4-0. Proc. Natl. Acad. Sci. USA *74*, 5453-5457.

Green, C. J., Sohel, I., and Vold, B. S. (1990). The discovery of new intron-containing human tRNA genes using the polymerase chain reaction. J. Biol. Chem. *265*, 12139-12142.

Greer, C. L. (1986). Assembly of a tRNA splicing complex: evidence for concerted excision and joining steps in splicing *in vitro*. Mol. Cell. Biol. *6*, 635-644.

Greer, C. L., Soll, D., and Willis, I. (1987). Substrate recognition and identification of splice sites by the tRNA-splicing endonuclease and ligase from *Saccharomyces cerevisiae*. Molecular Cellular Biology *7*, 76-84.

Grosjean, H., Szweykowska-Kulinska, Z., Motorin, Y., Fasiolo, F., and Simos, G. (1997). Intron-dependent enzymatic formation of modified nucleosides in eukaryotic tRNAs: a review. Biochimie *79*, 293-302.

Ho, C. K., Rauhut, R., Vijayraghavan, U., and Abelson, J. (1990). Accumulation of pre-tRNA splicing 2/3 intermediates in a *Saccharomyces cerevisiae* mutant. Embo Journal *9*, 1245-1252.

Hopper, A. K., Banks, F., and Evangelidis, V. (1978). A yeast mutant which accumulates precursor tRNAs. Cell *14*, 211-219.

Johnson, J., Ogden, R., Johnson, P., Abelson, J., Dembeck, P., and Itakura, K. (1980). Transcription and processing of a yeast tRNA gene containing a modified intervening sequence. Proceedings of the National Academy of Sciences, USA *77*, 2564-2568.

Kjems, J., and Garrett, R. A. (1985). An intron in the 23S rRNA gene of the Archaebacterium *Desulfurococcus mobilis*. Nature *318*, 675-677.

Kjems, J., Jensen, J., Olesen, T., and Garrett, R. A. (1989). Comparison Of tRNA and rRNA intron splicing in the extreme thermophile and Archaebacterium *Desulfurococcus mobilis*. Canadian Journal Of Microbiology *35*, 210-214.

Kleman-Leyer, K., Armbruster, D. A., and Daniels, C. J. (1997). Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Cell *89*, 839-848.

Klenk, H., Clayton, R., Tomb, J., White, O., Nelson, K., Ketchum, K., Dodson, R., Gwinn, M., Hickey, E., Peterson, J., Richardson, D., Kerlavage, A., Graham, D., Kyrpides, N., Fleischmann, R., Quackenbush, J., Lee, N., Sutton, G., Gill, S., Kirkness, E., Dougherty, B., McKenney, K., Adams, M., Loftus, B., Peterson, S., Reich, C., McNeil, L., Badger, J., Glodek, A., Zhou, L., Overbeek, R., Gocayne, J., Weidman, J.,

McDonald, L., Utterback, T., Cotton, M., Spriggs, T., Artiach, P., Kaine, B., Sykes, S., Sadow, P., D'Andrea, K., Bowman, C., Fujii, C., Garland, S., Mason, T., Olsen, G., Fraser, C., Smith, H., Woese, C., and JC, V. (1998). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature *390*, 364.

Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A., and Abelson, J. N. (1978). Transcription and processing of intervening sequences in yeast tRNA genes. Cell *14*, 221-236.

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

Kuhsel, M. G., Strickland, R., and Palmer, J. D. (1990). An ancient group I intron shared by eubacteria and chloroplasts. Science *250*, 1570-1573.

Laski, F., Fire, A., RajBhandary, U., and Sharp, P. (1983). Characterization of tRNA precursor splicing in mammalian extracts. J. Biol. Chem. *258*, 11974-11980.

Lee, M. C., and Knapp, G. (1985). tRNA splicing in *Saccharomyces cerevisiae*: secondary and tertiary structures of the substrates. J. Biol. Chem. *260*, 3108-3115.

Li, H., Trotta, C. R., and Abelson, J. N. (1998). Crystal structure and evolution of a tRNA splicing enzyme. Science *280*, 279-284.

Lykke-Andersen, J., and Garrett, R. A. (1997). RNA-protein interactions of an archaeal homtetrameric splicing endoribonuclease with an exceptional evolutionary history. EMBO Journal *16*, 6290-6300.

McCraith, S. M., and Phizicky, E. M. (1991). An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. J. Biol. Chem. *266*, 11986-11992.

McCraith, S. M., and Phizicky, E. M. (1990). A highly specific phosphatase from *Saccharomyces cerevisiae* implicated in tRNA splicing. Mol. Cell. Biol. *10*, 1049-1055.

Nishikura, K., and De Robertis, E. (1981). RNA processing in microinjected *Xenopus* oocytes. Sequential addition of base modification in the spliced transfer RNA. J. Mol. Biol. *145*, 405-420.

Ogden, R. C., Lee, M. C., and Knapp, G. (1984). Transfer RNA splicing In Saccharomyces cerevisiae: defining the substrates. Nucleic Acids Res. *12*, 9367-9382.

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

Phizicky, E. M., Consaul, S. A., Nehrke, K. W., and Abelson, J. (1992). Yeast tRNA ligase mutants are nonviable and accumulate tRNA splicing intermediates. J. Biol. Chem. *267*, 4577-4582.

Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986). *Saccharomyces cerevisiae* tRNA ligase: purification of the protein and isolation of the structural gene. J. Biol. Chem. *261*, 2978-2986.

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

Reinhold-Hurek, B., and Shub, D. A. (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. Nature *357*, 173-176.

Reyes, V. M., and Abelson, J. (1988). Substrate recognition and splice site determination in yeast tRNA splicing. Cell *55*, 719-730.

Schneider, A., Perry-McNally, K., and Agabian, N. (1993). Splicing and 3'-processing of the tyrosine tRNA of *Trypanosoma brucei*. J. Biol. Chem. *268*, 21868-21874.

Schwartz, R., Greer, C., Gegenheimer, P., and Abelson, J. (1983). Enzymatic mechanism of an RNA ligase from wheat germ. J. Biol. Chem. *258*, 8374-8383.

Sidrauski, C., Cox, J. S., and Walter, P. (1996). tRNA ligase is required for regulated mRNA splicing in the unfolded protein response. Cell *86*, 405-413.

Sidrauski, C., and Walter, P. (1997). The transmembrane kinase Ire1p is a site-specific endonuclease that initiates mRNA splicing in the unfolded protein response. Cell *90*, 1031-1039.

Spinelli, S., Consaul, S., and Phizicky, E. (1997). A conditional lethal yeast phophotransferase (tpt1) mutant accumulates tRNAs with a 2'-phosphate and an undermodified base at the splice junction. RNA *3*, 1388-1400.

Stange, N., and Beier, H. (1986). A gene for the major cytoplasmic $tRNA^{Tyr}$ from *Nicotiana rustica* contains a 13 nucleotides long intron. Nucleic Acids Res. *14*, 8691.

Strobel, M., and Abelson, J. (1986). Effect of intron mutations on processing and function of Saccharomyces cerevisiae SUP53 tRNA in vitro and in vivo. Mol. Cell. Biol. *6*, 2663-2673.

Swerdlow, H., and Guthrie, C. (1984). Structure of intron-containing tRNA precursors: analysis of solution conformation using chemical and enzymatic probes. J. Biol. Chem. *259*, 5197-5207.

Thompson, J. E., and Raines, R. T. (1994). Value of general acid-base catalysis to ribonucleaseA. Journal of the American Chemical Society *116*, 5467-5468.

Thompson, L. D., Brandon, L. D., Nieuwlandt, D. T., and Daniels, C. J. (1989). Transfer RNA intron processing in the halophilic archaebacteria. Canadian Journal Of Microbiology *35*, 36-42.

Thompson, L. D., and Daniels, C. J. (1988). A $tRNA^{trp}$ intron endonuclease from Halobacterium volcanii: unique substrate recognition properties. J. Biol. Chem. *263*, 17951-17959.

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.

Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R., and Rutter, W. J. (1978). Structure of yeast phenylalanine-tRNA genes: an intervening DNA segment within the coding region for the tRNA. Proceedings of the National Academy of Sciences USA *75*, 190-194.

Walker, G., Uhlenbeck, O. C., Bedows, E., and Gumport, R. I. (1975). T4-induced RNA ligase joins single-stranded oligoribonucleotides. Proceedings of the National Academy of Sciences, USA *72*, 122-126.

Walsh, C. (1979). Enzymatic reaction mechanisms (San Francisco: W.H. Freeman and Company).

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

Willis, I., Hottinger, H., Pearson, D., Chisholm, V., Leupold, U., and Soll, D. (1984). Mutations affecting excision of the intron from a eukaryotic dimeric tRNA precursor. EMBO J. *3*, 1573-1580.

Winey, M., and Culbertson, M. R. (1988). Mutations affecting the tRNA splicing endonuclease activity of *Saccharomyces cerevisiae*. Genetics *118*, 49-63.

Xu, Q., Teplow, D., Lee, T. D., and Abelson, J. (1990). Domain structure in yeast tRNA ligase. Biochemistry *29*, 6132-6138.

Zillmann, M., Gorovsky, M. A., and Phizicky, E. M. (1991). Conserved mechanism of tRNA splicing in Eukaryotes. Mol. Cell. Biol. *11*, 5410-5416.

Zillmann, M., Gorovsky, M. A., and Phizicky, E. M. (1992). Hela cells contain a 2'-phosphate-specific phosphotransferase similar to a yeast enzyme implicated in tRNA splicing. J. Biol. Chem. *267*, 10289-10294.

Chapter II - The Yeast tRNA Splicing Endonuclease: A Two Active Site Tetrameric Enzyme Homologous to Archaeal tRNA Splicing Endonucleases

(Christopher R. Trotta, Feng Miao, Eric A. Arn, Scott W. Stevens, Calvin K. Ho, Reinhard Rauhut, and John N. Abelson. Published in Cell 89:849-858, 1997; Appendix II)

Abstract

The splicing of tRNA precursors is essential for the production of mature tRNA in organisms from all major phyla. In yeast, the tRNA splicing endonuclease is responsible for identification and cleavage of the splice sites in pre-tRNA. We have cloned the genes encoding all four protein subunits of endonuclease. Each gene is essential. Two subunits, Sen2p and Sen34p, contain a homologous domain of approximately 130 amino acids. This domain is found in the gene encoding the archaeal tRNA splicing endonuclease of *H. volcanii* and in other Archaea. Our results demonstrate that the eucaryal tRNA splicing endonuclease contains two functionally independent active sites for cleavage of the 5' and 3' splice sites, encoded by SEN2 and SEN34 respectively. The presence of endonuclease in Eucarya and Archaea suggests an ancient origin for the tRNA splicing reaction.

Introduction

Although RNA splicing is a required step for gene expression in all three taxonomic kingdoms, at least four different mechanisms of splicing have evolved. The most primitive mechanism, autocatalytic self splicing of Group I and Group II introns, involves two phosphotransfer reactions (Cech, 1993). Pre-mRNA splicing requires ATP hydrolysis and more than 100 proteins, but it proceeds via the same chemical mechanism

used by Group II introns suggesting that it is also RNA-catalyzed and that it shares a common origin with Group II intron splicing (Madhani and Guthrie, 1994; Moore, 1993). In contrast to these RNA-catalyzed reactions, splicing of eucaryotic tRNA introns occurs by a mechanism involving three proteinaceous enzymes (Abelson, 1991; Culbertson and Winey, 1989; Phizicky and Greer, 1993). In the archaeal kingdom tRNA splicing proceeds by a similar protein-based enzymatic mechanism, but this splicing is distinct by a number of functional criteria (Thompson et al., 1989). Splicing of all known prokaryotic tRNA introns occurs via the primitive autocatalytic mechanism (Biniszkiewicz et al., 1994; Reinhold-Hurek and Shub, 1992). So while tRNA splicing occurs in all three kingdoms, the mechanism for this process is not conserved. Major questions regarding the origin of tRNA introns and the relationship between the disparate enzymatic machineries required for their removal therefore remain. Do the different splicing mechanisms reflect independent origins or a single origin and later divergence? The results of this work together with those in the accompanying papers of Di Nicola Negri et al. (1997) and Kleman-Leyer et al. (1997) provide new insight into the origin of the tRNA spicing machinery.

Much of our knowledge of the mechanism of eucaryotic tRNA splicing comes from studies in *Saccharomyces cerevisiae.* The *S. cerevisiae* genome contains 272 tRNA genes of which 59, encoding ten different tRNAs, contain introns (CRT, unpublished data). These introns are 14 to 60 bases in length and all interrupt the anticodon loop immediately 3' to the anticodon (Ogden et al., 1984). Splicing of introns from precursor tRNA (pre-tRNA) is accomplished through the action of a site-specific endonuclease, an RNA ligase, and a phosphotransferase. The tRNA splicing endonuclease (referred to herein as "endonuclease") cleaves pre-tRNA at the 5' and 3' splice sites to release the

intron. The products of the endonuclease reaction are an intron and two tRNA half-molecules bearing 2',3' cyclic phosphate and 5'-OH termini (Peebles et al., 1983). tRNA ligase covalently joins the half-molecules through a complex series of ATP- and GTP-dependent reactions (Belford et al., 1993; Westaway et al., 1993; Westaway et al., 1988). After ligation, a 2' phosphate remains at the splice junction and is removed by a specific phosphotransferase (Culver et al., 1993; McCraith and Phizicky, 1991).

Figure II-1A summarizes the conserved features of the ten yeast pre-tRNAs containing introns (as depicted by Lee and Knapp, 1984). There are no conserved sequences at the splice sites, but the intron is invariably located at the same site in the gene, placing the splice sites an invariant distance from the constant structural features of the tRNA body. This conserved placement implies that recognition of the splice sites may be accomplished via interaction with the mature tRNA domain. The hypothesis that the endonuclease measures the distance from the mature domain to the splice sites to correctly place cleavages (Figure II-1B) was confirmed by experiments where insertions in the anticodon stem of pre-tRNA which changed the distance between the mature domain and the splice sites resulted in a predictable shift of the cleavage sites (Reyes and Abelson, 1988). Further evidence for a strict measuring mechanism was provided by the fact that an artificial tRNA intron composed almost entirely of uridine residues was spliced correctly in the yeast system (Reyes and Abelson, 1988).

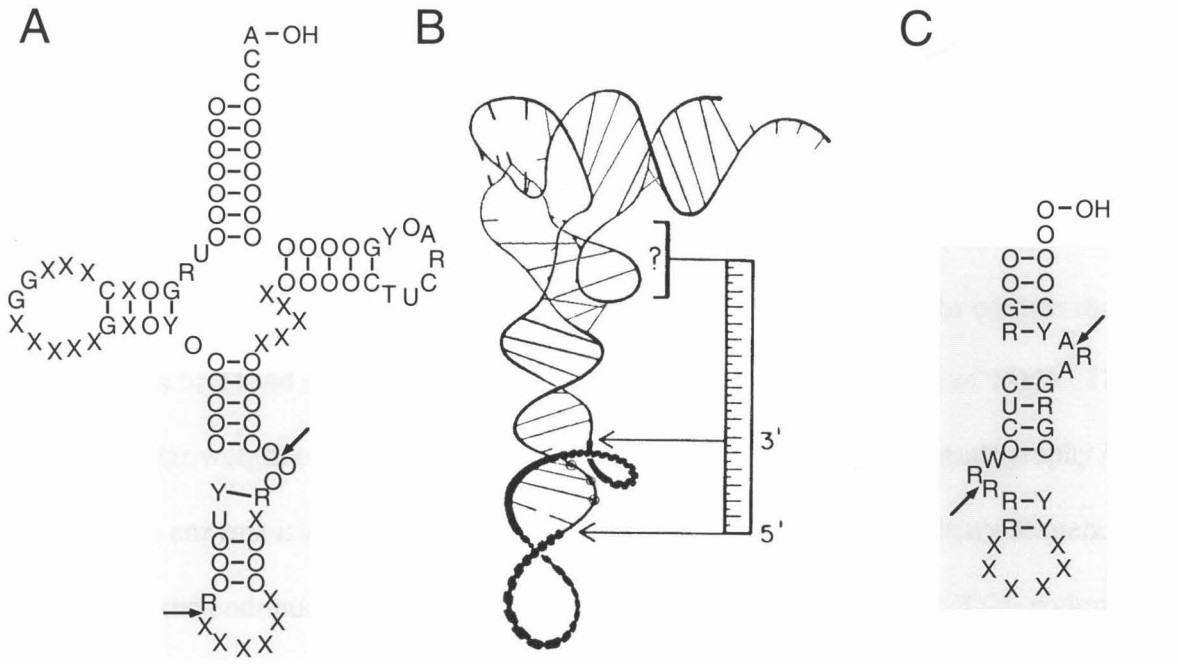
The intron is not, however, completely passive in the recognition of the splice sites. The *Xenopus* tRNA endonuclease has been shown to recognize a crucial structural element in tRNA introns (Baldi et al., 1992). Yeast tRNA introns contain a conserved purine residue three nucleotides upstream of the 3' splice site. This base must be able to pair with a pyrimidine at position 32 in the anticodon loop in order for the intron to be

FIGURE II-1. Conserved Sequence and Structure Elements of Yeast Pre-tRNA and Archaeal Minimal Pre-tRNA

A. A composite, modified from Lee and Knapp (1985), of yeast pre-tRNA sequences (Ogden et al., 1984). X = unconserved base in a region of variable length or secondary structure, O = an unconserved base in a region of conserved secondary structure. G, C and A indicate conserved bases conserved among pre-tRNAs. Y and R = conserved pyrimidines and purines, respectively. The conserved intron-anticodon tertiary interaction is depicted by a line between the conserved R and Y bases.

B. Taken from Reyes and Abelson (1988). Presumed tertiary structure of pre-tRNA^{Phe} (from Lee and Knapp, 1984). Endonuclease is proposed to interact with the mature domain and measure the distance to the two splice sites.

C. A composite of archaeal pre-tRNA sequences, modified from Thompson and Daniels (1988). Position designations are as in panel A. This represents the minimal substrate for cleavage by the archaeal endonuclease.



Yeast Splicing Substrate Consensus

Yeast Endonuclease Ruler Model

Archaeal Splicing Substrate Consensus

Figure II-1

recognized by either yeast or *Xenopus* endonuclease (Figure II-1A) (Baldi et al., 1992; Di Nicola Negri et al., 1997). This data suggested there could be different requirements for the recognition of the 5' and 3' splice sites. More recent evidence in this paper and the accompanying paper by Di Nicola Negri et al. support the hypothesis that the endonuclease contains two independent active sites with different substrate recognition mechanisms.

The mechanism of intron removal in tRNA splicing has also been investigated through purification of the endonuclease enzyme from yeast (Rauhut et al., 1990). Endonuclease behaves as a membrane-bound protein present in a low abundance of ~100 molecules per (Rauhut et al., 1990). The purified enzyme was seen to contain three subunits of 54, 44 and 34kD and, as we report here, a fourth subunit of 15kD. The molecular weight of the holoenzyme as judged by gel exclusion chromatography implies that the enzyme is an $\alpha\beta\gamma\delta$ tetramer. In this paper we report the primary sequences of all four of the endonuclease subunits. Analysis of the peptide sequence of the endonuclease subunits presented here sheds light on the architecture of the enzyme and its possible interactions with other components of the nucleus.

The pre-tRNAs of Archaea are similar to those of Eucarya in that they contain introns which must be removed by protein enzymes in *trans*. Superficially, these introns are similar to yeast tRNA introns in that they are often (though not always) located in the anticodon loop. Removal of tRNA introns by either eucaryal or archaeal endonucleases yield 2',3' cyclic PO_4 and 5'-OH terminii. Studies of removal of tRNA introns in Archaea have, however, shown notable differences between the yeast and archaeal tRNA endonucleases. Yeast pre-tRNAs are not substrates for the *Haloferax volcanii* endonuclease (Palmer et al., 1994), and the mature domain of a *Halobacterium volcanii*

pre-tRNA is not required for correct excision of the intron (Thompson and Daniels, 1988). Rather, archaeal tRNA splice sites must be located in bulged loops separated by a helix of four base-pairs (Figure II-1C). This evidence would seem to suggest a different origin of the splicing process in these two kingdoms. The results presented here, however, together with those in the accompanying paper by Kleman-Leyer et al. (1998) establish that two of the yeast endonuclease subunits define a gene family that includes the single homodimeric 40kDa *Haloferax volcanii* endonuclease as well as related genes from other Archaea. We discuss the evolutionary and mechanistic implications of these findings.

Results

The gene for the 44kD subunit, *SEN2*.

Our primary goal in the characterization of the yeast tRNA splicing endonuclease has been to clone the genes for each of the subunits of this enzyme. We previously reported the isolation and characterization of the cold sensitive *sen2-3* mutation, an allele of the *SEN2* gene found by Winey and Culbertson (1988) in their search for yeast tRNA mutants defective in tRNA splicing *in vitro*. At the non-permissive temperature *sen2-3* accumulates "2/3" pre-tRNA molecules cleaved at the 3' splice site only (Ho et al., 1990) strongly suggesting that Sen2p is a component of endonuclease. In this paper we report the cloning of the *SEN2* gene by complementation of the *sen2-3* mutant and the determination of its nucleotide sequence (Genbank accession #M32336). Analysis of the amino acid sequence of Sen2p reveals the presence of a probable transmembrane sequence between residues 225 and 243 (Figure II-5B). During purification endonuclease behaves as a membrane-bound enzyme (Rauhut et al., 1990). None of the other subunits (see below) contain a convincing transmembrane sequence, so it is possible that Sen2p is the

subunit which anchors the endonuclease complex to the nuclear membrane. Sen2p also contains a domain predicted to form a canonical anti-parallel coiled-coil structure between residues 122 and 184 (Figure II-5E). Sequencing of the mutant *sen2-3* allele reveals a single point mutation causing the change of glycine 292 to glutamate (Figure II-3B).

A new purification strategy for the tRNA splicing endonuclease

We had previously developed a chromatographic procedure for the purification of endonuclease which resulted in homogenous enzyme (Rauhut et al., 1990), and provided sufficient material for the characterization of two of the endonuclease subunits. However, this procedure was extremely time consuming and gave a low yield of enzyme. We have developed a highly efficient affinity purification procedure for large-scale preparation of the enzyme based on the cloned *SEN2* gene.

The *SEN2* gene was modified by addition of an amino-terminal peptide tag consisting of eight histidine residues followed by the 4 residue FLAG epitope. This gene, cloned in a yeast 2μ vector, complemented the haplo-lethal phenotype of a *SEN2* deletion, proving that the affinity tag did not affect the activity of Sen2p *in vivo*. The complemented *sen2* deletion strain, CRT44, was used as a source of tagged enzyme for purification by a procedure utilizing standard chromatographic methods and two specific affinity steps, as diagrammed in Figure II-2A. This rapid procedure resulted in an efficient (10-20%) yield of 5 μg of pure enzyme per 150 grams of cells.

A fourth subunit in the tRNA splicing endonuclease

Pure endonuclease preparations had previously been characterized by SDS-PAGE and silver staining, which detected three subunits with apparent molecular masses of

FIGURE II-2. Affinity Purification of Endonuclease from Yeast

A. Vector YpH8/FLAG SEN2 constructed as per Experimental Procedures. HIS8 = eight histidine tag, FLAG = FLAG peptide epitope tag. This vector was used to complement a *SEN2* chromosomal deletion.

B. Purification procedure for isolation of affinity tagged endonuclease from yeast strain CRT44 (see Experimental Procedures for details).

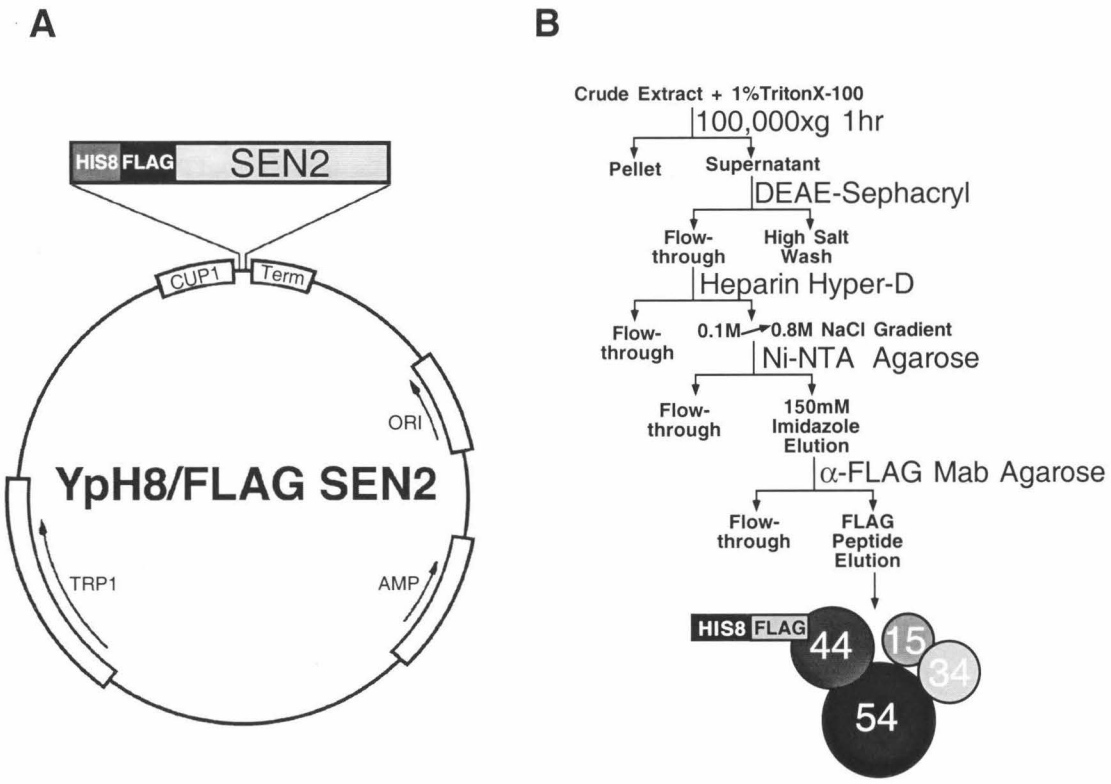

A
HIS8 FLAG
SEN2
CUP1
Term
ORI
YpH8/FLAG SEN2
TRP1
AMP
B
Crude Extract + 1%TritonX-100
100,000xg 1hr
Pellet
Supernatant
DEAE-Sephacryl
Flow-through
High Salt Wash
Heparin Hyper-D
Flow-through
0.1M
0.8M NaCl Gradient
Ni-NTA Agarose
Flow-through
150mM Imidazole Elution
α-FLAG Mab Agarose
Flow-through
FLAG Peptide Elution
HIS8 FLAG
44
15
34
54

Figure II-2

51kD, 42kD and 31kD (Rauhut et al., 1990). To compare endonuclease purified by the affinity method to that obtained in our original purification, we employed a high-resolution gradient gel procedure. This procedure revealed that both preparations contain proteins of 54, 44 and 34kD (cf. lanes A and R Figure II-6A), although the 44kD band in lane A has a lower mobility due to its affinity tag (the molecular weights of the proteins are designated here as determined by comparison with markers in gradient gel electrophoresis). Both preparations contained, in addition, a fourth band with an apparent molecular weight of 15kD. This band would have run with the dye front in our original procedure. The fact that all four proteins are present in stoichiometric amounts in endonuclease purified by different purification procedures and that all four proteins cosediment in glycerol density gradients (data not shown) strongly suggests that all four polypeptides are subunits of the enzyme. The molecular mass of the enzyme as measured by gel filtration chromatography in our earlier study was 140kD (Rauhut et al., 1990). With the fourth subunit, the predicted mass of a 1:1:1:1 endonuclease complex is 146kD, very close to the observed molecular mass. It is therefore likely that the subunit constitution of this enzyme is $\alpha\beta\gamma\delta$.

The genes for the 34kD, 54kD, and 15kD subunits

A 50 microgram aliquot of purified endonuclease, produced by the original Rauhut procedure, was fractionated by SDS-PAGE, and individual bands were excised from the gel. These polypeptides were digested with lysyl-protease and a peptide resulting from digestion of the 34kD subunit yielded the sequence FIAYPGDPLRFX(R?)XLTIQ. A search of the yeast genome database revealed that this peptide is uniquely found in the open reading frame (ORF) YAR008w on *S. cerevisiae* Chromosome I (Figure II-3C).

This ORF potentially encodes a protein with a predicted molecular weight of 34kD. We have renamed this gene *SEN34*. Further evidence described below confirms that this protein is the 34kD subunit of the endonuclease.

Five micrograms of endonuclease purified by the affinity method described above were also fractionated by SDS-PAGE, the 54kD band was digested with lysyl-protease and one unambiguous peptide sequence was obtained, NDDLQHFPTYK, which matches the sequence encoded by ORF YPL083c of *S. cerevisiae* Chromosome XVI (Figure II-3A). This 467 amino acid protein has a predicted molecular weight of 54kD. We have renamed this gene *SEN54*. Sen54p is overall a very basic protein (pI=9.5) but it contains several highly acidic regions, e.g., the sequence EEEEEEEEELQQD (residues 19-31) and EFLSEIEDDDDWE (residues 346-358). The significance of these extreme disparities in charge remains to be investigated. Further evidence described below suggests that this protein is a subunit of endonuclease.

Using a variation of the affinity purification method (see Experimental Procedures), several micrograms of pure 15kD subunit were prepared from SDS-PAGE-fractionated endonuclease. This material was analyzed by trypsin digestion followed by LC-MS/MS analysis and three peptides were sequenced. Two of these were identical except for a possible acetylation modification of the methionine in the peptide sequence MATTDIISLVK, suggesting that this peptide is the amino terminus of the protein (K. Swiderek, pers. comm). The two unique peptides are found in a protein encoded by ORF YMR059w on Chromosome XIII (Figure II-3D). This ORF potentially encodes a protein of 17kD. However, if the methionine codon at position 20 in this ORF is the translation start site for this protein, as suggested by the peptide sequencing, the resulting protein would be predicted to have a length of 124 amino acids with a molecular weight of

FIGURE II-3. Primary Amino Acid Sequence of the Four Subunits of the Yeast tRNA Splicing Endonuclease

A. Sen54p

B. Sen2p

C. Sen34p

D. Sen15p.

Residues which are underlined represent peptides which were directly sequenced. Residues shown in gray boxes represent either bi-partite or basic nuclear localization sequences as described by Robbins et al., (1991) or Kalderon et al., (1984), respectively. Residues shown in a black box represent a predicted transmembrane domain (see figure II-5 for details). Residues shown in open boxes represent potential nuclear export signals as described by Taylor and co-workers (1995) and Lührmann and co-workers (1995). In panel B residue 292, the underlined and bold G represents the mutation in the *sen2-3* mutant (Ho et al., 1990). In panel D the * represent the presumed initiator methionine, and the italicized residues represent the presumed mitochondrial pre-sequence (see text for details).

A Sen54p

```
MQFAGKKTDQVTTSNPGFEEEEEEEEELQQDWSQLASLVSKNAALSLPKR  50
GEKDYEPDGTNLQDLLLYNASKAMFDTISDSIRGTTVKSEVRGYYVPHKH 100
QAVLLKPKGSFMQTMGRADSTGELWLDFHEFVYLAERGTILPYYRLEAGS 150
NKSSKHETEILLSMEDLYSLFSSQQEMDQYFVFAHLKRLGFILKPSNQEA 200
AVKTSFFPLKKQRSNLQAITWRLLSLFKIQELSLFSGFFYSKWNFFFRKY 250
TTSPQLYQGLNRLVRSVAVPKNKKELLDAQSDREFQKVKDIPLTFKVWKP 300
HSNFKKRDPGLPDFQVFVYNKNDDLQHFPTYKELRSMFSSLDYKFEFLSE 350
IEDDDDWETNSYVEDIPRKEYIHKRSAKSQTEKSESSMKASFQKKTAQSS 400
TKKKRKAYPPHIQQNRRLKTGYRSFIIAIMDNGLISFVKMSEADFGSESV 450
WYTPNTQKKVDQRWKKH                                   467
```

B Sen2p

```
MSKGRVNQKRYKYPLPIHPVDDLPELILHNPLSWLYWAYRYYKSTNALND  50
KVHVDFIGDTTLHITVQDDKQMLYLWNNGFFGTGQFSRSEPTWKARTEAR 100
LGLNDTPLHNRGGTKSNTETEMTLEKVTQQRRLQRLEFKKERAKLERELL 150
ELRKKGGHIDEENILLEKQRESLRKFKLKQTEDVGIVAQQQDISESNLRD 200
EDNNLLDENGDLLPLESLELMPVEAMFLTFALPVLDISPACLAGKLFQFD 250
AKYKDIHSFVRSYVIYHHYRSHGWCVRSGIKFGCDYLLYKRGPPFQHAEF 300
CVMGLDHDVSKDYTWYSSIARVVGGAKKTFVLCYVERLISEQEAIALWKS 350
NNFTKLFNSFQVGEVLYKRWVPGRNRD                         377
```

C Sen34p

```
MAEEGGTRIAINIYAKRTAKGEEVFMPPLVFDIDHIKLLRKWGICGVLSG  50
TLPTAAQQNVFLSVPLRLMLEDVLWLHLNNLADVKLIRQEGDEIMEGITL 100
ERGAKLSKIVNDRLNKSFEYQRKFKKDEHIAKLKKIGRINDKTTAEELQR 150
LDKSSNNDQLIESSLFIDIANTSMILRDIRSDSDSLSRDDISDLLFKQYR 200
QAGKMQTYFLYKALRDQGYVLSPGGRFGGKFIAYPGDPLRFHSHLTIQDA 250
IDYHNEPIDLISMISGARLGTTVKKLWVIGGVAEETKETHFFSIEWAGFG 300
```

D Sen15p

```
MKDIINCRREITSRRLRKIRMATTDIISLVKNNLLYFQMWTEVEILQDDL  50
SWKGNSLRLLRGRPPHKLSNDVDTEHENSLSSPRPLEFILPINMSQYKEN 100
FLTLECLSQTFTHLCSPSTERILLAIINDDGTIVYYFVYKGVRKPKRN    144
```

Figure II-3

14.8kD, exactly the size determined by SDS-PAGE. We have renamed this gene *SEN15*. Further evidence described below suggests that this protein is a subunit of endonuclease.

Chromosomal disruptions of each of the subunit genes are lethal in haploid yeast cells, indicating that all of their gene products are essential for vegetative growth, and employing an essential function for each in the tRNA splicing process (Figure II-4A-D).

Analysis of the tRNA splicing endonuclease subunit sequences

During all purification procedures endonuclease behaves as a membrane-bound protein. We therefore searched for predicted membrane-spanning sequences in each of subunits. Figure II-5 A-D shows hydropathy plots for each of the four subunits. The most probable transmembrane sequence is found in Sen2p, the 44kD subunit, between residues 225 and 243 (see Figure II-3B). Less convincing but still potential transmembrane sequences are found in Sen54p (residues 226 to 247) and Sen34p (residues 42 to 60).

Sen2p contains a domain predicted to form a canonical anti-parallel coiled-coil structure (see Figure II-5E). This structure can form distinct protrusions from the surface of a protein, as seen in the *E. coli* seryl-tRNA synthetase (Cusack et al., 1996), and may mediate protein-protein or RNA-protein interactions.

All four subunits contain probable basic and bipartite nuclear localization sequences of the type described by Kalderon et al., (1984) and Robbins et al., (1991) (Figure II-3 A-D shaded boxes). Sen2p and Sen54p also contain potential nuclear export signals (NES) similar to those originally described by Taylor and co-workers for export of the PKI protein and Lührmann and co-workers for export of the HIV Rev-1 protein (Fischer et al., 1995; Wen et al., 1995) (Figure II-3A,B, open boxes). When fused to a

FIGURE II-4. All Four Subunits of Endonuclease are Essential

(**A-D**) Diploid strains containing single locus deletion of the ORFs of the various endonuclease subunits were sporulated and tetrads were dissected on YPD. Image was taken after 4 days growth at 30°C. All spores were examined for germination and only those in which all four spores germinated are depicted.

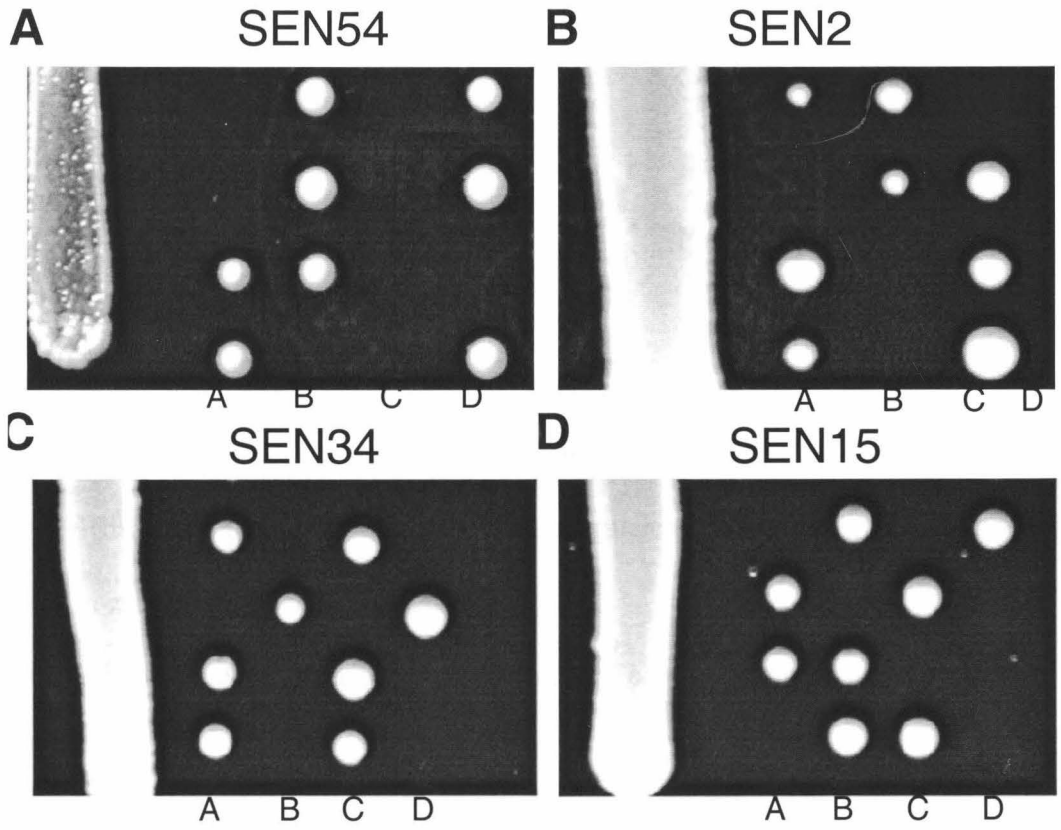


Figure II-4

FIGURE II-5. Sequence Analysis of the Subunits of Endouclease

(**A-D**) Hydropathy plots of each subunit of endonuclease using the Kyte and Doolittle algorithm and a window of 19 amino acids.

(**E**) Coiled coil prediction for Sen2p utilizing the Coils program (Lupas et al., 1991).

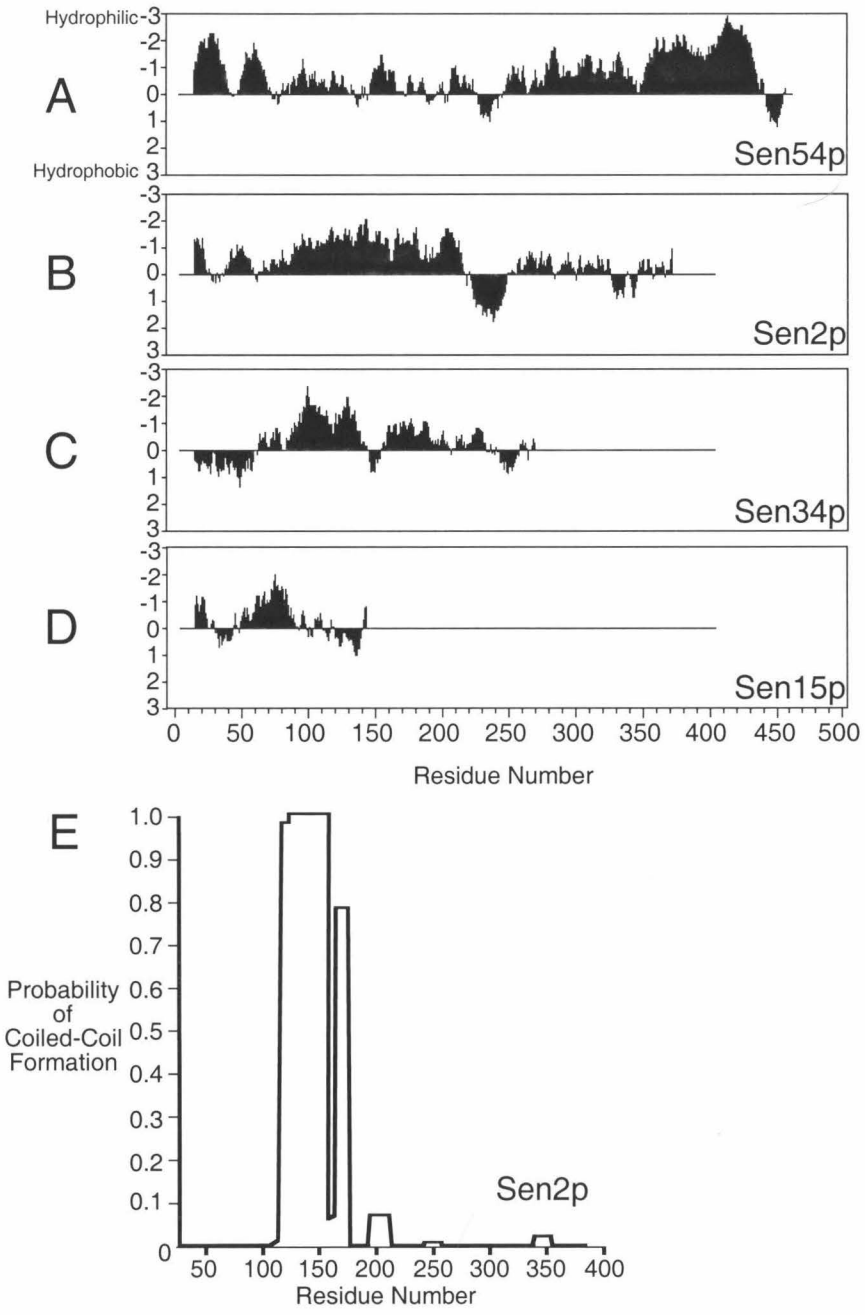

Hydrophilic
Hydrophobic
A
B
C
D
E
Sen54p
Sen2p
Sen34p
Sen15p
-3
-2
-1
0
1
2
3
0 50 100 150 200 250 300 350 400 450 500
Residue Number
Probability of Coiled-Coil Formation
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
50 100 150 200 250 300 350 400
Residue Number

Figure II-5

heterologous protein, NES's direct export of a protein from the nucleus to the cytoplasm (Fischer et al., 1995; Wen et al., 1995). *SEN15* encodes a possible mitochondrial import presequence in the first 20 amino acids (Figure II-3D italicized residues) (Attardi and Schatz, 1988; Kiebler et al., 1993). This feature is surprising and suggests the possibility that Sen15p may also localize to the mitochondria. All four genes are characterized by a low codon usage bias, in accordance with the very low abundance of this enzyme (Rauhut et al., 1990).

Antibodies to the subunits of endonuclease

To unambiguously determine that each of these genes encodes a subunit of endonuclease, we prepared polyclonal antisera specific for each of the four gene products expressed in *E. coli*. A Western blot experiment was performed in which purified endonuclease was fractionated by SDS-PAGE and individually probed with antisera specific for each of the four gene products. As shown in Figure II-6, probing of endonuclease purified by the affinity procedure (lane A), and endonuclease purified by the Rauhut procedure (lane R), reveals that each of the antisera react specifically with the appropriate polypeptide. The antibody to Sen15p was produced to a histidine/FLAG tagged antigen and as can be seen, the polyclonal antisera also contains antibodies to the affinity tag present on Sen2p and an apparent degradation product of Sen2p (compare 2C α-Sen15, lane A to 2C α-Sen2, lane A). Since the two preparations of pure endonuclease were prepared by different purification protocols, these results strongly support the conclusion that each of these four genes encodes a subunit of endonuclease.

FIGURE II-6. Western Blot of Pure Endonuclease Utilizing Antibodies to Endonuclease Subunits

A. Silver stained gel of: Lane 1, 5μg of crude yeast extract; lane 2, ~0.5μg of affinity tagged endonuclease purified as in Figure II-2; lane 3, ~0.5μg of endonuclease purified by Rauhut et al.(1990); MW represents molecular weight standards (Novex) and are labeled to the left in kiloDaltons.

Panels **B**, **C**, **D**, and **E** represent duplicate blots of gel from panel A probed with 1:500 dilution of antibodies to Sen54p, Sen2p, Sen34p, and Sen15p, respectively.

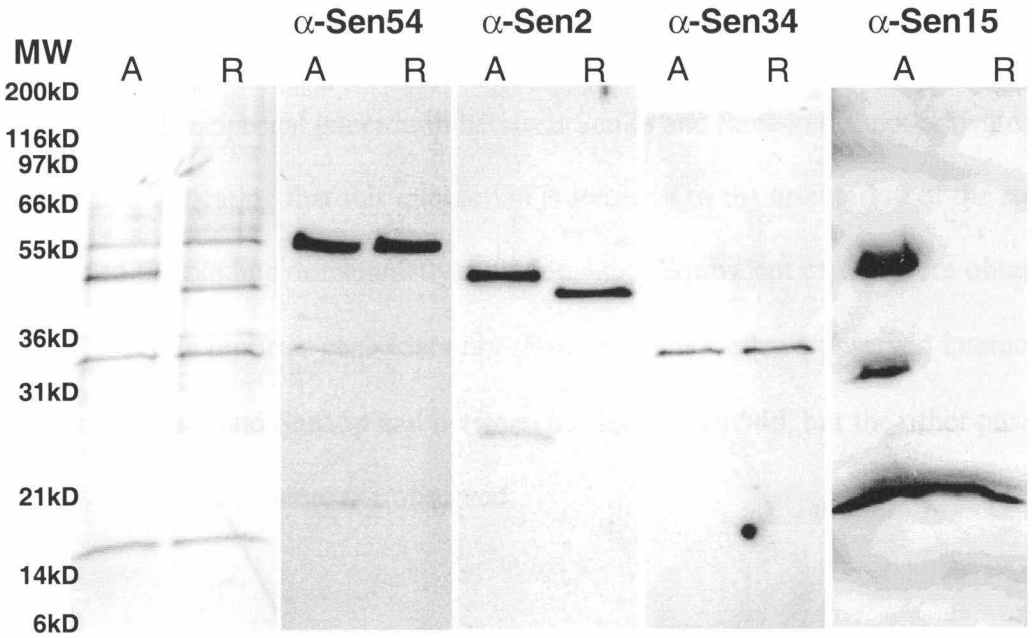

MW
A
R
α-Sen54
A
R
α-Sen2
A
R
α-Sen34
A
R
α-Sen15
A
R
200kD
116kD
97kD
66kD
55kD
36kD
31kD
21kD
14kD
6kD

Figure II-6

Two-hybrid analysis of the four subunits of endonuclease

To begin to understand the functional interactions of the endonuclease subunits, we have tested the four subunit genes in a two-hybrid matrix experiment. Each of the four genes was translationally linked to either a DNA-binding or transcriptional activation domain for analysis in the Fields two-hybrid system (Bartel and Fields, 1995; Fields and Song, 1989). Figure II-7 shows the results of combinatorial expression of subunit gene pairs using rescue of histidine auxotrophy as an assay of interaction. Two interactions were detected *in vivo*: interaction between Sen2p and Sen54p, and between Sen34p and Sen15p. The reciprocal interaction between Sen2p and Sen54p did not activate expression, indicating that this interaction is sensitive to the orientation of the subunits relative to the binding domain/activation domains. Equivalent results were obtained with a β-galactosidase reporter gene (data not shown). Thus we have observed interactions between Sen34p and Sen15p and between Sen2p and Sen54p, but the other possible pairwise interactions were not observed.

Sen2p and Sen34p are members of a gene family of endonucleases

A search of the Genbank database (ver. 97) for sequences similar to the four endonuclease subunit genes did not reveal any potential homologs. However, similarity was detected between the Sen2 and Sen34 proteins themselves. The similarity occurs in a region of approximately 130 amino acids (see below). A sequence profile of the region of similarity between these proteins was used to search the translated Genbank database and was found to match one sequence, a 205 amino acid ORF contained in a presumed intron of an HMG-like gene of *Zea mays* (Yanagisawa and Izui, 1993). C. Daniels and colleagues have determined the nucleotide sequence of the gene for the tRNA splicing

FIGURE II-7. Two-Hybrid Matrix Analysis of Endonuclease Subunits

Activation of histidine reporter gene by various hybrid combinations of Gal4 binding domain and Gal4 activation domain to the subunits of endonuclease. Reporter strain was transformed and then duplicate colonies from each were picked to a new selective plate. This plate was then replica plated to plates lacking histidine and containing 5mM 3-aminotriazole (to suppress background growth) to determine activation of the reporter gene.

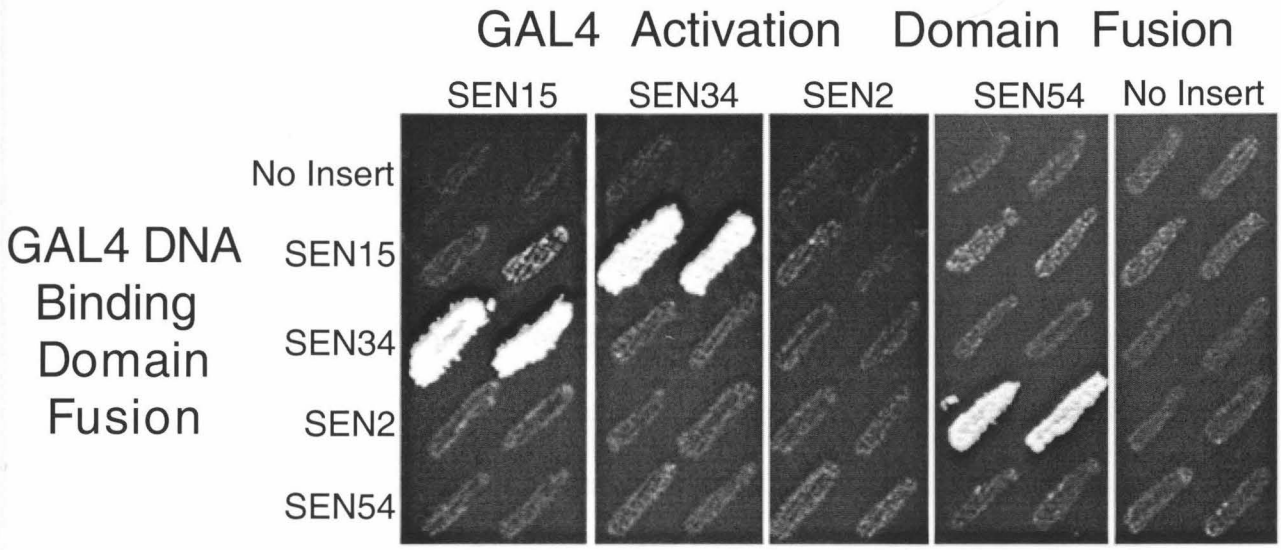

GAL4 Activation Domain Fusion
SEN15
SEN34
SEN2
SEN54
No Insert
GAL4 DNA Binding Domain Fusion
No Insert
SEN15
SEN34
SEN2
SEN54

Figure II-7

FIGURE II-8. Sequence Alignment Between the Archaeal Endonucleases and Two Subunits of the Yeast Endonuclease Identifies a Family of tRNA Splicing Endonucleases Sequence alignment between endonuclease identified by Daniels and co-workers from *Haloferax volcanii* and *Methanobacterium thermoautotrophicum*, endonuclease from *Methanococcus jannaschii*, putative endonuclease from *Pyrobaculum aerophilum*, putative endonuclease from *Zea mays* and two subunits, Sen34p and Sen2p, of the yeast tRNA splicing endonuclease. Sequence alignment was performed using the GCG program Pileup.

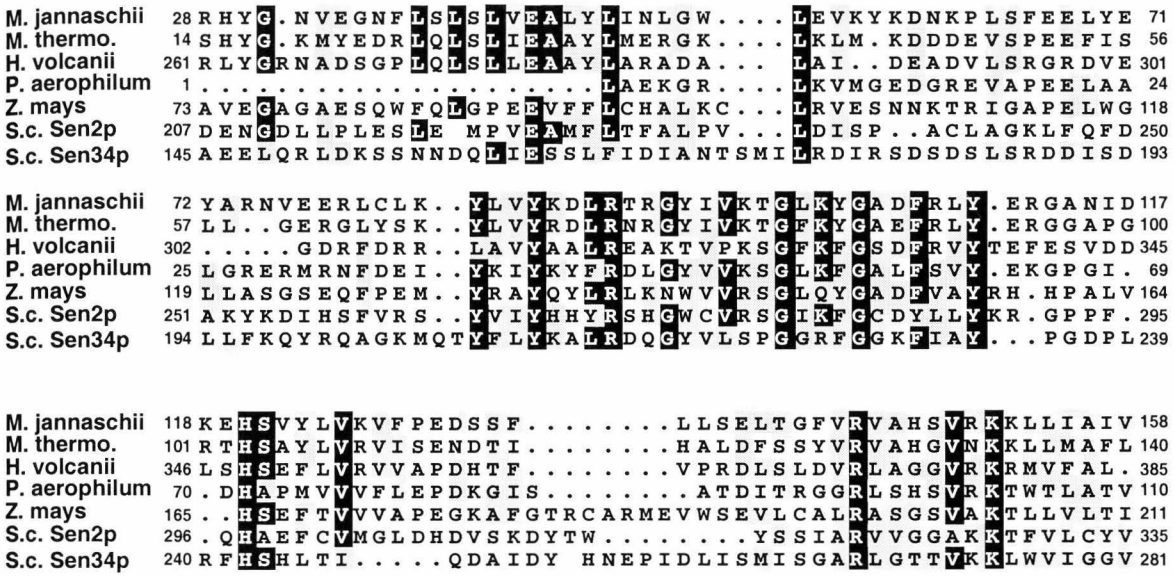

M. jannaschii 28 RHYG.NVEGNFLSLSLVEALYLINLGW....LEVKYKDNKPLSFEELYE 71
M. thermo. 14 SHYG.KMYEDRLQLSLIEAAYLMERGK....LKLM.KDDDEVSPEEFIS 56
H. volcanii 261 RLYGRNADSGPLQLSLLEAAYLARADA... LAI..DEADVLSRGRDVE 301
P. aerophilum 1 .....LAEKGR....LKVMGEDGREVAPEELAA 24
Z. mays 73 AVEGAGAESQWFQLGPEEVFFLCHALKC...LRVESNNKTRIGAPELWG 118
S.c. Sen2p 207 DENGDLLPLESLE MPVEAMFLTFALPV...LDISP..ACLAGKLFQFD 250
S.c. Sen34p 145 AEELQRLDKSSNNDQLIESSLFIDIANTSMILRDIRSDSDSLSRDDISD 193
M. jannaschii 72 YARNVEERLCLK..YLVYKDLRTRGYIVKTGLKYGADFRLY.ERGANID 117
M. thermo. 57 LL..GERGLYSK..YLVYRDLRNRGYIVKTGFKYGAEFRLY.ERGGAPG 100
H. volcanii 302 .....GDRFDRR..LAVYAALREAKTVPKSGFKFGSDFRVYTEFESVDD 345
P. aerophilum 25 LGRERMRNFDEI..YKIYKYFRDLGYVVKSGLKFGALFSVY.EKGPGI. 69
Z. mays 119 LLASGSEQFPEM..YRAYQYLRLKNWVVRSGLQYGADFVAYRH.HPALV 164
S.c. Sen2p 251 AKYKDIHSFVRS..YVIYHHYRSHGWCVRSGIKFGCDYLLYKR.GPPF. 295
S.c. Sen34p 194 LLFKQYRQAGKMQTYFLYKALRDQGYVLSPGGRFGGKFIAY...PGDPL 239
M. jannaschii 118 KEHSVYLVKVFPEDSSF.....LLSELTGFVRVAHSVRKKLLIAIV 158
M. thermo. 101 RTHSAYLVRVISENDTI.....HALDFSSYVRVAHGVNKKLLMAFL 140
H. volcanii 346 LSHSEFLVRVVAPDHTF.....VPRDLSLDVRLAGGVRKRMVFAL. 385
P. aerophilum 70 .DHAPMVVVFLEPDKGIS.....ATDITRGGRLSHSVRKTWTLATV 110
Z. mays 165 ..HSEFTVVVAPEGKAFGTRCARMEVWSEVLCALRASGSVAKTLLVLTI 211
S.c. Sen2p 296 .QHAEFCVMGLDHDVSKDYTW.....YSSIARVVGGAKKTFVLCYV 335
S.c. Sen34p 240 RFHSHLTI.....QDAIDY HNEPIDLISMISGARLGTTVKKLWVIGGV 281

Figure II-8

endonuclease of the Archaea *Haloferax volcanii* and *Methanobacterium thermoautotrophicum* (Kleman-Leyer et al., 1997). Homologs of this gene occur in other Archaeon, *Pyrobaculum aerophilum* and *Methanococcus jannaschii* (Bult et al., 1996). In order to determine if the *M. jannaschii* homolog is a functional homolog of the *H. volcanii* endonuclease, the gene product was expressed in *E. coli*, purified and shown to cleave intron containing archaeal pre-tRNAs (CRT, unpublished). Interestingly, the endonucleases of the Archaea align with the yeast endonuclease subunits, Sen2p and Sen34p, and the ORF of *Z. mays* (Figure II-8). In all, with sequences for seven members of this gene family available, a convincing homologous core domain of ~50 amino acids is now apparent with lesser regions of homology extending over ~130 amino acids.

Two active sites in endonuclease

The sequence similarity between Sen2p and Sen34p suggests that these proteins have a similar function. It seems that this function is very likely to be the catalysis of tRNA cleavage, since these proteins belong to a family which includes the homodimeric tRNA splicing endonuclease of *H. volcanii*. Interestingly, members of the endonuclease family found in the genomes of *M. thermoautotrophicum* and *M. jannaschii* are both small proteins made up entirely of the sequence domain found in the yeast and *H. volcanii* endonucleases. This further supports the notion that the sequence alignment reveals the catalytic core of the endonuclease enzyme. Previous results have indicated that Sen2p carries the active site for 5' splice site cleavage. Ho et al., 1990, showed that the *sen2-3* mutation results in endoncuclease that is defective in cleavage only at the 5' splice site of tRNA precursors *in vivo* (Ho et al., 1990). As reported here, the glycine to glutamate change in this allele is located at position 292, within the domain conserved between the

eucaryal and archaeal endonucleases. This residue itself is not conserved in the endonuclease family, as perhaps expected for a conditional lethal mutation. Given that Sen34p is homologous to Sen2p and to the archaeal endonucleases we predicted that the Sen34p contains the active site for 3' splice site cleavage. This prediction led to the model shown in Figure II-9 in which the 5' splice site is cleaved by Sen2p and the 3' splice site by Sen34p.

This model clearly predicts that endonuclease containing a mutant Sen34p subunit should fail to cleave the 3' splice site of pre-tRNA while retaining full activity in cleavage of the 5' splice site. To test this prediction we changed the conserved histidine at position 242 of *SEN34* to an alanine residue (H242A). Extrachromosomal expression of a HIS/FLAG tagged mutant Sen34p H242A in wild type haploid cells had a dominant deleterious effect, increasing the doubling time from 2.5 hours to 6 hours in minimal media. Extrachromosomal expression of wild type Sen34p in the same strain had little or no effect on growth. Endonuclease was purified from each of these strains by the affinity purification method described in the Experimental procedures to select only endonuclease complexes containing the tagged Sen34p subunit. tRNA splicing assays were performed with endonuclease containing wild type Sen34p and Sen34p H242A. Endonuclease containing affinity tagged wild type Sen34p cleaves pre-tRNA^{Phe} normally indicating that the affinity tagged enzyme functions as wild type endonuclease (Figure II-10). At the low enzyme concentrations used in this experiment, a slight accumulation of a 2/3 molecule consisting of the intron joined to the 3' exon, derived from cleavage at the 5' splice site only, is normally seen (Miao and Abelson, 1993). In contrast, endonuclease containing Sen34p H242A shows a marked accumulation of 5' exon and the intron-3' exon 2/3 molecule. There is a nominal amount of 3' exon and intron product, but we cannot

FIGURE II-9. Two Active Site Model for Cleavage of Pre-tRNA by Yeast tRNA Splicing Endonuclease

Yeast pre-tRNA^{Phe} tertiary structure is depicted with a schematic diagram of the yeast endonuclease. In this model Sen2p and Sen34p each contain an active site for cleavage of the 5' and 3' splice sites, respectively. The Sen2p-Sen54p and Sen34p-Sen15p interaction, detected in the two-hybrid analysis, is depicted as well as the potential tRNA recognition mediated by Sen54p (see text for details).

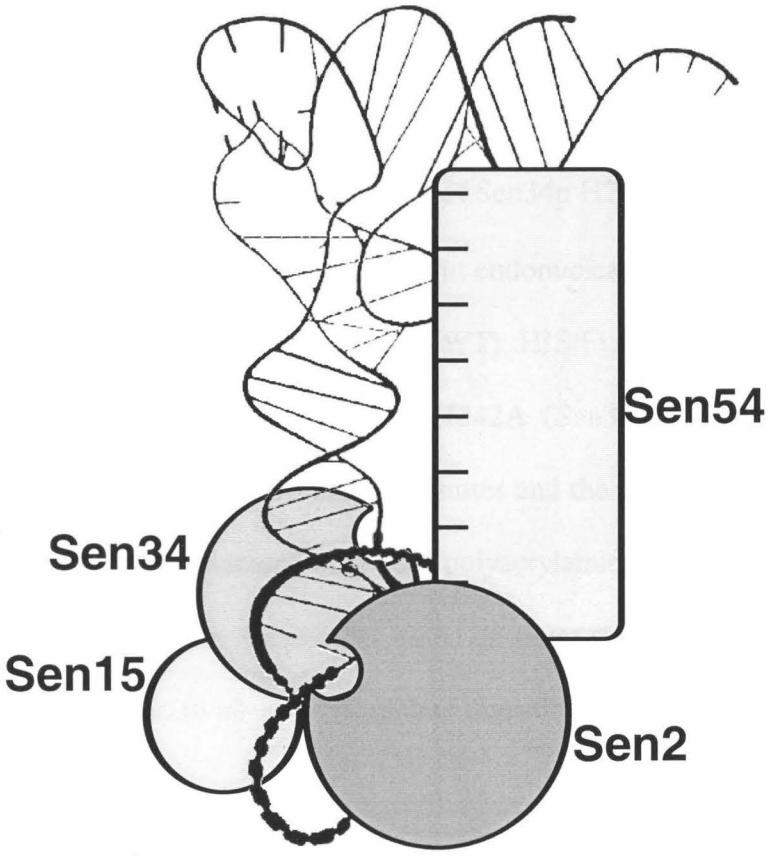


Figure II-9

FIGURE II-10. tRNA Splicing Activity of Sen34p H242A Containing Endonuclease Yeast pre-tRNA^{Phe} was incubated without endonuclease (pre) or with endonuclease purified from HIS/FLAG tagged Sen2p (WT), HIS/FLAG tagged wild type Sen34p (Sen34p) or HIS/FLAG tagged Sen34p H242A (Sen34p H242A) expressing yeast. Pre-tRNA was incubated for 5, 10 and 15 minutes and the resulting cleavage products (shown to the right) were fractionated on a 10% polyacrylamide gel. The amount of endonuclease used in each time course was similar based on silver staining (WT enzyme was used at 100X concentration to allow for complete digestion of pre-tRNA).

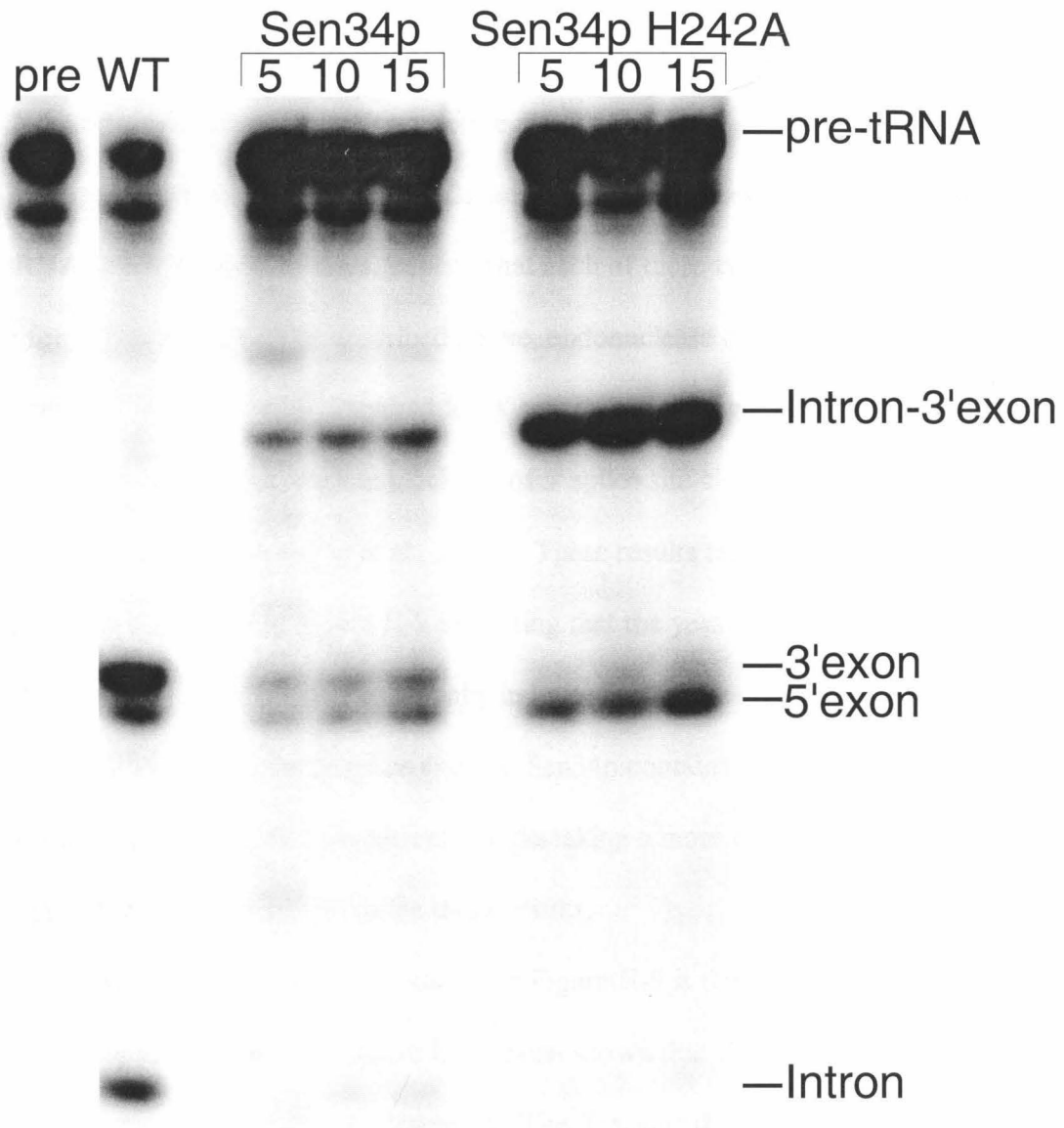

Sen34p
Sen34p H242A
pre WT
5 10 15
5 10 15
—pre-tRNA
—Intron-3'exon
—3'exon
—5'exon
—Intron

Figure II-10

determine if this is derived from retention of some 3' splice site cleavage activity in Sen34p H242A or from the presence of some wild type endonuclease in the purified mutant extract. However, the assay clearly demonstrates that endonuclease containing Sen34p H242A is, as predicted by the model, impaired in 3' splice site cleavage.

Discussion

Yeast tRNA endonuclease contains two active sites

The sequence similarity between Sen2p and Sen34p, and their homology to the archaeal tRNA splicing endonucleases, implied that each of these two subunits contains an active site for tRNA cleavage. As described above, endonuclease containing Sen34p H242A appears to be significantly impaired for 3' splice site cleavage, but 5' splice site cleavage occurs normally. A complete blockage of 5' splice site cleavage was seen in endonuclease containing *sen2-3* (Ho et al., 1990). These results taken together provide strong support for the model in Figure II-9 indicating that the yeast tRNA splicing endonuclease contains two distinct, functionally independent active sites. Sen2p contains the active site for cleavage at the 5' splice site and Sen34p contains the active site for cleavage at the 3' splice site. We are currently undertaking a more extensive mutagenesis of *SEN34* and *SEN2* in order to generalize these results.

Additional support for the model shown in Figure II-9 is derived from previous work on the *Xenopus* endonuclease where it has been shown that the two cleavage sites in yeast tRNA precursors have different properties. The *Xenopus* tRNA splicing endonuclease recognition of the 3' splice site requires a base-pair between a purine located 3 bases upstream of the 3' splice site and the conserved pyrimidine at position 31 in the anticodon loop (the A-I basepair) (Baldi et al., 1992). Cleavage at the 5' splice site

requires only a purine as the adjacent 5' nucleotide. It is conceivable that different specificities in cleavage could be obtained by a conformational change in a single active site occurring between cleavage reactions. For this to occur one might expect an obligatory reaction pathway, i.e., cleavage at a primary site followed by cleavage at the second site. However, we have shown that the yeast endonuclease can and does cleave the 5' and 3' splice sites in a random order (Miao and Abelson, 1993).

More recently, the accompanying paper by Tocchini-Valentini's laboratory has shown that accurate 3' splice site cleavage does not require the mature tRNA domain of the pre-tRNA (Di Nicola Negri et al., 1997). A small substrate containing only an intron and anticodon stem is cleaved by the *Xenopus* endonuclease at the 3' splice site but not at the 5' splice site. While both splice sites in an intact pre-tRNA are subject to the measuring mechanism, it appears that only the 5' splice site absolutely requires this mechanism.

The *H. volcanii* tRNA splicing endonuclease cleaves with the same general mechanism as the eucaryal endonuclease, producing 2',3' cyclic PO4 and 5'-OH termini, but has a different splice site recognition mechanism which fails to cleave yeast pre-tRNAs (Thompson and Daniels, 1990; Thompson and Daniels, 1988). The mature tRNA domain is not a requirement for the archaeal endonuclease. Archaeal tRNA splice sites are located in bulged regions separated by four helical base-pairs (Fig. II-1C). In the accompanying paper by Kleman-Leyer et al., the *H. volcanii* endonuclease is shown to exist as a homodimer, and thus contains two active sites. The view that each of these active sites is functional is strengthened by the evidence presented here for two distinct active sites in the yeast endonuclease. However, because of the near symmetry of the archaeal splice site structure, we cannot rule out that a single active site could act interchangeably.

Interestingly, most archaeal pre-tRNAs appear to contain the conserved A-I basepair of eucaryotic pre-tRNAs in the context of the four basepair stem separating the bulged loops of the splice sites. In this regard, the requirements for 3' splice cleavage are more similar between Archaea and Eucarya than 5' splice site cleavage. In both cases the splice site must be in a single strand bulge near a base-paired element (in the eucaryotic pre-tRNAs it is the A-I basepair).

Subunit interactions in endonuclease

In a complete matrix of two-hybrid interactions between the four subunits of endonuclease, strong interactions were seen only between the Sen2p and Sen54p pair and the Sen34p and Sen15p pair. Since all four subunits remain tightly associated through two different purification regimens, it seems likely that there are also interactions between the two presumed heterodimers. These interactions, however, would not be detected in a two-hybrid experiment if they exist only within the context of the intact enzyme, or if three of the subunits are required for the interaction. The involvement of each of the active site subunits in a specific heterodimeric interaction suggests that the different specificities of the two active sites may be determined by these interactions. For example the measuring mechanism, seen primarily in the recognition of the 5' splice site, may be determined through the interaction between Sen2p and Sen54p as indicated in Figure II-9. Sen54p is a very basic protein and as such might be expected to interact with the phosphate backbone of the pre-tRNA substrate, presumably in the mature tRNA domain. Sen15p could be primarily responsible for the recognition of the A-I basepair.

Evolutionary implications of the tRNA splicing endonuclease gene family

The existence of homologous tRNA splicing endonuclease genes in Eucarya and Archaea suggest that this enzyme was present in their common ancestor. Since endonuclease participates in tRNA splicing in both Eucarya and Archaea, it seems likely that the original function of this enzyme was the same. The discussion of the time of appearance of introns in evolution has been a lively one and it is safe to say that a consensus has not emerged. The simplest hypothesis consistent with the data presented here and in the accompanying papers is that the function of endonuclease in the common eucaryal-archaeal ancestor was the removal of introns from tRNA. A less parsimonius hypothesis is that the tRNA splicing function of this enzyme evolved independently in both kingdoms from an earlier, more generalized RNA processing function. In some present day Archaea the tRNA splicing endonuclease can apparently also function to remove introns from ribosomal RNA (Burggraf et al., 1993; Kjems and Garrett, 1985; Kjems and Garrett, 1991). Recently, Walter and co-workers have detected an intron in a yeast transcription factor gene that is not processed by the normal mRNA spliceosomal machinery (Cox and Walter, 1996; Sidrauski et al., 1996). They have demonstrated that tRNA ligase is involved in the splicing of this intron, but the tRNA splicing endonuclease has not as of yet been implicated. Clearly, members of this family of proteins have a number of functions, but it is not yet possible to determine which are original and which have been co-opted later in evolution.

Although the endonuclease active site is apparently a very primitive protein domain, it is clear that its substrate specificity has been altered in evolution because the recognition of the splice sites is different in Eucarya and Archaea. We speculate that the peculiarities of the eucaryal system are in part explained by the heterodimeric interaction

of the divergent subunits-Sen2p with Sen54p and Sen34p with Sen15p. Many fascinating questions remain concerning this evolutionary specialization. Since Sen54p and Sen15p have no known homologs, where did these proteins come from? Do Sen54p and Sen15p play a role in splice site selection? Can the active site subunits function in conjunction with other protein subunits, thereby modifying endonuclease specificity for other RNA processing reactions? We are now able to begin to address these questions.

A further complexity of the evolution of the tRNA splicing process is indicated by the lack of a yeast tRNA ligase (or T4 RNA ligase) homolog in the entire *M. jannaschii* genome (Bult et al., 1996; CRT and EAA, unpublished data). Interestingly, several archaeal genomes contain a clear homolog of a bacterial RNA ligase which can form a 2'-5' linkage when joining tRNA half-molecules derived from endonucleolytic cleavage of yeast pre-(Arn and Abelson, 1996; Greer et al., 1983). The complete ligation mechanism for archaeal tRNA splicing has not yet been determined, but it has been demonstrated that an RNA splicing ligase activity present in *H. volcanii* extracts does not require ATP or exogenous phosphate (Gomes and Gupta, 1997). The *E. coli* 2'-5' RNA ligase is the only known RNA ligase activity which does not require a nucleotide triphosphate cofactor (Arn and Abelson, 1998). We are presently attempting to ascertain whether the archaeal 2'-5' RNA ligase homolog functions in tRNA splicing *in vivo*.

Experimental Procedures

Yeast Methods

Yeast were transformed by the lithium acetate procedure (Gietz and Schiestl, 1995). Yeast minimal media (SD) contained Bacto yeast nitrogen base without amino acids (YNB - aa, 6.7g/l), dextrose (2%w/v) and a supplemental amino acid mix (Bio101)

lacking various amino acids for auxotroph selection. All other yeast media was prepared according to Sherman et al., 1986. For fermentation growth of CRT44, SD-histidine-tryptophan, prepared as described above, was used and supplemented with 25ug/ml Kanamycin. The fermentation was fed using SD + 10X-histidine-tryptophan amino acid mix with constant flow for 24 hours.

Two-hybrid strains were obtained by a generous gift from P. Bartel of S. Fields laboratory and are described in Table I. All two hybrid techniques used are described by Hannon et al. (1995). Plates used for analysis contained 5mM 3-aminotriazole which suppressed background growth associated with histidine reporter assay (Hannon and Bartel, 1995).

Strains

E.coli strains utilized - HB101. XL1-Blue (Novagen). DH5 (Gibco-BRL). *Saccharomyces cerevisiae* strains are described in Table I (Appendix II). Knockout strains YPH274ΔSEN2 and CRT54 were created utilizing methods described in Guthrie and Fink, 1992. Knockout strains of *SEN15* (CRT15) and *SEN34* (CRT34) were created utilizing methods and plasmids generously obtained from A. Wach and described by Wach et al. (1996). All knockouts were confirmed by PCR analysis and each knockout was complemented with the corresponding ORF only.

Strain CRT44 was prepared by transforming YPH274ΔSEN2pC10 with YpH8/FLAG SEN2, plating on selective media (SD-his-trp-ura), and streaking out isolates on selective media (SD-his-trp) + 5 FOA to select for loss of pC10 (URA3 CEN/ARS).

Plasmid Preparation

All restriction enzymes used were obtained from New England Bio-Labs and Boehringer Mannheim and used in accordance with the manufacturer's instructions. Restriction digests, DNA ligations and electrophoresis of DNA were performed by standard methods of Sambrook et al. (1990).

pC10 was prepared by subcloning a EcoRV/EcoRI fragment of DNA derived from the YcP50 *sen2-3* complementing clone, into the vector pRS416 (URA3)(Sikorski and Hieter, 1989). This fragment contains an open reading frame coding for the 377 amino acid *SEN2* gene plus 250 base pairs in the promoter region and 345 base pairs of downstream region. pC10 was capable of complementing a knockout of the *SEN2* ORF.

YpH8/FLAG was prepared by PCR of the *CUP1* promoter, from JD51, to allow cloning of the *CUP1* promoter into the HindIII/EcoRI site of pBluescript (Stratgene). This plasmid was then digested with XbaI, filled-in and cut with PstI and the PstI-NaeI fragment from pGBT9 (from S. Fields), containing the *ADH1* terminator sequence, was cloned into this site, yielding pBSCUP/TERM. This plasmid was then cut with HindIII/SacI and the resulting fragment was subcloned into pRS424 (Sikorski and Hieter, 1989), to yield YpCUP/TERM. This plasmid was cut with NdeI and EcoRI and the HIS/FLAG linker was cloned into this site, generating YpH8/FLAG. YpH8/FLAG SEN2 was prepared by digestion of pGBT9-SEN2 with EcoRI and PstI, gel isolation of the fragment containing SEN2 and subcloning into EcoRI/PstI digested YpH8/FLAG. YpH8FLAG SEN34 was prepared by digestion of pGBT9-SEN34 with EcoRI and PstI, gel isolation of the fragment containing SEN34 and subcloning into EcoRI/PstI digested YpH8/FLAG. Oligonucleotide directed mutagenesis was performed on single stranded

YpH8/FLAG-SEN34 to effect an H242A mutation according to the protocol provided by the manufacturer (Amersham).

Two-hybrid vectors were prepared by cloning PCR products of *SEN15* (oligos SEN15-N 5'-GGAATTCGCAACGACAGATATCATATC-3' and -C 5'-TTCTGCAGATCTTCAATTTCTTTTCGGTTTTCG- 3'), *SEN54* (oligos SEN54-N 5'-GGAATTCCAATTCGCTGGGAAG-3' and -C 5'-TTCTGCAGTTAATGCTTCTTCCATCTTTG-3'), *SEN2* (oligos SEN2-N 5'-GGAATTCATGTCTAAAGGGAGGG and -C 5'-CGGGATCCCAATGAAAAGTTTAGAGC) and *SEN34* (oligos SEN34-N 5'-CGGAATTCGGGATCCATATGCCACCGCTAGTATTTGAC-3' and -C 5'-CGGGATCCCGGGAAGCTTAACCAAATCCAGCCC-3') digested with EcoRI and BamHI and cloned into EcoRI/BamHI digested pGBT9 and pGAD424 (Hannon and Bartel, 1995) (these constructs were sequenced).

Antibodies and Western Blots

Polyclonal antibodies to Sen34p were produced by injection of rabbit (CoCalico Biosciences) with peptides NH_2-DKTTAEELQRLDKSSC-COOH, and NH_2-DIRSDSDSLSRDDIC-COOH. Polyclonal antibodies to Sen2p were produced by injection of rabbits with peptide derived from the C-terminus of Sen2p (NH_2-MSKGRVNQKRYPLC-COOH). All peptides were produced by the Caltech Peptide Facility. The antibodies were used directly in rabbit serum at the indicated dilution. Polyclonal antibodies to both Sen54p and Sen15p were prepared in the following manner: Antigen was derived from purification of HIS8/FLAG tagged protein from *E.coli*. Each gene was cloned into the EcoRI/PstI site of YpH8/FLAG by subcloning from pGBT9

SEN54 and pGBT9 SEN15. Each construct was then cut with NdeI/PstI and subcloned into Pet11a. Purification was carried out under denaturing conditions by Ni-NTA chromotography as suggested by Qiagen. Protein eluted from the Ni-NTA resin was gel purified by Prep-Cell Chromatography (Biorad) as per the manufacturers suggested protocol with elution into buffer containing 0.05% SDS. The eluate was directly injected into Swiss Webster mice as per the protocol of Susan Ou of the Caltech Monoclonal Antibody Facility.

SDS-PAGE was performed on 10%-20% gradient gels (Biorad). Proteins were transferred to PVDF (Immobilon) by semi-dry transfer (Biorad). Western blots were performed as described by Harlow and Lane (1988) and were visualized by enhanced chemiluminescence (ECL, Amersham).

Endonuclease Assay

Endonuclease was assayed by incubation of 2µl of extract for specified amount of time at 30°C in 10µl reaction containing 20mM Na-HEPES pH 7.5, 5mM $MgCl_2$, 2.4mM spermidine-HCl pH 7.5, 0.1mM DTT, 0.4% Triton X-100 and 10 fmole pre-$tRNA^{Phe}$ body-labeled with [α-^{32}P]UTP (specific activity ~500 cpm per femtomole tRNA) by T7 polymerase transcription according to Sampson and Saks (1993). RNA products were phenol/chloroform extracted, precipitated and fractionated on a 10% polyacrylamide gel containing 8M urea. Gels were exposed to phosphorimager for 8 hours.

Purification of tRNA Endonuclease

Purification of endonuclease was carried out utilizing three separate purification strategies. The first is described by Rauhut et al. (1990). This protein was used to obtain the gene for the 34kD subunit.

Affinity purification was carried out utilizing yeast strain CRT44 grown by fermentation at 30°C. The cells were harvested and stored at -70°C. All subsequent steps were performed at 4°C. 150g of yeast cells resuspended in 1.5X breaking buffer (37.5mM Tris pH8.0, 15%(v/v) Glycerol, 5mM B-ME, 1mM EDTA, 0.1% (v/v) Triton X-100, 405 mM NH_4SO_4) to 300mls was ground in a Bead-Beater (Bio-Spec Corp) for 20x1 minute intervals and Triton X-100 (Boehringer-Mannheim) was added to 1%. The extract was incubated for 1 hour and then centrifuged (40,000 rpm X 1.5 h). The supernatant was loaded onto a 300ml DEAE-sepharose CL-6B equilibrated in 1X breaking buffer + 1% Triton X-100. DEAE effluent was diluted approximately three-fold and loaded onto a 100ml Heparin-HyperD (BioSepra) column equilibrated in buffer A (25mM MOPS pH7.1, 10% Glycerol, 1mM EDTA, 5mM BME, 0.9% Triton X-100, 100mM NaCl), at 1ml/minute. The column was washed with 500ml buffer B (25mM Tris pH8.0, 10% Glycerol, 5mM BME, 0.3% Triton X-100, 150mM NaCl). A 2X200ml gradient of 0.1-0.8mM NaCl in buffer B was applied. Active fractions were pooled and bound in batch to 5ml Ni-NTA agarose (Quiagen, equilibrated in buffer B) for 1.5h. The suspension was poured into a 2 X 10 mm column and washed extensively with 5mls of each of the following buffers: buffer N (25mM Tris-HCl ph 7.5, 150mM NaCl, 10% Glycerol, 5mM B-mercaptoethanol, 0.3% Triton X-100)+ 650mM NaCl + 15mM imidazole; buffer N + 650mM NaCl + 20mM imidazole; buffer N + 20mM imidazole; buffer N + 40 mM imidazole. Endonuclease was eluted in Buffer N + 150mM imidazole. The eluate was

loaded onto a 1ml M2 monoclonal anti-FLAG antibody column (Kodak) equilibrated in Buffer N. The column was washed in buffer N and eluted with Buffer N + 150ug/ml FLAG peptide. This endonuclease was used to prepare the 54 kD subunit.

A modified purification protocol was developed to prepare the 15 kD subunit and purification of mutant containing endonuclease. Yeast cells (strain CRT44 or as indicated in the text) were resuspended and broken as above. The extract was then spun at 40,000 rpm X 2h to pellet the nuclear membranes. The pellet was resuspended in 50ml of Buffer R (50mM Tris, pH7.5, 150mM NaCl, 5% Glycerol, 1mM EDTA, 5mM BME, 0.9% Triton X-100) and incubated for 1 h. The extract was subjected to centrifugation as before and the supernatant was loaded directly onto a 2.5ml M2 monoclonal anti-FLAG antibody column. The column was washed with buffer R and endonuclease was eluted with Buffer S (30mM Tris, pH8.0, 500mM NaCl, 10% Glycerol, 5mM BME, 0.3% Triton X-100, 5mM Imidazole) containing 150ug/ml FLAG peptide. The elution was incubated with 1ml of Ni-NTA agarose, equilibrated in Buffer S, for 1.5h and poured into a 3ml column. The column was washed with Buffer S + 40mM imidazole and endonuclease was eluted in Buffer S + 150mM imidazole.

Protein Sequencing

Three separate procedures were utilized for sequencing the subunits of endonuclease. In all three cases, protein was submitted as a gel slice with various amounts of protein in each. For the 34kD subunit, ~25ug of protein was submitted to the protein sequencing facility at Cold Spring Harbor and was digested *in gel* with lysly-endoprotease followed by extraction and HPLC purification of the resulting peptides following the

protocols of the facility (Dan Marshak, pers.comm.). Fractions containing single peaks were sequenced by Edman degradation.

For the 54kD subunit, ~5ug of protein was digested *in gel* with trypsin as described by Hellman et al. (1995) (Hellman et al., 1995), except that 0.3ug of Promega modified pig trypsin and speedvac drying of the gel slice was used. Peptides were HPLC purified and fractions containing single mass peaks, as determined by mass spectrometry on a Voyager Mass Spec, were submitted for Edman Degradation sequencing on an Applied Biosystems sequencer (Gary Hathaway, pers.comm.).

For the 15kD subunit, ~5ug of protein was submitted to the protein sequencing facility at the City of Hope (Duarte, CA) for *in gel* trypsin digestion followed by peptide sequencing utilizing a Beckman LC/MS-MS protein sequencer (K. Swiderek, pers. comm). Data obtained was used to search the yeast genome database (YGD).

Acknowledgements

We wish to acknowledge first and foremost the protein sequencing work done by Gary Hathaway and Dirk Krapf at Caltech Protein Analysis Laboratory, Dr. Terry Lee and Dr. Kristine Swiderek at the City of Hope Protein Sequencing Facility and Dr. Dan Marshak at the Cold Spring Harbor Protein Sequencing Facility. Special thanks to Paul Bartel and Stan Fields for two-hybrid strains and plasmids. We also want to thank Chuck Daniels for communicating the endonuclease sequences of *Haloferax volcanii* and *Methanobacterium thermoautotrophicum* prior to publication and Drs. Mel Simon and Ung-Jin Kim for sequence of the *Pyrobaculum aerophilum* homolog. We wish to thank Drs. Christine Guthrie, Eric Phizicky, Chris Greer and Anita Hopper for their insight and critical review of the manuscript and members of the Abelson laboratory for their advice

and support. This work was supported by generous grant from the American Cancer Society.

References

Abelson, J. (1991). RNA splicing in yeast. Harvey Lecture *85*, 1-42.

Arn, E. A., and Abelson, J. (1998). RNA ligases: function, mechanism, and sequence conservation. In RNA structure and function (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 695-726.

Arn, E. A., and Abelson, J. N. (1996). The 2'-5' ligase from Escherichia coli: purification, cloning and genomic disruption. J. Biol. Chem. *271*, 31145-31153.

Attardi, G., and Schatz, G. (1988). Biogenesis of mitochondria. Ann. Rev. Cell Biol. *4*, 289-333.

Baldi, M. I., Mattoccia, E., Bufardeci, E., Fabbri, S., and Tocchini-Valentini, G. P. (1992). Participation of the intron in the reaction catalyzed by the *Xenopus* tRNA splicing endonuclease. Science *255*, 1404-1408.

Bartel, P. L., and Fields, S. (1995). Analyzing protein-protein interactions using 2-hybrid system. Meth. Enz. *254*, 241-263.

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.

Biniszkiewicz, D., Cesnaviciene, E., and Shub, D. A. (1994). Self-splicing group-I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. EMBO Journal *13*, 4629-4635.

Bult, C. J., White, O., Olsen, G. J., Zhou, L. X., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weidman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, H. P., Fraser, C. M., Smith, H. O., Woese, C. R., and Venter, J. C. (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 rRNA gene of the archaeon Pyrobaculum aerophilum. Proceedings of the National Academy of Sciences *90*, 2547-2550.

Cech, T. (1993). Structure and mechanism of the large catalytic RNAs: group I and group II introns and ribonuclease P. In RNA World, R. F. Gesteland and J. F. Atkins, eds. (Cold Spring Harbor: Cold Spring Harbor Laboratory Press), pp. 206-208.

Cox, J. S., and Walter, P. (1996). A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. Cell *86*, 391-404.

Culbertson, M. R., and Winey, M. (1989). Split tRNA genes and their products: a paradigm for the study of cell-function and evolution. Yeast *5*, 405-427.

Culver, G. M., McCraith, S. M., Zillmann, M., Kierzek, R., Michaud, N., Lareau, R. D., Turner, D. H., and Phizicky, E. M. (1993). An NAD derivative produced during tRNA splicing: ADP-ribose 1"-2" cyclic phosphate. Science *261*, 206-208.

Cusack, S., Yaremchuk, A., and Tukalo, M. (1996). The crystal structure of the ternary complex of Thermus thermophilus seryl-tRNA synthetase with tRNAser and a seryl-adenylate analog reveals a conformational switch in the active site. Biological Chemistry Hoppe Seyler *377*, 343-356.

Di Nicola Negri, E., Fabbri, S., Bufardeci, E., Baldi, M. I., Gandini Attardi, D., Mattoccia, E., and Tocchini-Valentini, G. P. (1997). The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements. Cell *89*, 859-866.

Fields, S., and Song, O. K. (1989). A novel genetic system to detect protein- protein interactions. Nature *340*, 245-246.

Fischer, U., Huber, J., Boelens, W. C., Mattaj, I. W., and Luhrmann, R. (1995). The HIV-1 rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. Cell *82*, 475-483.

Gietz, R. D., and Schiestl, R. H. (1995). Transforming yeast with DNA. Methods In Molecular and Cellular Biology *5*, 255-269.

Gomes, I., and Gupta, R. (1997). RNA splicing ligase activity in the archaeon *Haloferax volcanii*. Biochemical and Biophysical Review *237*, 588-594.

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

Guthrie, C., and Fink, G. R. (1991). Guide to yeast genetics and molecular biology (San Diego: Academic Press).

Hannon, G., and Bartel, P. (1995). Identification of interacting proteins using the 2-hybrid system. Methods In Molecular and Cellular Biology *5*, 289-297.

Harlow, E. (1988). Antibodies: a laboratory manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

Hellman, U., Wernstedt, C., Gonez, J., and Heldin, C. H. (1995). Improvement of an in-gel digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. Analytical Biochemistry *224*, 451-455.

Ho, C. K., Rauhut, R., Vijayraghavan, U., and Abelson, J. (1990). Accumulation of pre-tRNA splicing 2/3 intermediates in a *Saccharomyces cerevisiae* mutant. Embo Journal *9*, 1245-1252.

Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984). A short amino-acid sequence able to specify nuclear location. Cell *39*, 499-509.

Kiebler, M., Becker, K., Pfanner, N., and Neupert, W. (1993). Mitochondrial protein import: specific recognition and membrane translocation of preproteins. Journal Of Membrane Biology *135*, 191-207.

Kjems, J., and Garrett, R. A. (1985). An intron in the 23S rRNA gene of the Archaebacterium *Desulfurococcus mobilis*. Nature *318*, 675-677.

Kjems, J., and Garrett, R. A. (1991). rRNA introns in Archaea and evidence for RNA conformational changes associated with splicing. Proceedings of the National Academy of Sciences *88*, 439-443.

Kleman-Leyer, K., Armbruster, D. A., and Daniels, C. J. (1997). Properties of *H. volcanii* tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Cell *89*, 839-848.

Lee, M. C., and Knapp, G. (1985). tRNA splicing in *Saccharomyces cerevisiae*: secondary and tertiary structures of the substrates. J. Biol. Chem. *260*, 3108-3115.

Lupas, A., Vandyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162-1164.

Madhani, H. D., and Guthrie, C. (1994). Dynamic RNA-RNA interactions in the spliceosome. Ann. Rev. Genet. *28*, 1-26.

McCraith, S. M., and Phizicky, E. M. (1991). An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2'-phosphate from ligated tRNA to an acceptor molecule. J. Biol. Chem. *266*, 11986-11992.

Miao, F., and Abelson, J. (1993). Yeast tRNA-splicing endonuclease cleaves precursor tRNA in a random pathway. J. Biol. Chem. *268*, 672-677.

Moore, M., and Sharp, P. (1993). Splicing of precursor mRNAs by the spliceosome. In RNA World, R. F. Gesteland and J. F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 303-349.

Ogden, R. C., Lee, M. C., and Knapp, G. (1984). Transfer RNA splicing In Saccharomyces cerevisiae: defining the substrates. Nucleic Acids Res. *12*, 9367-9382.

Palmer, J. R., Nieuwlandt, D. T., and Daniels, C. J. (1994). Expression of a yeast intron-containing tRNA in the Archaeon Haloferax volcanii. J. Bact. *176*, 3820-3823.

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

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.

Reinhold-Hurek, B., and Shub, D. A. (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. Nature *357*, 173-176.

Reyes, V. M., and Abelson, J. (1988). Substrate recognition and splice site determination in yeast tRNA splicing. Cell *55*, 719-730.

Robbins, J., Dilworth, S. M., Laskey, R. A., and Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell *64*, 615-623.

Sherman, F., Fink, G. R., and Hicks, J. B. (1986). Methods in yeast genetics (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sidrauski, C., Cox, J. S., and Walter, P. (1996). tRNA ligase is required for regulated mRNA splicing in the unfolded protein response. Cell *86*, 405-413.

Sikorski, R. S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19-27.

Thompson, L. D., Brandon, L. D., Nieuwlandt, D. T., and Daniels, C. J. (1989). Transfer RNA intron processing in the halophilic archaebacteria. Canadian Journal Of Microbiology *35*, 36-42.

Thompson, L. D., and Daniels, C. J. (1988). A $tRNA^{trp}$ intron endonuclease from Halobacterium volcanii: unique substrate recognition properties. 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.

Wach, A. (1996). PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in Saccharomyces cerevisiae. Yeast *12*, 259-265.

Wen, W., Meinkoth, J. L., Tsien, R. Y., and Taylor, S. S. (1995). Identification of a signal for rapid export of proteins from the nucleus. Cell *82*, 463-473.

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.

Westaway, S. K., Phizicky, E. M., and Abelson, J. (1988). Structure and function of the yeast tRNA ligase gene. J. Biol. Chem. *263*, 3171-3176.

Winey, M., and Culbertson, M. R. (1988). Mutations affecting the tRNA splicing endonuclease activity of *Saccharomyces cerevisiae*. Genetics *118*, 49-63.

Yanagisawa, S., and Izui, K. (1993). Molecular cloning of 2 DNA-binding proteins of maize that are structurally different but interact with the same sequence motif. J. Biol. Chem. *268*, 16028-16036.

Chapter III – Three-Dimensional Structure and Evolution of a tRNA Splicing Enzyme (Hong Li, Christopher R. Trotta, John Abelson published in Science 280:279-284, 1998; Appendix III; Chris Trotta's contributions include, identifying, cloning and purification of the *M. jannaschii* endonuclease, data collection from the crystals, extensive writing of the manuscript and intellectual contributions related to the understanding and generation of the models to describe the active site and enzyme architecture)

Abstract

The splicing of transfer RNA precursors is similar in Eucarya and Archaea. In both kingdoms an endonuclease recognizes the splice sites and releases the intron, but the mechanism of splice site recognition is different in each kingdom. The crystal structure of the endonuclease from the archaeon *Methanococcus jannaschii* was determined to a resolution of 2.3 angstroms. The structure indicates that the cleavage reaction is similar to that of ribonucleaseA and the arrangement of the active sites is conserved between the archaeal and eucaryal enzymes. These results suggest and evolutionary pathway for splice site recognition.

Introduction

Introns are found in the tRNA genes of organisms in all three of the great lines of descent: the Eucarya, the Archaea and the Bacteria. In Bacteria tRNA introns are self-splicing group I introns and the splicing mechanism is autocatalytic (1). In Eucarya tRNA introns are small and they invariably interrupt the anticodon loop one base 3' to the anticodon. They are removed by the step-wise action of three enzymes: i) an endonuclease that cleaves the pre-tRNA, ii) a ligase that joins the exons together in a

reaction that requires the hydrolysis of both ATP and GTP (2), and iii) a phosphotransferase that transfers a 2'-phosphate left at the splice site to nicotinamide adenosine dinucleotide (NAD) acceptor (3). In Archaea the introns are also small and often reside in the same location as eucaryal tRNA introns. In some cases, however, they interrupt the tRNA gene at sites other than the anticodon (4). As in Eucarya, splicing in Archaea is also catalyzed by proteins. An endonuclease recognizes and cleaves the pre-tRNA. It is virtually certain, however, that the mechanism of ligation is different from its counterpart in Eucarya as there is no obvious homolog of the eucaryal tRNA splicing ligase in the complete genome sequence of several members of the Archaea (5). Furthermore, the ligation mechanism in *Haloferax volcanii* extracts has been shown to proceed in an ATP-independent manner (6).

The tRNA splicing endonucleases of both the Eucarya and the Archaea cleave the pre-tRNA substrate leaving 5'-hydroxyl and 2',3' cyclic phosphate termini, but these enzymes recognize their substrates differently. The eucaryal enzyme uses a measuring mechanism to determine the position of the universally positioned splice sites relative to the conserved domain of pre-tRNA (Figure III-1A)(7). In Archaea, where introns are not equivalently positioned, the enzyme recognizes a pseudo-symmetric substrate in which two bulged loops of three bases are separated by a stem of four base pairs (Figure III-1A) (8) . These observations suggested that the tRNA splicing mechanisms of the three major kingdoms evolved independently.

Recently, however, characterization of eucaryal and archaeal endonucleases has provided new insights that reveal a common origin for this step in splicing. The yeast endonuclease is an $\alpha\beta\gamma\delta$ heterotetramer with subunit molecular weights of 15kD, 34kD, 44kD and 54kD (9). The first archaeal endonuclease to be characterized, from *Haloferax*

FIGURE III-1. **A**. Consensus sequence and secondary structure of precursor tRNA substrate for yeast and archaeal endonucleases (modified from reference 9). Splice sites are indicated by short arrows. (O), (X) = unconserved bases in regions of conserved or variable secondary structure respectively, (Y) = pyrimidines, (R) = purines. The endonuclease in yeast is proposed to interact with the mature domain and measure the distance to the splice sites. In contrast, only a minimum region containing two three-base nucleotides separated by a four-base-paired helix (gray shaded region) is both necessary and sufficient for the splice site recognition by an archaeal endonuclease (16).
B. Comparison of endonuclease models in Eucarya (heterotetramer) and Archaea (homodimer).

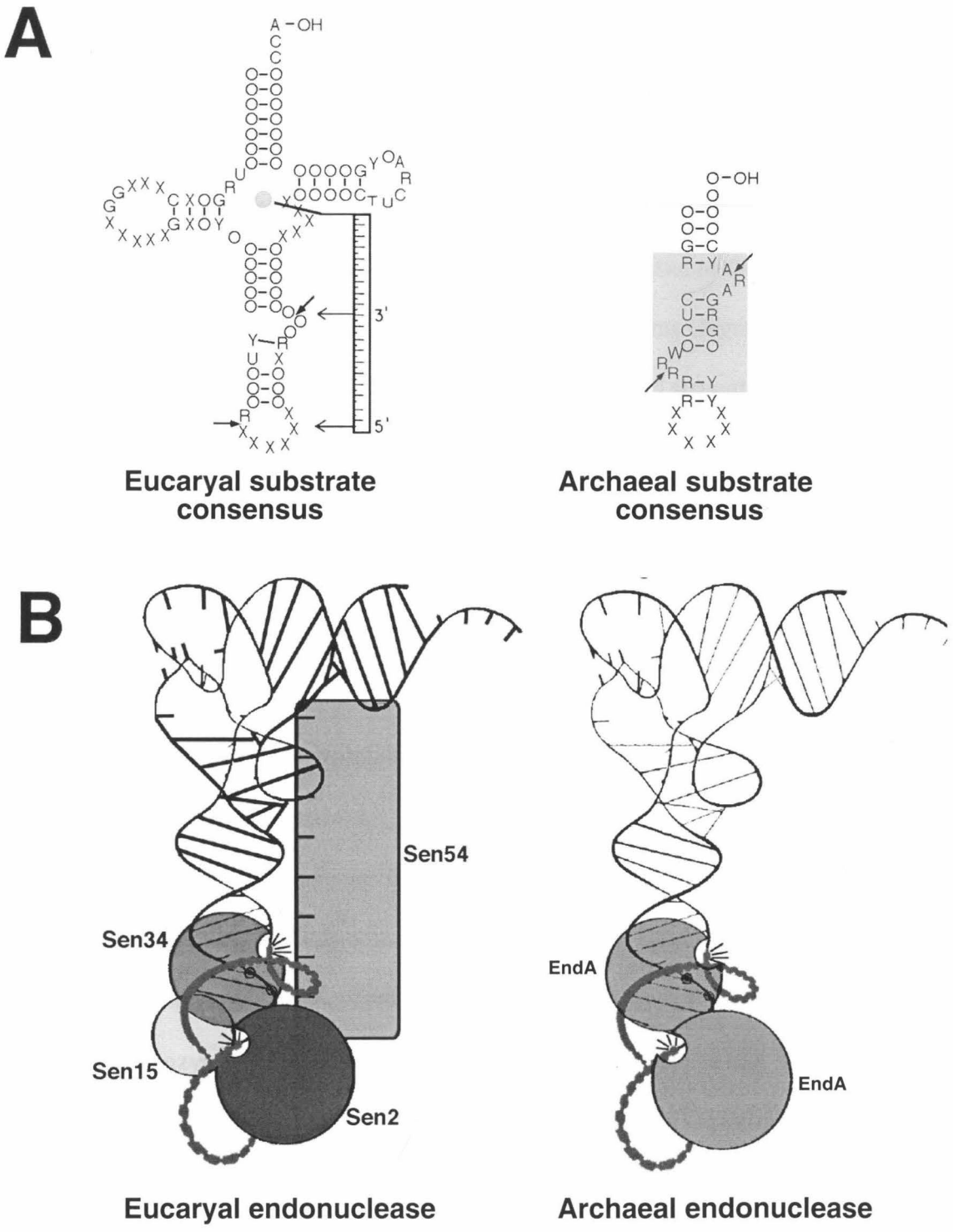

A
Eucaryal substrate consensus
Archaeal substrate consensus
B
Sen54
Sen34
Sen15
Sen2
EndA
EndA
Eucaryal endonuclease
Archaeal endonuclease

Figure III-1

volcanii, is a dimer of identical 37kD subunits (10). The *H. volcanii* enzyme was found to be homologous to both the 34kD and 44kD subunits of the yeast enzyme, suggesting that these distinct subunits contain active sites for tRNA splicing. This hypothesis was strongly supported by the finding that a mutant in the 44kD subunit, *sen2-3*, is selectively defective in 5' splice site cleavage (11) while a mutant in the 34kD subunit, H242A, is selectively defective in 3'-splice site cleavage (9). These results suggested related models for tRNA splice site recognition and cleavage in the two kingdoms (Figure III-1B). In yeast, the 34kD subunit recognizes the 3'-splice site and interacts strongly with the 15kD subunit (as seen in a yeast two-hybrid experiment) (9). The 44kD subunit recognizes the 5'-splice site and interacts with the 54kD subunit which, we have suggested, may embody the molecular ruler (9). In *H. volcanii*, the homodimeric enzyme is proposed to interact with the pseudo-symmetric substrate so that each splice site is cleaved by a symmetrically disposed active site monomer. Thus it is likely that the spatial disposition of the active site subunits has been conserved in evolution and in the eucaryal lineage a distinct mechanism of splice site recognition has evolved.

There are apparent advantages in choosing the simpler archaeal system to begin to understand the enzymology of this conserved enzyme family. The euryarcheota group of Archaea which includes *Methanococcus jannaschii* contains genes that are homologous to the *H. volcannii* enzyme but whose products are about half the size as that of *H. volcanii* (179 amino acids in the case of *M. jannaschii*). *M. jannaschii* is a thermophilic organism collected at a thermal vent on the East Pacific Rise at a depth of 2600 meters. It can withstand 200 atmosphere of pressure and grows at an optimum temperature of about 85°C. Its complete genome sequence has been determined by the TIGR group (12). We believed that a high resolution structure of the simpler archaeal enzyme would shed light

on the mechanism of the more complicated but related eucaryal endonuclease and thus embarked on a structural study of the *M. jannaschii* endonuclease. We report here the structure of the *M. jannaschii* endonuclease determined by X-ray diffraction and refined to a resolution of 2.3Å.

Structure Determination

The *M. jannaschii* endonuclease gene with an 8xHis and Flag tag sequences (9) at its 5' end was cloned by PCR from genomic DNA into the pET11a vector. Soluble protein was obtained by overexpression in *E coli* (13). The structure was determined by the multiwavelength anomalous diffraction (MAD) method (14) on crystals derivatized by $Au(CN)_2$. Details of the structure determination are described in the legend to Table 1. MAD data were treated as a special case of multiple isomorphous replacement with inclusion of anomalous scattering (Table 1). Phases were initially computed using a maximum likelihood procedure (Table 1) and were markedly improved by non-crystallographic averaging and solvent flattening (Table 1). A typical section of averaged electron density map is shown in Figure III-2. Four subunit polypeptide chains could be unambiguously placed in the averaged electron density. The final model, which includes the tetrameric endonuclease of residues 9-179, 53 water oxygen atoms, 2 partially occupied SO_4^- ions, and 4 gold atoms, was refined to 2.3 Å with good sterochemistry to an R factor of 0.204 and a free R of 0.268 (Table 1).

A Two Domain Monomer

The endonuclease monomer folds into two distinct domains, the NH_2-terminal domain (residues 9-84) and the COOH-terminal domain (residues 85-179) (Figure III-3).

FIGURE III-2. A section of the averaged electron density map contoured at 1σ shown together with the final refined structure model. The real space correlation coefficient of the final endonuclease structure model with the averaged map is 0.68 as calculated by O (32).

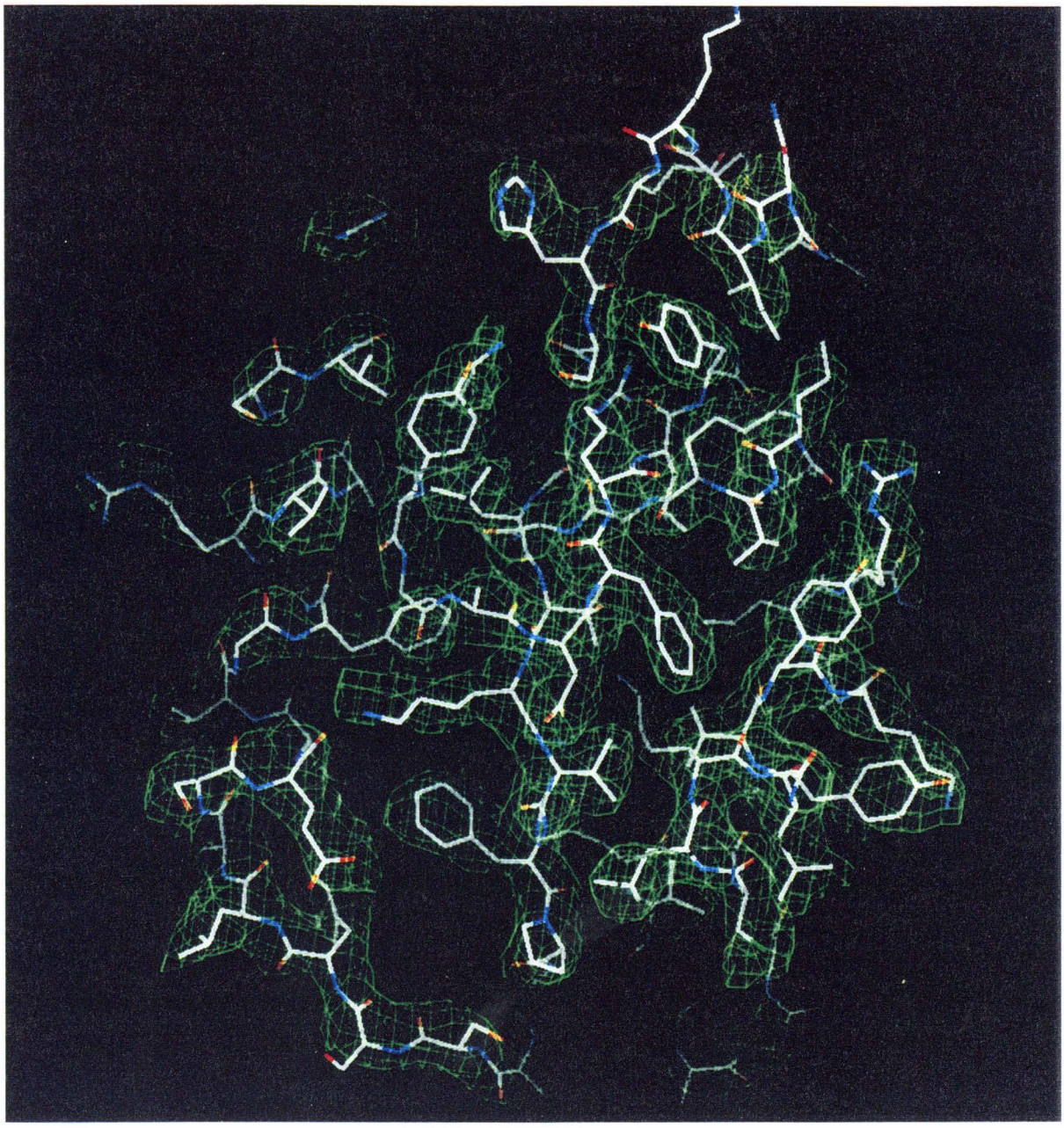


Figure III-2

FIGURE III-3. Ribbon representation of one subunit of *M. jannaschii* endonuclease as obtained with the program RIBBONS (37). Secondary structure elements as referred in text are indicated.

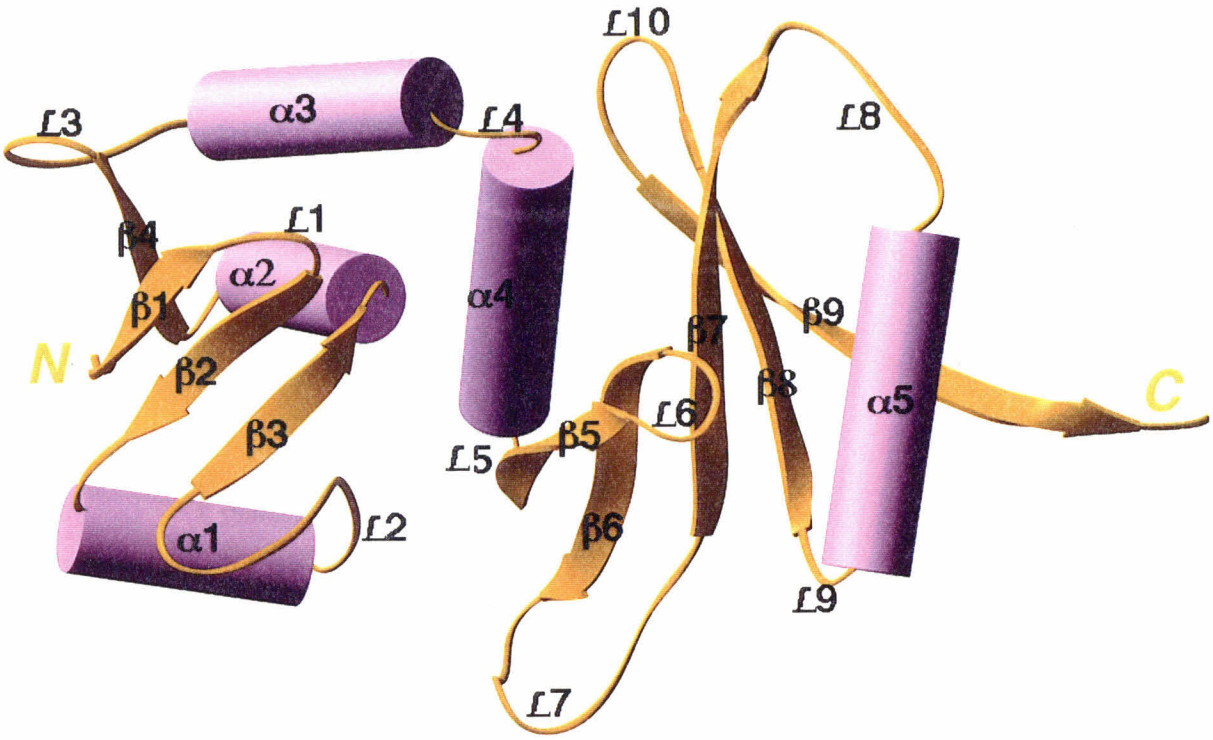

L10
L3
α3
L4
L8
β4
L1
α2
α4
β9
β1
β7
N
β2
β8
α5
C
L6
β3
β5
L5
α1
L2
β6
L9
L7

Figure III-3

The NH_2-domain consists of a mixed antiparallel/parallel β-sheet and three α helices. The first three β strands (β1-β3) are antiparallel, and together with β4, are packed against two nearly perpendicular α-helices, α1 and α2. The third α-helix, α3, is associated with α2 via an antiparallel coiled-coil. The inclination angle between α2 and α3 is 20° such that apolar residues at the interface mesh together with a "*knobs into holes*'" packing (15). Hydrophobic residues packed at the interface of the coiled-coil include Leu^{47}, Leu^{51}, and Ile^{54} from α2 and Phe^{71}, Leu^{74}, and Ala^{78} from α3. The connection between β4 and α3 (residues 64-67) is disordered, implying that this loop is flexible. This feature is consistent with the proteolytic sensitivity of this region (16). A short loop consisting of three charged residues (Glu^{81}, Glu^{82}, Arg^{83}) links the N-terminal domain to the C-terminal domain. The C-terminal domain is the α/β type and consists of a central four-stranded mixed β-sheet containing strands β5-β6-$\overline{\beta 7}$-$\overline{\beta 8}$ (upperlines denote parallel strands) cradled between helices α4 and α5. Two basic folding motifs lead to this fold; the rarely seen αββ (α4β5β6) and the commonly seen βαβ (β7α5β8). The last β strand, β9, is partially hydrogen bonded with β8 and has little involvement in the C-terminal domain folding itself; rather, it participates in dimerization interactions (see below). The interface between the NH_2-domain and the COOH-domain consists mainly of aromatic and polar residues, all of which are from helices α2 and α4. One notable feature is a hydrogen bond triad mediated by a single water molecule linking three conserved residues, Glu^{49}, Tyr^{92}, and Arg^{96}.

The Homotetramer: A Dimer of Two Dimers

When viewed along a non-crystallographic two-fold axis, the homotetrameric *M. jannaschii* endonuclease resembles a parallelogram with the four subunits occupying the four corners (Figure III-4A). The quaternary structure of the endonuclease is unusual in that it does not assume the D_2 point group symmetry observed in most homotetrameric proteins. It therefore may be more precise to describe endonuclease as a dimer of dimers. Subunits A1 and B1, related by a true two-fold axis, are closer in space than subunits A2 and B2 which are related by the same two-fold axis. Subunits A1 and A2 or B1 and B2 are related by two non-orthogonal pseudo two-fold axes plus a 1.6 Å translation along the rotation axis. The resulting dimers, A1-A2 and B1-B2, lie on two opposite sides of the parallelogram. Each of the pseudo two-fold axes relating monomers within the A1-A2 and B1-B2 dimers are at an angle of 32° with respect to the two-fold axis which relates the two dimers. In addition the three two-fold axes do not intersect. Each pseudo two-fold axis is shifted ~10 Å in the opposite direction relative to the two-fold axis that relates the two dimers. The Cα's of the A1-A2 dimer can be superimposed with a root-mean-square-difference (RMSD) of 0.47 Å with those of the B1-B2 dimer. There are larger structural differences observed between the subunits within each dimer, their Ca carbons have a RMSD of 1.2 Å. The largest deviation is observed on the loop L7 which harbors the conserved His^{125} residue. The asymmetry observed in the organization of the four identical subunits suggests non-equivalent roles for the two subunits in a dimer.

FIGURE III-4. The tetrameric *M. jannaschii* endonuclease and subunit-subunit interactions. **A**. Each subunit in the tetramer is given a distinct color and a label. The tetramer is viewed along the true two-fold axis relating the A1-A2 and B1-B2 dimers. **B**. The isologous dimer formed between subunits A1 and A2 viewed approximately perpendicular to the pseudo two-fold axis between them. The main chain hydrogen bonds formed between β9 and β9' and between loops L8 and L8' are drawn as black lines. Side chains of the hydrophobic residues enclosed at the dimer interface are shown as blue ball-and-stick models.

C. The heterologous interaction between subunits A1 and B2 (or B1 and A2) via the acidic loops L10 and L8. An electrostatically positive cleft (blue mesh surface) on subunit A1 (or B1) formed between the NH_2- and COOH-domains is complementary to the protruding L10 and L8 (both in ball-and-stick model) from subunit B2 (or A2). Secondary elements that form and surround the cleft are shown in yellow ribbon and labeled. Negatively charged residue side chains are colored red for L10 and L8 residues. Backbones and hydrophobic residues are colored green. Amino acid from L10 and L8 buried in the cleft are numbered. This figure was generated using the program GRASP (38).

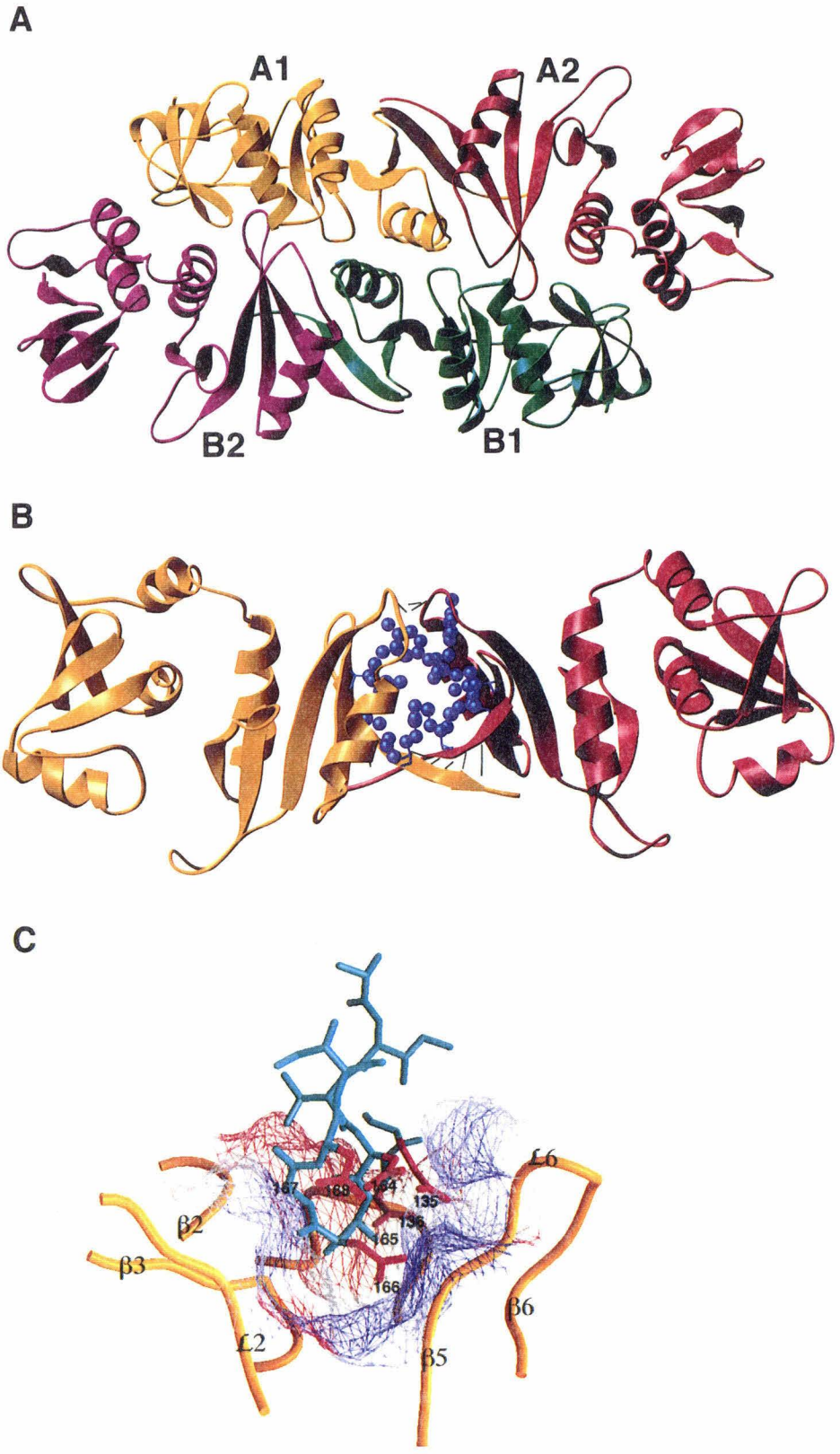


Figure III-4

Subunit Interfaces

The endonuclease dimers A1-A2 and B1-B2 associate isologously in a tail-to-tail fashion (Figure III-4B). The dimerization interface buries extensive surface area (2422 $Å^2$) (17) and is formed by the interaction of the central COOH-terminal β-sheets from each monomer via main chain hydrogen bonding and hydrophobic interactions. The pseudo two-fold axis that relates the two monomers lies between the two β9 strands and between the two L8 loops in each subunit. The NH_2 portion of β9 (169-172) is hydrogen bonded to β8 via the main chain atoms. The COOH-terminal half of β9 (residues 172-177) bends away from the plane of the pleated central β sheet to form main chain hydrogen bonds with the symmetry related residues of β9 in the isologous subunit, leading to a two-stranded β-sheet spanning the subunit boundary. The symmetrically related L8 loops form another layer on top of the two-stranded β9 sheet, and together with the β9 sheet enclose a hydrophobic core at the intersubunit surface. Residues in the hydrophobic core include Phe^{139}, Leu^{141}, Leu^{144}, Gly^{146}, Val^{148}, Leu^{158}, Ile^{160}, Met^{174}, and Tyr^{176}. These hydrophobic residues are important for stabilizing the dimer interface. The buried surface area at the dimer interface is 74 percent hydrophobic (18), slightly higher than the average 65 percent observed for 18 oligomeric proteins (19).

The interface between the two dimers is between subunits A1 and B2 and subunits B1 and A2, each pair contributing 40 percent of 5420 $Å^2$ total buried surface area (17). The remaining 20 percent buried surface area can be accounted for by the interface between subunits A1 and B1. The subunits A2 and B2 are not in contact. Interestingly, while the buried non-polar surface area at the tetramer interface is 64 percent (18), comparable to the observed average value of 65 percent (19), some strikingly close

contacts between polar residues of opposite charges are observed between the two heterologously associated subunits (Figure III-4C). A positively charged cleft with a dimension of 9 Å wide, 12 Å long and 14 Å deep is formed between the two domains of subunit A1 (or B1). It accommodates a protruding negative surface on the subunit B2 (or A2) formed primarily by loop L10 and partially by L8. Residues from L10 and L8 that are deeply buried by this inter-domain cleft are mostly acidic and are comprised of Asp^{164}-Ala^{165}-Asp^{166}-Gly^{167}-Asp^{168} and Glu^{135}-Asp^{136}. Several positively charged residues (Lys^{88}, Lys^{103}, Lys^{107}, Arg^{113}) line the cleft and interact electrostatically with the acidic loop residues. The interactions via loops L10 and L8 to the inter-domain cleft contribute nearly all the buried surface area at the tetramerization interface. In addition, three completely buried water molecules at the base of the cleft form hydrogen bonds with polar residues from both subunits, indicating solvent accessibility of this cleft prior to tetramer association. Most interestingly, one of the water molecules at the base of the cleft is part of the hydrogen bond network formed among the conserved Glu^{49}, Tyr^{92}, and Arg^{96} residues that stabilizes the interaction between the NH_2- and the COOH-domains (see above).

The isologous dimer interaction formed via the C-terminal interface is expected to be more stable than the heterologous polar interaction mediated by L10 and L8, suggesting that the tetramer assembles through dimer intermediates. Thus the structure appears to be a dimer of two dimers. This interpretation is in agreement with the results obtained from protein cross-linking (16). The prevalent cross-linked species seen in this study were dimers and tetramers. It is plausible that the observed dimers in solution resemble isologous dimers A1-A2 or B1-B2 in the crystal structure.

Active Sites

The tRNA splicing endonucleases cleave pre-tRNA leaving 5'-OH and 2',3' cyclic phosphate termini. This is the same specificity as in ribonucleases such as RNaseA and T1. The RNaseA mechanism has been studied extensively and a first guess would be that the endonuclease mechanism should be chemically similar. Figure III-5 shows the accepted reaction pathway and mechanism for RNaseA. The two step reaction is acid-base catalyzed. A general base abstracts a proton from the 2'-OH of ribose leading to an in-line attack on the adjacent phosphodiester bond and the formation of a pentacovalent intermediate (20). The general acid protonates the 5' leaving group leading to the 2',3' cyclic phosphate product. In a second step a proton is abstracted from a water molecule, the hydroxyl ion attacks the 2',3 cyclic phosphate which is then hydrolyzed to give the 3' phosphate. Figure III-5 also diagrams the accepted structure of the transition state. His^{12} is the general base in the first step, His^{119} the general acid and the pentacovalent transition state is stabilized by Lys^{41}.

In the splicing endonuclease family there is a conserved histidine residue. There is strong evidence that this histidine is part of the active site. Changing this histidine at position 242 in Sen34 to alanine was shown to markedly decrease 3' splice site cleavage (9) . Daniels (personal communication) has shown that the equivalent histidine mutant in the *H. volcanii* enzyme greatly impairs cleavage activity as have Lykke-Anderson and Garrett (16) for a His^{125} to Ala^{125} mutant in the *M. jannaschii* endonuclease. For reasons outlined below we propose that His^{125} together with the absolutely conserved

FIGURE III-5. RNaseA reaction mechanism and the accepted transition state. Adapted from reference 20.

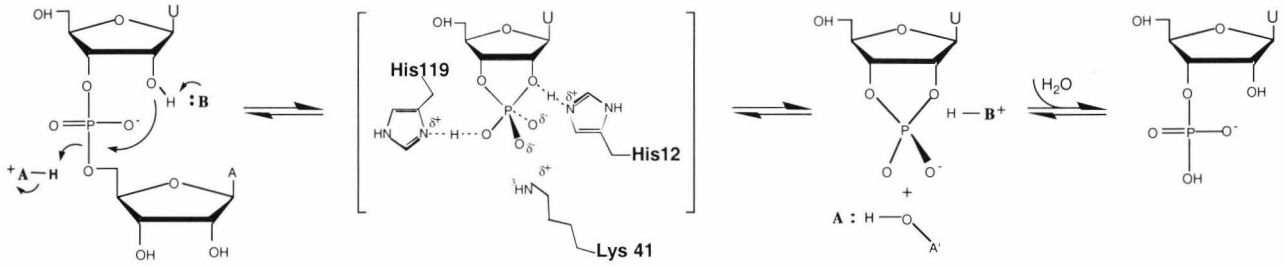


Figure III-5

FIGURE III-6. The spatial arrangement of three strictly conserved residues, Tyr^{115}, His^{125}, Lys^{156} in relation to the density attributed to a bound negative ion. The NCS-averaged electron density at 3.0 Å, shown around a modeled sulfate ion, is contoured at 3σ.

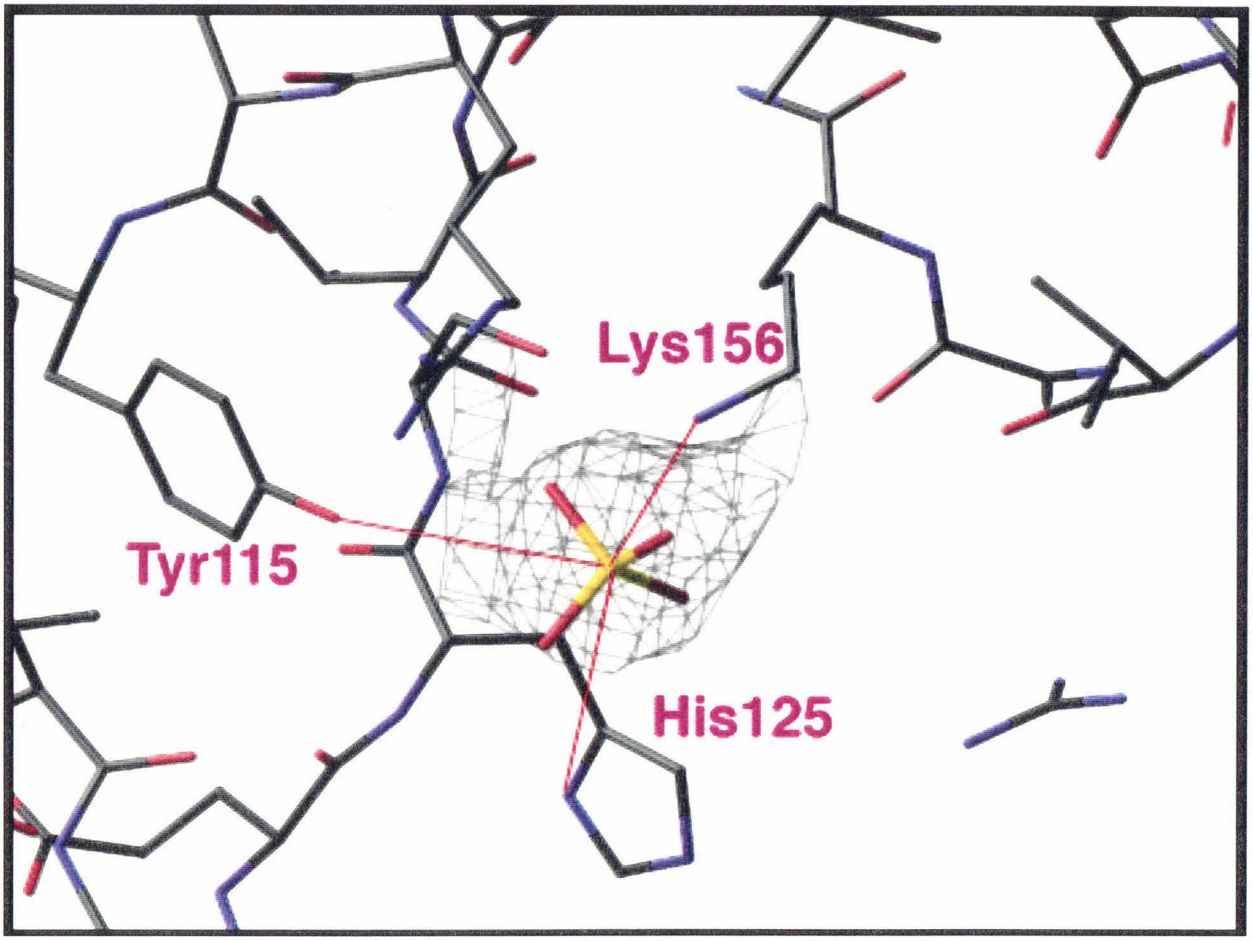


Figure III-6

Tyr^{115} and Lys^{156} constitute a catalytic triad for acid-base catalysis carried out by the tRNA endonucleases.

The His^{125} in *M. jannaschii* endonuclease is located on the boundary between L7 and β7. Even though His^{125} is well ordered in the structure in all four subunits, the rest of the loop L7 (residues 119-124) has weak electron density suggesting flexibility of this loop. The average B factor of L7 is 66 $Å^2$ in comparison with the average value of 44 $Å^2$ for the protein as a whole. Several striking structural features are seen in this region. There are primarily positively charged or polar residues in the vicinity of His^{125}, including two other strictly conserved residues, Tyr^{115} and Lys^{156}. The side chains of His^{125}, Tyr^{115} and Lys^{156} are within 4-7 Å of each other and form a His-Tyr-Lys triad (5.5 Å between His^{125} and Lys^{156}, 6.7 Å between Lys^{156} and Tyr^{115}, 6.1 Å between Tyr^{156} and His^{125}) (Figure III-6). A distinct electron density feature (> 5σ in the averaged F_O map) is found in the middle of the triad. This density cannot be attributed to the polypeptide chain but is compatible with the size of a sulfate ion (Figure III-6). A metal ion is ruled out due to the positive environment of this site and we have currently modeled this site with a sulfate ion which is present in the crystal stabilization buffer, although it could be a chloride ion also present in crystallization solutions.

The side chains of the His-Tyr-Lys triad in *M. jannaschii* endonuclease can be spatially aligned with those of the catalytic triad in RNaseA (Figure III-7). The orientation of this alignment results in an overlap of His^{125} with His^{12} of RNase A, Tyr^{115} with His^{119}, and Lys^{156} with Lys^{41} (Figure III-8). This leads to a model in which the *M. jannaschii* endonuclease employs His^{125} as the general base, Tyr^{115} as the general acid

FIGURE III-7. Spatial arrangement of the putative catalytic triad of *M. jannaschii* (red residues) in comparison with that of those in RNaseA (blue residues). The three catalytic residues from RNaseA were manually brought to best overlap with those in *M. jannaschii* endonuclease using the program O (32) . *M. jannaschii* endonuclease is represented by orange ribbon and RNaseA (coordinates are from protein data bank code 5rsa) is shown in silver ribbon.

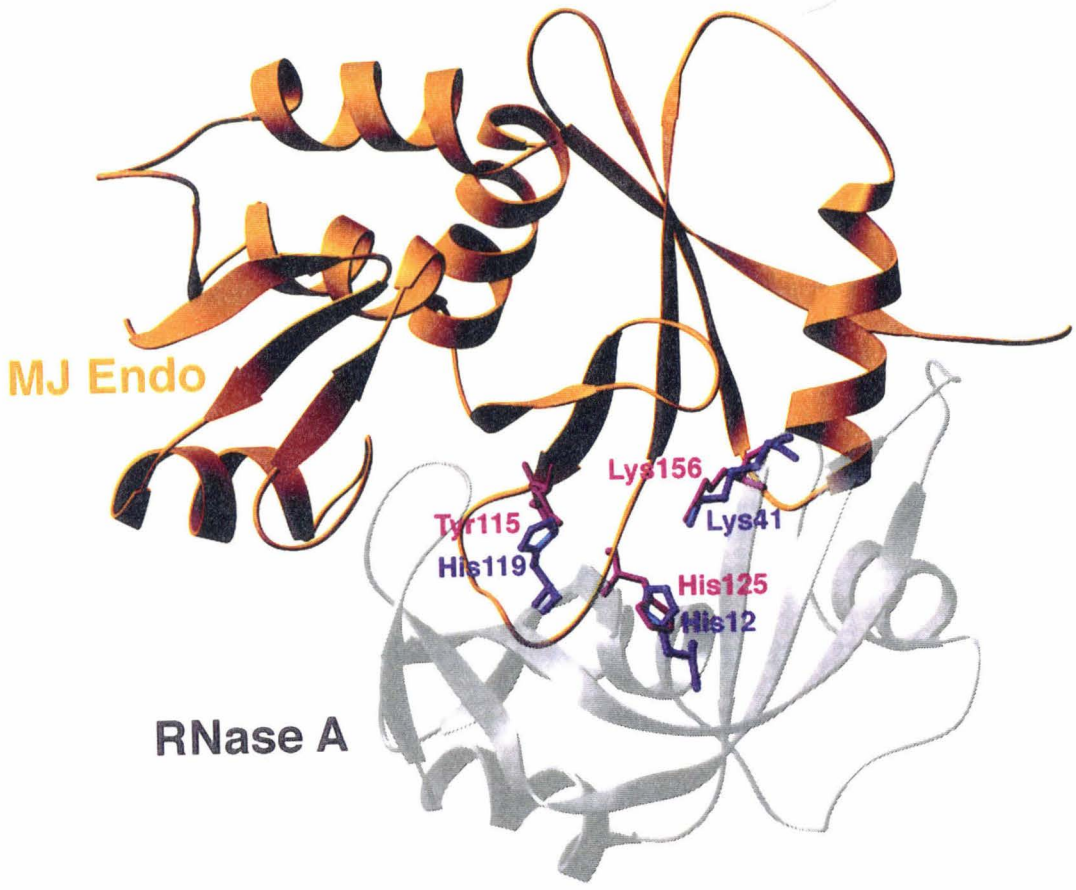


Figure III-7

and Lys^{156} functions to stabilize the transition state. Consistent with this model, single mutations in each of these three residues lead to impaired RNA cleavage activity (5).

Two out of the three putative active site residues (Tyr^{115} and His^{125}) are in the long loop L7. This loop is flexible, as commonly seen in proteins that bind to the phosphodiester backbone of nucleic acid. The intron-exon junctions in archaeal pre-tRNA substrates are all located in single stranded regions. The flexible L7 loop could potentially participate in correct positioning of the scissile phosphate for cleavage. Interestingly, the Gly^{292} to Glu mutation in the *sen2-3* mutant in yeast occurs within the aligned sequences of this loop and is impaired in cleavage of the 5' splice site (9).

The *M. jannaschii* endonuclease tetramer contains four active sites but the following reasons make it likely that only two of these participate in cleavage of the pre-tRNA substrate. The yeast endonuclease tetramer contains two functionally independent active sites for cleavage of the 5' and 3' splice sites, Sen2 and Sen34, respectively. The *H. volcanii* endonuclease is a homodimer and thus contains only two functional active sites. The two intron-exon junctions of the archaeal tRNA substrate are related by a pseudo-two-fold symmetry axis in the consensus bulge-helix-bulge substrate. It is therefore expected that the *M. jannaschii* endonuclease contains two functional active sites related by two-fold symmetry. *M jannaschii* endonuclease is protected to a maximum of 50% from proteolytic cleavage upon RNA substrate binding, consistent with the notion that this enzyme contains only two functional active sites (16). There are four pairs of active sites in the *M. jannaschii* endonuclease that are related by two-fold symmetry. These are found in subunits A1 and B1, A2 and B2, A1 and A2, and B1 and B2 (Figure III-8). We favor the choice of subunits A1 and B1 whose His-Tyr-Lys triads are separated by ~27 Å. The location of amino acid residues in subunits A1 and B1

protected from proteolytic cleavage upon RNA substrate binding (16) are shown in Figure III-8. The protected residues cluster around the two proposed catalytic sites and more significantly fall on a path between the two sites. This region of the enzyme, formed by L9 and the COOH portions of α5 and β9 in A1 and B1, has a positive electrostatic potential suitable for binding of the phosphodiester backbone of the pre-tRNA substrate. By contrast, the distance between the His-Tyr-Lys triads of subunits A2 and B2 is 75 Å and there is a less appealing distribution of substrate protected residues between this pair of active sites (data not shown).

We attempted to dock a model of the substrate with the tetramer, keeping in mind the possibility that both the protein and the RNA substrate may undergo significant conformational changes upon association. The consensus substrate has two unpaired loops of three nucleotides separated by a helix of four basepairs (Figure III-1A). A model for the substrate was constructed from a related structure solved by NMR, that of HIV TAR RNA (21, 22). The TAR RNA structure is a stem loop structure with a bulge of three bases in the stem (Figure III-9A). Calnan et al. (23) showed that the binding of Tat to TAR is primarily via one arginine residue. Puglisi et al. (21) and later Aboul-ela et al. (22) solved the structures of both free RNA and RNA bound to arginamide. The two structures are different. In the free RNA the bulged bases are stacked within the helix. In the arginamide complex one of the bulged bases forms a base triple with an adjacent basepair and the other two bases are rotated out of the helix. The substrate model was constructed by rotating a section of Tar RNA 180° so as to create a pseudo-symmetric structure of two three base bulges separated by a four base pair helix conforming to the archaeal endonuclease consensus substrate (Figure III-9B). Interestingly, it is the model

FIGURE III-8. Electrostatic potential surface of the *M. jannaschii* tetramer generated by the program GRASP (38) and the location of proteolytically protected residues by an RNA substrate analog on subunits A1 and B1. The surface viewed along the strict two-fold axis relating the two dimers, but in opposite direction of that in Figure III-4A. The catalytic His^{125} for all four subunits, Tyr^{115} and Lys^{156} for subunits A1 and B1, are shown in red stick models. The locations of residues for subunits A1 and B1 protected from proteolytic cleavage by an RNA substrate analog as found by Lykke-Adnersen and Garrett (16) are indicated by yellow spheres.

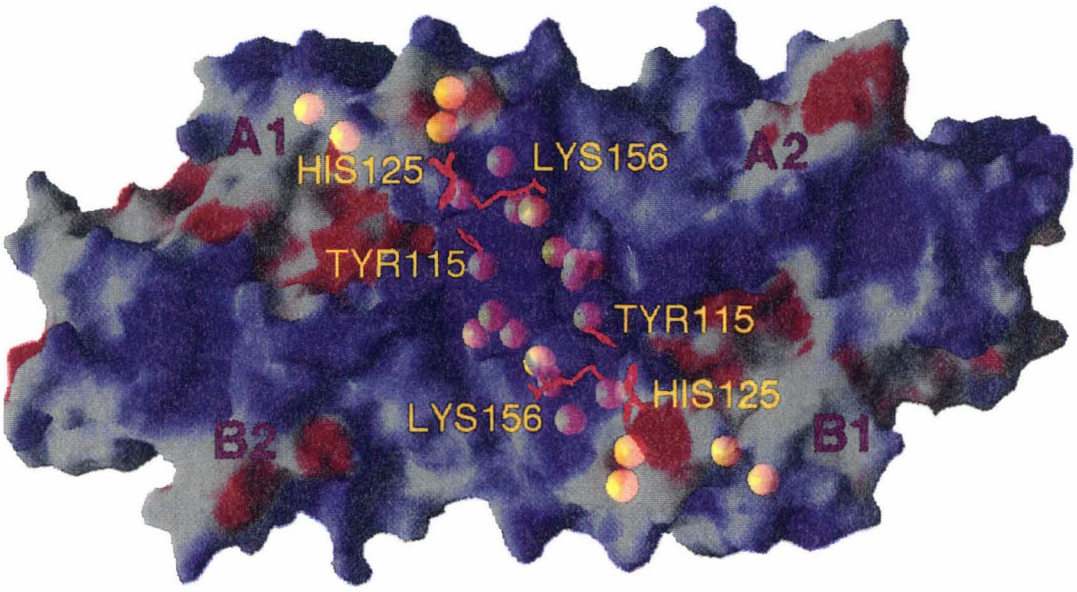


Figure III-8

FIGURE III-9. A modeled endonuclease-RNA substrate complex.

A. Sequence and secondary structure of HIV TAR RNA (PDB code 1arj) used to model *M. jannaschii* pre-tRNA minimum substrate (B).

B. A two-fold symmetry operation as indicated by the filled ellipse is applied to coordinates of the TAR RNA NMR structure (light gray region) to obtain a structure model for the second three-base bulge and the acceptor stem (dark gray region). The corresponding splice sites on the final model are indicated by arrows.

C. Modeled complex of *M. jannaschii* endonuclease with RNA substrate obtained by manually aligning the phosphate backbone of the four base-paired helix with the positively charged surface between subunits A1 and B1. This view is perpendicular to that of Figure III-9. The two putative active subunits A1 and B1 are shown in orange and green ribbons and catalytic triad residues are shown in magenta. The phosphodiester bonds at the splice sites on RNA are depicted in blue and fit nearly perfectly with the putative sulfate density shown by blue dots. The distance from the center of the catalytic triad to the surface where RNA binds is around 7Å.

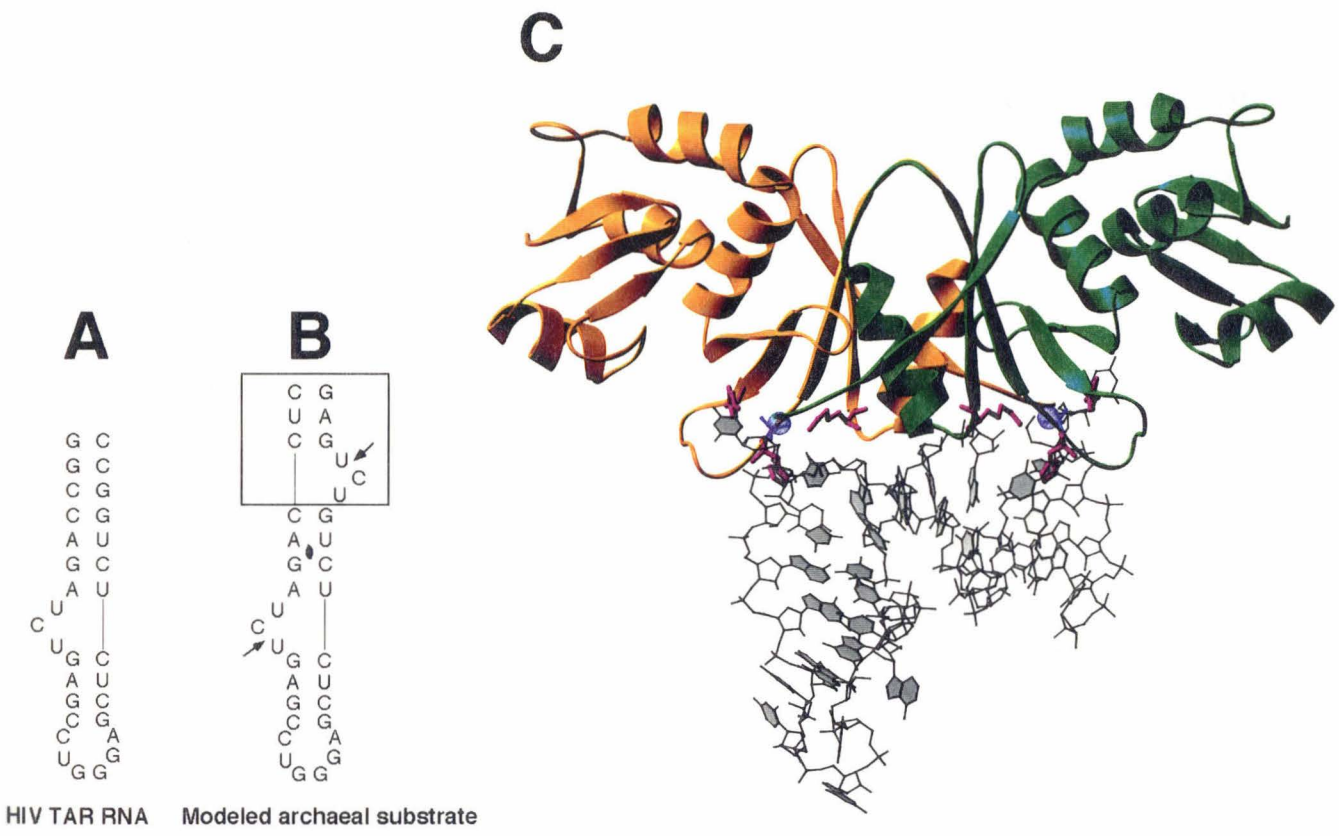


Figure III-9

derived from the arginamide complex of TAR RNA that best docks with the *M. jannaschii* endonuclease tetramer. Figure III-9C shows that the two phosphodiester bonds that are to be cleaved fit exactly into the active sites and superimpose with the putative sulfate ion found there. The four base-pair helix separating the two bulges docks on the basic side of the enzyme and could interact with a number of the residues that Lykke-Andersen and Garrett (16) found to be protected from protease cleavage upon binding (see Figure III-8) of a consensus substrate analog. Though this is a speculative exercise, it suggests that the splice sites must be bulged in order to fit into the ~7Å deep active site cavity (Figure III-9C). This enzyme must be specific for its substrate. The requirement for bulged bases, precisely separated by four base pairs, would insure that the enzyme could not cleave helical substrates. Work by Thompson and Daniels (8) has shown that the elimination of the bulges at either splice site of an archaeal RNA substrate blocks cleavage by *H. volcanii* endonuclease.

Implications for the Structure of Other Splicing Endonucleases

As the *H. volcanii* endonuclease dimer and the *M. jannaschii* tetramer recognize the same consensus substrate, their active sites must be arrayed similarly in space. This was hard to understand until Lykke-Andersen and Garrett pointed out (16) that the *H. volcanii* endonuclease monomer is in fact a tandem repeat of the consensus sequence of the endonuclease gene family. The two repeats are more similar to the *M. jannaschii* endonuclease sequence than to each other. In particular the N-terminal repeat does not contain a convincing NH_2-domain and lacks two of the three putative active site residues (Figure III-10B). It does, however, contain many of the features of the COOH-domain important for structural arrangement of the enzyme, in particular the loop L10 sequence.

FIGURE III-10. Structure models of endonuclease from *M. jannaschii* and *H. volcanii*.

A. Schematic drawing shows the *M. jannaschii* endonuclease sequence features followed by structural arrangement model of the endonuclease homotetramer. Several important structural features as discussed in the text are indicated: the loop L10, the C-terminus β9 strands as indicated by arrows, and the conserved catalytic residue His^{125} shown as a pentagon. The two putative active sides that participate in cleavage reaction are those in subunits labeled A1 and B1.

B. Schematic drawing shows the *H. volcanii* endonuclease sequence features followed by the proposed structural arrangement model. At top, a bar diagram shows the sequence features conserved between *H. volcanii* and *M. jannaschii* endonucleases. The N-terminal repeat lacks the catalytic residues but has kept L10 in order to mediate the dimer-dimer interaction. The C-terminal repeat contains all the sequence features as seen in *M. jannaschii* enzyme. At the bottom, a proposed structure of *H. volcanii* homodimer is schematized. Dashed lines represent the polypeptide chain connecting the C- and N-terminal repeats.

C. Proposed yeast endonuclease structural model. Conserved amino acid sequences near the C-termini of archaeal enzymes, M.jann. (*Methanococcus jannashcii*), H.v. Nt (*Haloferax volcanii* N-terminal repeat) and yeast Sen54 (Sc Sen54p) and Sen15 (Sc Sen15p) subunits are aligned. Important hydrophobic residues that stabilize the isologous C-terminus interaction observed between *M. jannaschii* subunits A1 and A2 or B1 and B2 are highlighted in green and circled on the structure models of heterodimers derived from the two-hybrid matrix analysis (9). Sen54 and Sen15 are aligned with the loop L10 sequences in *M. jannaschii* and *H. volcanii* and are highlighted in red.

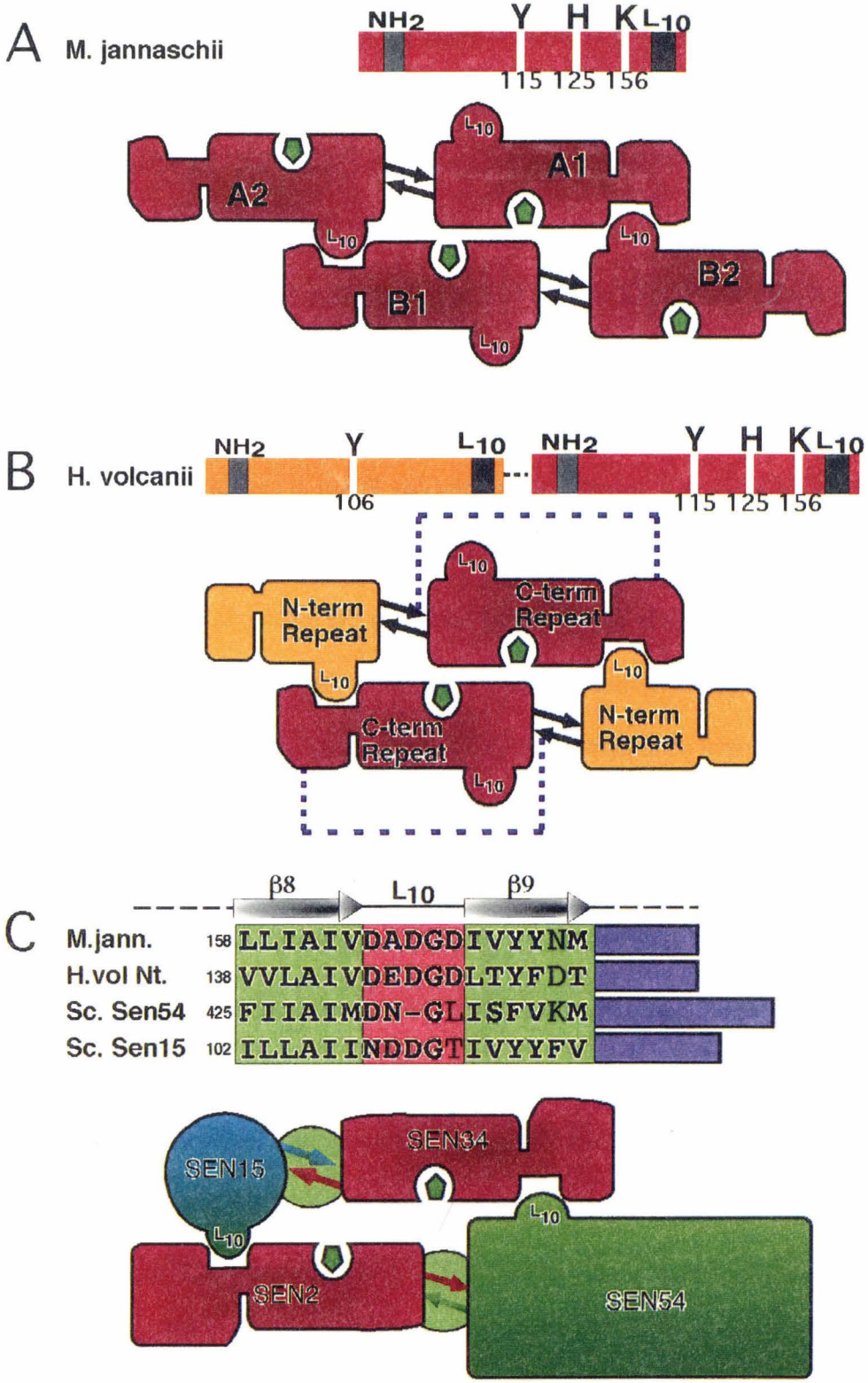


Figure III-10

We propose that the *H. volcanii* enzyme forms a pseudo-tetramer of two pseudo-dimers (Figure III-10B). The structure of the pseudo-dimer is predicted to contain a two stranded β9-β9' pleated sheet, an important structural feature of the *M. jannaschii* endonuclease dimer. The tandem repeat must be connected by a stretch of polypeptide (dotted line in Figure III-10B). Within each dimer of the *M. jannaschii* endonuclease structure, the N-terminus of subunit A2 (or B2) and the C-terminus of subunit A1 (or B1) is separated by 28Å. This would require a span of at least 10 amino acid residues of an extended polypeptide chain in order to connect the region between the N-terminal and the C-terminal tandem repeats in the *H.volcanii* endonuclease. The *H. volcanii* enzyme model contains only two of the proposed active site triads- in the C-terminal repeat of each pseudo-dimer- and these are proposed to occupy an identical spatial configuration to those in the A1 and B1 subunits of the *M. jannaschii* endonuclease. The pseudo-dimers are proposed to interact via the conserved L10 loop sequences found in the N-terminal repeat-equivalent to those in the A2 and B2 subunits in the *M. jannaschii* enzyme. This *H. volcanii* endonuclease structural model reveals that only two of the active sites are necessary but to array these in space correctly one must retain important features of both the isologous dimer interactions (β9-β9') and the dimer-dimer interactions mediated by the loop L10 found in the *M. jannashcii* crystal structure.

The yeast endonuclease contains two active site subunits, Sen2 and Sen34. The other two subunits did not initially appear to belong to the endonuclease gene family; however, Lykke-Andersen and Garrett (16) pointed out sequence similarity of both Sen15 and Sen54 near the C-termini to the *M. jannaschii* endonuclease. This region contains both an L10 loop sequence and the conserved hydrophobic residues near the C-terminus

as seen in β8 and β9 in the crystal structure of the *M. jannaschii* endonuclease. We propose that the strong Sen2-Sen54 and Sen34-Sen15 interactions seen in the two-hybrid experiment are mediated by C-terminal β9-β9'-like interactions (Figure III-10C). These dimers are associated to form the tetramer via the L10 loop sequences of Sen15 and Sen54 (Figure III-10C).

We conclude that the Archaea and the Eucarya have inherited from a common ancestor an endonuclease active site and the means to array two sites in a precise and conserved spacial orientation. This conclusion is supported by the results of Fabbri et al., (39), which demonstrate that both the eucaryal and the archaeal endonucleases can accurately cleave a universal substrate containing the BHB motif. The eucaryal enzyme appears to dispense with the more complex ruler mechanism for tRNA substrate recognition when it cleaves the universal substrate. Thus the precise positioning of two active sites in endonuclease appears to have been conserved in evolution. Subunits A1 and B1 comprise the active site core of all tRNA splicing endonucleases, and subunits A2 and B2 position the two active sites precisely in space. The eucaryal enzyme has evolved a distinct measuring mechanism for splice site recognition by specialization of the A2 and B2 subunits, and it has retained the ability to recognize and cleave the primitive consensus substrate.

Table 1. Statistics for MAD data collection and phase determination from an $Au(CN)_2$ derivatized endonuclease crystal, flash cooled to 100K. Crystals of recombinant endonuclease were grown at 22°C by the vapor diffusion method at pH 5.5-6.0 using KCl as a salting-in precipitant with a starting concentration of 1M. Crystals were soaked for 3 days in a stabilization/cryoprotection solution containing $Au(CN)_2$ (20mM Na-cacodylate,

pH 6.0, 20mM $(NH_4)_2SO_4$, 25% glucose, and 2mM $Au(CN)_2$) prior to data collection. Crystals belong to the space group $P2_12_12_1$, with unit cell dimensions a = 61.8 Å, b = 79.8 Å, and c = 192.8 Å. Three wavelengths near the L_{III} absorption edge of Au were collected from a single crystal at National Synchrotron Light Source beamline X4A using Fuji image plates: λ_1, the absorption edge corresponding to a minimum of the real part (f') of the anomalous scattering factor for Au; λ_2, the absorption peak corresponding to a maximum of the imaginary part (f'') of the scattering factor for Au; and λ_3, a high energy remote to maximize f'. The crystal was oriented with its c^* axis inclined slightly from the spindle axis (~16 degrees) to avoid a blind region. The a^* axis was placed initially within the plane of the spindle and the beam direction using the STRATEGY option in the program MOSFLM (version 5.4) (24). A total of 180 consecutive 2° images were collected with every 8° region repeated for all three wavelengths. All data were processed using the program DENZO and scaled with SCALEPACK (25). A previously collected data set at 1.5418 Å on a crystal soaked with the same $Au(CN)_2$ solutions was included as the fourth wavelength. This data set was measured with CuKα X-rays generated by a Rigaku RU200 rotating anode using a SIEMENS multiwire detector and was indexed and integrated using the program XDS (26) and reduced and scaled with ROTAVATA and AGROVATA in CCP4 (27). All reflection data were then brought to the same scale as those of wavelength 1 using the scaling program SCALEIT followed by the Kraut scaling routine FHSCAL (26) as implemented in CCP4 (27). The wavelength 1 (edge) has minimum value of f' for Au and this data set was taken as the "native" data with intrinsic anomalous signals. The data sets collected at wavelengths 2 (peak) and 3 (remote1) were treated as individual "derivative" date sets. Anomalous signals are included in the data set

of wavelength 2 where the X-ray absorption of Au is expected to be maximum. Due to the absence of a sharp white line at the Au L_{III} edge, the observed diffraction ratio at wavelength 2 is small, 0.061 (Table 1), as compared to the error signal ratio of 0.036 computed from centric reflections. The data set at wavelength 4 (remote 2) was taken at different time and from a different crystal, and was included as another "derivative" as it is near the maximum f' value. Two gold atoms in each asymmetric unit could be clearly located by both dispersive difference (λ_3 or λ_4 with respect to λ_1) and anomalous (λ_2) Patterson functions using the program HASSP (28). Two additional sites were later located by difference Fourier. We treated MAD data as a special case of the multiple isomorphous replacement with the inclusion of anomalous signals (29). Initial phases were computed at 3.0 Å using the program MLPHARE as implemented in the CCP4 program suite, which uses a maximum-likelihood algorithm to refine heavy atom parameters (30). We exploited the fact that there are four non-crystallographically related monomers per asymmetric unit and performed four-fold averaging in parallel with solvent flattening and histogram mapping using the program DM (31). The subsequent electron density map was markedly improved and the polypeptide chain could be unambiguously traced. A model representing the monomeric endonuclease containing all but the first 25 amino acids (the 8 Histidine residues, the 9 FLAG-tag amino acids and the first 8 residues in the wild type endonuclease sequence) was built into the electron density map using the interactive graphics program O (32) . The assignment of side-chains for the *M. jannaschii* endonuclease polypeptite was confirmed by a difference Fourier map between the $Au(CN)_2$ derivatized protein and a selenomethionine substituted protein. The final model, which includes the tetrameric endonuclease of residues 9-179, 53 water oxygen atoms, 2

partially occupied SO_4^- atoms, and 4 gold atoms, was refined to 2.3 Å against the observed anomalous data of wavelength 3 using X-PLOR version 3.8 (33). Non-crystallographic restraints were applied to atoms between dimers A1-A2 and B1-B2 excluding those from loops L3 and L7. The structure was assessed with the structure verification programs PROCHECK (34) and VERIFY3D (35) and was shown to be consistent with a correct structure. There are more than 90 percent residues in the most favored and no residues in disallowed regions of a Ramachandran plot (34).

table 1 continued

Data Statistics*

Item	λ_1 (edge)	λ_2 (peak)	λ_3 (remote 1)	λ_4 (remote 2)
Wavelength (Å)	1.03997	1.03795	1.02088	1.5418
Resolution (Å) (last shell)	2.28 (2.36-2.28)	2.28 (2.36-2.28)	2.28 (2.36-2.28)	3.0 (3.16-3.0)
measured reflections	165549	161691	164948	50259
Unique reflections	75778	75310	76385	16679
R_{merge}†	0.052 (0.153)	0.051 (0.151)	0.054 (0.167)	0.085 (0.359)
$\langle I \rangle / \langle \sigma(I) \rangle$	11.3 (2.6)	10.6 (2.6)	10.3 (2.4)	8.7 (2.2)
Completeness (%)	89.2 (62.2)	89.0 (62.0)	90.4 (65.4)	83.3 (42.6)
f' (electrons)	-20.21	-15.21	-10.76	-5.10
f''(electrons)	10.20	10.16	9.87	7.30
Dispersive ratios (30-3Å)‡	-	0.059	0.060	0.165
Anomalous ratio (30-3Å)‡	0.061 [0.041]	0.061 [0.036]	0.062 [0.036]	-

Phasing and Symmetry

Averaging Statistics

Resolution range (Å)	50-3.0
Phasing power for acentric reflections	
$\lambda_2 \rightarrow \lambda_1$	0.11
$\lambda_3 \rightarrow \lambda_1$	0.24
$\lambda_4 \rightarrow \lambda_1$	1.02
Figure of merit	
Initial	0.22
After solvent flattening and averaging	0.78
Average correlation coefficient of the four-fold NCS operations	
Initial	0.369
After solvent flattening and averaging	0.861

Refinement Statistics[§]

Resolution range (Å)	50-2.28
R value	0.204
R_{free} value (10% reflections)	0.268
non-hydrogen protein atoms	5584
water molecules	53
SO_4^- atoms	9
Au atoms	4
root-mean-square-deviations from ideal geometry (36)	
Bond (Å)	0.011
Angle (°)	1.586
Improper Dihedral (°)	0.733
Torsion (°)	27.195

table 1 continued

*Numbers in parentheses correspond to those in the last resolution shell. [†]R_{merge} indicates agreement of individual reflection measurements over the set of unique averaged reflections. $R_{merge} = \Re_{h,i} |I(h)_i - \langle I(h)\rangle| / \Re_{h,i} I(h)_i$ where h is the Miller index, i indicates individually observed reflections, and $\langle I(h)\rangle$ is the mean of all i reflections of the Miller index h, Bijvoet mates are treated as independent reflections when computing R_{merge}. [‡]Diffraction ratios are defined as $\langle \bullet |F|^2\rangle^{1/2} / \langle |F|^2\rangle^{1/2}$ where $\bullet |F|^2$ are taken between the current and the reference wavelength (λ_1) for dispersive ratios and between matched Bijvoet pairs for anomalous ratios. Ratios for centric reflections at each wavelength are in shown in square parentheses. [§]Bulk solvent correction is included in the final cycles of refinement.

References

1. D. Biniszkiewicz, E. Cesnaviciene, D. A. Shub, *EMBO J.* **13**, 4629-4635 (1994); B. Reinholdhurek, D. A. Shub, *Nature* **357**, 173-176 (1992).

2. H. G. Belford, S. K. Westaway, J. N. Abelson, C. L. Greer, *J. Biol. Chem.* **268**, 2444-2450 (1993); S. K. Westaway, H. G. Belford, B. L. Apostol, J. N. Abelson, C. L. Greer, *J. Biol. Chem.* **268**, 2435-2443 (1993); G. M. Culver, S. M. Mccraith, S. A. Consaul, D. R. Stanford, E. M. Phizicky, *J. Biol. Chem.* **272**, 13203-13210 (1997).

3. G. M. Culver, et al., *Science* **261**, 206-208 (1993); S. M. McCraith, E. M. Phizicky, *J. Biol. Chem.* **266**, 11986-11992 (1991).

4. J. R. Palmer, T. Baltrus, J. N. Reeve, C. J. Daniels, *Biochimica Et Biophysica Acta* **1132**, 315-318 (1992); J. Kjems, J. Jensen, T. Olesen, R. A. Garrett, *Canadian J. Microbiol.* **35**, 210-214 (1989).

5. C. R. Trotta, unpublished results.

6. I. Gomes, R. Gupta, *Biochem. and Biophy. Rev.* **237**, 588-594 (1997).

7. V. M. Reyes, J. N. Abelson, *Cell* **55**, 719-730 (1988).

8. L. D. Thompson, C. J. Daniels, *J. Biol. Chem.* **265**, 18104-18111 (1990).

9. C. R. Trotta, et al., *Cell* **89**, 849-858 (1997).

10. K. Kleman-Leyer, D. A. Armbruster, C. J. Daniels, *Cell* **89**, 839-848 (1997).

11. C. K. Ho, R. Rauhut, U. Vijayraghavan, J. N. Abelson, *EMBO J.* **9**, 1245-1252 (1990).

12. C. J. Bult, et al., *Sci.* **273**, 1058-1073 (1996).

13. The recombinant endonuclease was purified by a single Ni-NTA (Qiagen) affinity chromotography step following a heat-denaturation step to high purity.

14. W. A. Hendrickson, *Sci.* **254**, 51 (1991).

15. C. Chothia, A. V. Finkelstein, *Annu. Rev. Biochem.* **59**, 1007-1039 (1990).

16. J. Lykke-Andersen, R. A. Garrett, *EMBO J.***16**, 6290-6300 (1997).

17. B. Lee, and F. M. Richards, *J. Mol. Biol.* **55**, 379-400 (1971).

18. Computed using the program Biograf (Molecular Simulations Incorporated, San Diego, California).

19. J. Janin, S. Miller, C. Chothia, *J Mol Biol* **204**, 155-64 (1988).

20. C. Walsh, *Enzymatic reaction mechanisms* (W.H. Freeman and Company, San Francisco, 1979).

21. J. D. Puglisi, L. Chen, A. D. Frankel, J. R. Williamson, *Proc. Nat. Acad. Sci.* **90**, 3680-3684 (1993).

22. F. Aboul-ela, J. Karn, G. Varani, *Nuc. Acids Res.* **24**, 3974-3981 (1996).

23. B. J. Calnan, B. Tidor, S. Biancalana, D. Hudson, A. D. Frankel, *Sci.* **252**, 1167-1171 (1991).

24. A. G. W. Leslie, *CCP4 AND ESF-EACMB Newsletter on Protein Crystallography No. 26*, Daresbury Laboratory, Warrington, U. K. (1992).

25. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend: "Data Collection and Processing," 29-30 January 1993* L. I. Sawyer, N. Bailey, S., Ed. (SERC Daresbury Laboratory, England) pp. 56-62; W. Minor, XDISPLAYF program, Purdue University (1993).

26. W. Kabsch, *J. App. Crystallogr.* **21**, 916 (1988).

27. Collaborative Computational Project, Number 4. "The CCP4 Suite: Programs for Protein Crystallography," *Acta Crystallogr. D* **50**, 760-763 (1994).

28. T. C. Terwilliger, S. H. Kim, D. E. Eisenberg, *Acta Crystallogr. A* **43**, 34-38 (1987).

29. V. Ramakrishnan, V. Biou, *Methods Enzymol* **276**, 538-57 (1997).

30. Z. Otwinowski, in *Isomorphous Replacement and Anomalous Scattering* W. E. Wolf, P. R. Leslie, A. G. W., Ed. (Science and Engineering Research Council, Daresbury Laboratory, Daresbury, U. K., 1991) pp. 80.

31. K. Cowtan, *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38 (1994).

32. T. A. Jones, J. Y. Zou, S. W. Cowan, Kjeldgaard, *Acta Crystallogr. A* **47**, 110-9 (1991).

33. A. T. Brünger, *X-PLOR Version 3.1, A system for crystallography and NMR*, (Yale Univ. Press, Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, 1996).

34. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. App. Crystallogr.* **26**, 283-291 (1993).

35. R. Luthy, J. U. Bowie, D. Eisenberg, *Nature* **356**, 83-5 (1992).

36. R. Engh, R. Huber, *Acta Crystallogr. A* **47**, 392-400 (1991).

37. M. Carson, *J. Mol. Graphics* **5**, 103-106 (1987).

38. A. Nicholls, K. A. Sharp and B. Honig, *Proteins: Structure, Function and Genetics* **11**, 282 (1991).

39. S. Fabbri et al., Science **280**, 284-286 (1998).

40. We thank D. Graham and C. Woese for *M.jannaschii* genomic DNA, P. Bjorkman and D. Rees for their helpful discussions and generosity in sharing their X-ray diffraction and computation equipment, C. M. Ogata for X-ray beam time allocation at NSLS X4A beamline, S. Diana for assistance in data collection, R. Story and M. Saks for critical review of the manuscript and other members of the Abelson laboratory for their advice and support. This work was supported by American Cancer Society grant NP802 and NIH NRSA individual postdoctoral fellowship F32 GM188930-01.

Chapter IV – Investigations into the Endonuclease Active Site and Enzyme Architecture

Introduction

With an understanding of the tRNA splicing endonuclease at the molecular level, we can now address two main lines of investigation. First, what is the reaction mechanism utilized by the tRNA splicing endonuclease for the cleavage reaction? Second, how does the eukaryotic enzyme recognize the pre-tRNA substrate in a mature domain dependent manner? The following chapter will present experimental designs which have been utilized to begin to answer these questions.

Dissection of the Interaction Domains of the Archaeal Endonuclease

As desribed in detail in Chapter III, the archaeal enzyme forms a homotetramer through the use of several evolutionary conserved interaction domains. Dimerization is mediated by an isologous interaction between two endonuclease monomers (see pg. 106). Dimerization of this dimeric unit leads to tetramer formation and correct orientation of the endonuclease active sites for the cleavage reaction. Tetramerization is accomplished through a heterologous interaction of loop L10 from one monomer with the groove formed between the N- and C- terminal domains of another monomer. This interaction is mediated by aspartic acid residues found on loop L10 contacting basic residues which line the interaction groove (Figure III-4C). In order to confirm that these interaction have functional relevance for the architecture of the active tRNA splicing endonuclease, we have undertaken a mutagenesis approach. As a first target we chose the aspartic acid residues found in loop L10. Aspartic acids 164 and 166 were mutated to lysine either

singly or as a double mutant. The proteins were expressed in *E.coli* with His/Flag tag and the mutant enzymes were purified. The purified endonuclease was tested for cleavage activity and the percent activity was qualitatively compared to wild type enzyme. For the mutant enzyme with single aspartic acid mutations, the enzyme was approximately 30% as active as wild type. Activity goes down to 2% of wild type for the double aspartic acid to lysine mutant. There are several possible explanations for the drastic effects of such mutants. The simplest and most consistent with our idea of how the endonuclease functions is that the change in charge interferes with the loop L10 and basic groove interaction. Thus the enzyme would be unable to tetramerize. A second possibility is that the mutation effects the overall architecture of the enzyme by causing misfolding. In order to distiguish between these two possiblities, we performed gel filtration on the mutant enzyme. Figure IV-1 shows that the double aspartic acid mutant fractionates as a dimer compared to the wild type tetramer. There is a small fraction of tetrameric enzyme, perhaps explaining why this mutant enzyme retains some activity. Thus this experiment represents a powerful demonstration that the interactions described in the crystal structure are indeed functionally relevant.

Trunctation mutants are also being utilized to probe the architecture of the enzyme. For example, the loop L10 sequence inserts into a basic groove formed between the N- and C-terminal domains of the monomer (see Figure III-4B). Truncation of the endonuclease which removes the N-terminal domain results in enzyme which is dead. However, upon addition of the N-terminal domain in *trans*, full activity can be restored (CRT unpub). This result again supports the importance of the loop L10 heterologous interaction for the architecture of the enzyme.

FIGURE IV-1. Gel filtration of DD mutant endonuclease and wild type endonuclease.

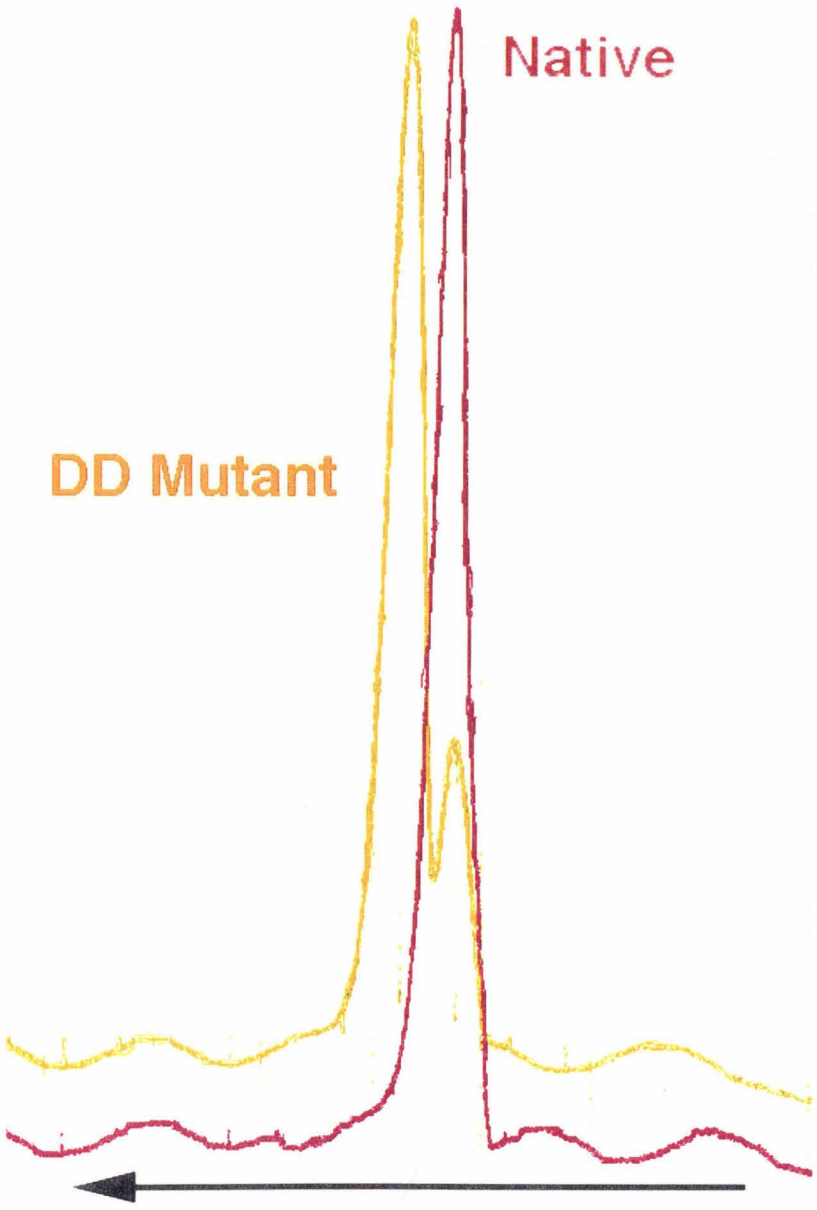


Figure IV-1

Conservation of the interaction domains in the eukaryotic enzyme makes a direct comparison of the effects of mutants on the archaeal and eukaryotic enzymes possible. The goal of such comparisons is to gain some insight into the architecture of the yeast endonuclease for which a crystal structure is surely a long way off due the heterotetrameric nature of the enzyme.

The Endonuclease Active Site

As discussed in Chapter III, the endonuclease active site, defined by the absolutely conserved histidine, appears to have a similar configuration as that of RNaseA. Thus the mechanism of cleavage by the tRNA splicing endonuclease is proposed to be acid-base catalysis. Experiments to determine the reaction mechanism have yet to be conducted in a rigorous fashion, but we have produced the reagents to allow such an examination. Again mutagenesis, of the crucial catalytic triad residues His125, Lys156 and Tyr115, is the tool of choice. Table IV-1 lists the mutants that have been produced. Also shown is the qualitative results of the cleavage efficiency of the mutants compared to the wild type enzyme. As can be seen from the table, both His125 mutants and Lys156 mutants exhibit a drastic effect on catalysis. Surprisingly, mutants in Tyr115 seem to have no effect on catalysis under the conditions which have been utilized for the assay. In the model of the active site triad, Tyr115 is proposed to be the general acid. Thus it is possible that under the reaction conditions utilized, we do not see an effect of the mutation due the ability of H_20 to substitute for the loss of the functional moiety normally presented by tyrosine. Careful kinetic analysis, including pH dependence of enzyme activity, should shed light on this discrepancy.

Table IV-1 List of endonuclease mutants and their qualitative splicing activity

Mutation	Splicing Phenotype
WT	+++
$H^{125}\rightarrow N$	++
$H^{125}\rightarrow A$	+/-
$K^{156}\rightarrow R$	+/-
$K^{156}\rightarrow Q$	+/-
$K^{156}\rightarrow A$	+/-
$Y^{115}\rightarrow A$	+++
$Y^{115}\rightarrow F$	+++
$Y^{115}\rightarrow H$	+++
$Y^{110}\rightarrow A$	+++
$Y^{115}Y^{110}\rightarrow FF$	+++
$D^{166}\rightarrow K$	+
$D^{166}D^{164}\rightarrow KK$	+/-

+++ is wild-type activity
++ is approximately half wild-type activity
+ is approximately 10% wild-type activity
+/- very little activity detected

The Molecular Ruler

Based on the model from Figure III-10C, we have constructed a working model for the interaction of the eukaryotic enzyme with the pre-tRNA substrate (Figure IV-2). We derived the pre-tRNA from modeling the yeast $tRNA^{Phe}$ with the archaeal substrate (see legend for Figure I-8). Using the known constraints for pre-tRNA cleavage, namely SEN2 contains the active site for the 5' splice site and SEN34, the active site for the 3' splice site, we docked the molecule in a similar fashion to the archaeal enzyme docking. Upon cloning of the subunits of the yeast endonuclease, we noted that SEN54 was a very basic protein and could thus interact with the phosphate backbone of the tRNA molecule. In our working model for interaction, we find that the SEN54 protein is positioned quite nicely for interaction with the mature domain of the tRNA. Of course this is all highly speculative, but it definitely leads to a testable model.

Since the domains for subunit interaction in SEN54 (and SEN15) were shown by homology to be present at the C-terminus, we propose that the N-terminal domain of the 54kD protein would embody the molecular ruler. Mutations in the N-terminal domain should not affect enzyme architecture, but should alter the ruler mechanism causing either abolishment or aberrant cleavage of yeast substrates by endonuclease containing mutant Sen54. Thus by employing a similar system as was used for purification of endonuclease containing mutant SEN34 H242A (see pg.71), enzyme decrepit for cleavage could be purified from yeast. The mutant enzyme could then be tested *in vitro* for cleavage ability. The first step of this screen has been conducted. Truncation mutants of SEN54 have been cloned and expressed in wild type haploid yeast. Interestingly, as was true for the catalytically decrepit endonuclease containing SEN34 H242A, expression of several

FIGURE IV-2. Working model for interaction of the yeast tRNA splicing endonuclease with pre-tRNA. The intron is shown in blue and the yeast endonuclease subunits are as labeled. The pre-tRNA is derived as described in the legend for Figure I-8.

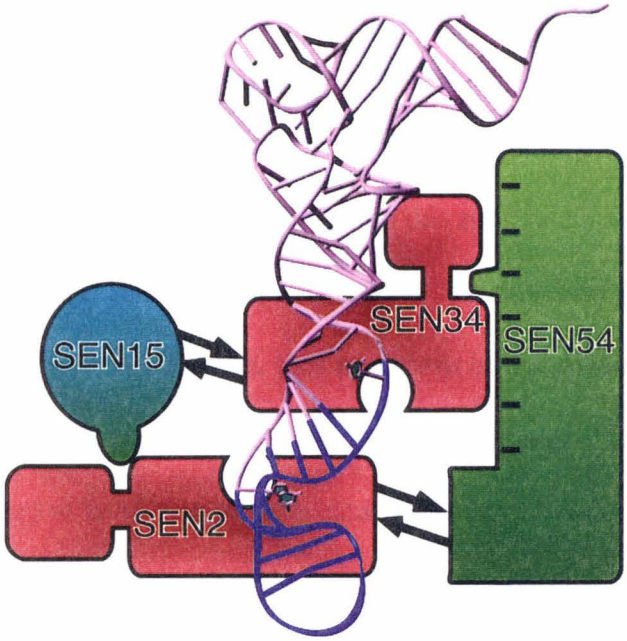


Figure IV-2

SEN54 truncation mutants result in a slow growth phenotype suggesting they are somehow affecting the normal processing events in a negative fashion.

Inability of the purified endonuclease mutants to cleave the yeast substrate is not conclusive proof that the ruler mechanism has been disabled. One possibility is that the mutations somehow affect the catalytic efficiency of the enyzme and not the substrate recognition. Careful kinetic analysis may allow an understanding of the defect. However, the model suggests that the mutant enzyme should contain two functional active sites and would thus be catalytically functional. In order to determine if this is the case, cleavage of pre-tRNA$^{Arch\text{-}Euka}$ could be used to test the mutant enzymes. It has been demonstrated by Tocchini-Valentini and co-workers (1998) that this substrate is cleaved by both the archaeal and eukaryotic enzymes. Upon cleaving this substrate, the eukaryotic enzyme does not utilize the ruler mechanism, thus mutations in the mature domain as well as deletions in the anticodon stem do not affect intron removal. This substrate represents a unique opportunity to be able to test the catalytic ability of endonuclease that contains the mutant SEN54 protein. The prediction is that the enzyme should possess the ability to cleave this substrate while being unable to cleave the yeast substrates.

Materials and Methods

Mutagenesis of *M. jannaschii* Endonuclease: Pet11a containing the wild type endonuclease gene was subjected to PCR utilizing KTLA under standard conditions (Barnes, 1996). PCR was carried out in two reactions. First the 5' and 3' half of the gene were amplified using a mutagenic oligonucleotide which consisted of one or two nucleotide changes flanked on either side by 10 nucleotides. The products were gel

purified on a 0.8% LMP agarose gel. 50ng of each product, as determined by gel quantitation, was added to a second PCR reaction. Oligonucleotides flanking the open reading frame were used and the full length gene was isolated by gel electrophoresis. The products were then cut with the original cloning sites, NdeI 5' and BamHI3', and were subcloned into pet11a for expression in *E. coli*. All mutants were sequenced to confirm the fidelity of the entire open reading frame at the Caltech Sequencing Facility.

Purification of Mutant Endonuclease: *E. coli* strain BL21 containing pet11a and pRI952 (a plasmid containing two tRNAs for the rare arginine and isoleucine bacterial codons) was inoculated into 3 mls of LB for overnight growth to saturation. 300mls of LB + 100ug ampicillin were then inoculated 1:100 with the culture and the cells were grown to OD600 = 0.6 – 0.8. IPTG was added to 1mM to induce expression from the plasmid and the cells were grown for 2 more hours. Cells were harvested by centrifugation and the pellet was resuspended in 8 mls. breaking buffer (50mM Phosphate Buffer pH6.7, 500mM KCl, 5% Glycerol). Cells were sonicated 8X10sec and the lysate was spun 10,000xg for 20 minutes. The supernatant was then heated to 65°C for 20 minutes, followed by a 10,000xg spin for 20 minutes. The supernatant was removed and 0.5 mls of Ni-NTA (Quiagen), equilibrated in breaking buffer, was added to the lysate. Incubation for 30 minutes at room temperature was followed by collection of the beads by filtration through a disposable column, and the beads were washed with breaking buffer. The beads were then washed with 10 vol. breaking buffer +15mM imidazole. Protein was eluted with breaking buffer +200mM imidazole. Enzyme was checked by SDS-PAGE gel electrophoresis and dialyzed against breaking buffer.

Appendix I

Minireview

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 273, No. 21, Issue of May 22, pp. 12685–12688, 1998

tRNA Splicing*

John Abelson‡, Christopher R. Trotta, and Hong Li

From the California Institute of Technology, Division of Biology, Pasadena, California 91125

Introns interrupt the continuity of many eukaryal genes, and therefore their removal by splicing is a crucial step in gene expression. Interestingly, even within Eukarya there are at least four splicing mechanisms. mRNA splicing in the nucleus takes place in two phosphotransfer reactions on a complex and dynamic machine, the spliceosome. This reaction is related in mechanism to the two self-splicing mechanisms for Group 1 and Group 2 introns. In fact the Group 2 introns are spliced by an identical mechanism to mRNA splicing, although there is no general requirement for either proteins or co-factors. Thus it seems likely that the Group 2 and nuclear mRNA splicing reactions have diverged from a common ancestor. tRNA genes are also interrupted by introns, but here the splicing mechanism is quite different because it is catalyzed by three enzymes, all proteins and with an intrinsic requirement for ATP hydrolysis.

tRNA splicing occurs in all three major lines of descent, the Bacteria, the Archaea, and the Eukarya. In bacteria the introns are self-splicing (1–3). Until recently it was thought that the mechanisms of tRNA splicing in Eukarya and Archaea were unrelated as well. In the past year, however, it has been found that the first enzyme in the tRNA splicing pathway, the tRNA endonuclease, has been conserved in evolution since the divergence of the Eukarya and the Archaea. Surprising insights have been obtained by comparison of the structures and mechanisms of tRNA endonuclease from these two divergent lines.

tRNA Precursors in Eukarya and Archaea

The earliest studies of tRNA splicing were in the yeast *Saccharomyces cerevisiae* where tRNA introns were first discovered (4, 5). With the completion of the *S. cerevisiae* genome sequence it is now known that yeast contains 272 tRNA genes of which 59, encoding 10 different tRNAs, are interrupted by introns (6). The introns are 14–60 nucleotides in length and interrupt the anticodon loop immediately 3′ to the anticodon (7). Among the 10 different yeast pre-tRNAs there is no conservation of sequence at the splice junctions although the 3′-splice junction is invariably in a bulged loop (8). Early studies on the structure of yeast tRNA precursors showed that the conformation of the mature domain is retained suggesting the model of the tertiary structure of eukaryal pre-tRNA shown in Fig. 1*A* (9, 10).

In the Archaea the introns are also small and often interrupt the anticodon loop, but they are found elsewhere as well, for example interrupting the dihydro U stem (11). In several of the Archaea, tRNA genes have been found that contain two introns. The splice sites are found in an absolutely conserved structural motif consisting of two loops of three bases separated by a four-base pair helix, the bulge-helix-bulge (BHB)[1] motif (12). This structure, modeled in Fig. 1*B* from the related TAR RNA structure (13), allows the archaeal splicing mechanism to be extended to introns in rRNA that also retain this motif. Thus, early on it was suggested that the eukaryal and archaeal splicing systems operate by a different mechanism.

The Pathway of tRNA Splicing in Eukarya

The early discovery by Hopper and co-workers (14) that pre-tRNAs accumulate in the yeast mutant rna1–1 provided a source of pre-tRNA substrates, which allowed the development of the first *in vitro* RNA splicing system (15, 16). Using this system the pathway of tRNA splicing was deduced (17–20).

The tRNA splicing reaction in yeast occurs in three steps; each step is catalyzed by a distinct enzyme, which can function interchangeably on all of the substrates (Fig. 2). In the first step the pre-tRNA is cleaved at its two splice sites by an endonuclease. The products of the endonuclease reaction are the two tRNA half-molecules and the linear intron with 5′-OH and 3′-cyclic PO_4 ends. The endonuclease has been purified to homogeneity (6, 21). The enzyme behaves as an integral membrane protein, and since splicing takes place in the nucleus, it may be an inner nuclear envelope protein. The two tRNA half-molecules, in essence a nicked tRNA, are the substrate for the ensuing ligase reaction. This baroque reaction, catalyzed by the 90-kDa tRNA ligase (22, 23), takes place in three steps. In the first step the cyclic PO_4 is opened to give a 2′-PO_4 and 3′-OH. In the second step the 5′-OH is phosphorylated with the γ-PO_4 of GTP (24, 25). tRNA ligase is adenylated at an active site lysine (26), and then the AMP is transferred to the 5′-PO_4 of the substrate. Formation of the 5′–3′-phosphodiester bond proceeds and AMP is released. The phosphate at the spliced junction is derived from the γ-phosphate of GTP, and the phosphate originally at the 5′-splice site remains at the spliced junction as a 2′-phosphate and must be removed to complete the splicing reaction. Phizicky and co-workers (27) have characterized the enzymology of the removal of the 2′-PO_4 from the spliced tRNA. A nicotinamide adenine dinucleotide (NAD)-dependent phosphotransferase catalyzes the transfer of the 2′-PO_4 to NAD (28, 29). Surprisingly the structure of the transfer product is ADP-ribose 1″–2″-cyclic phosphate (30). The nicotinamide moiety is displaced, apparently supplying the energy for cyclization. It is tempting to speculate that this unique and hitherto unknown compound goes on to play some crucial regulatory role in the cell.

Specificity in tRNA Splicing: Recognition of pre-tRNA by the Endonuclease

The eukaryal endonuclease is solely responsible for the recognition of the splice sites contained in the pre-tRNA. Since only the mature domain in the pre-tRNA is conserved, it was postulated that endonuclease recognizes the splice sites by measuring the distance from the mature domain to the splice sites (31). This hypothesis was confirmed by experiments in which insertion mutations in the pre-tRNA that changed the distance between the mature domain and the anticodon resulted in a predictable shift of the splice sites (31).

The intron is not completely passive in the recognition process. The *Xenopus* oocyte tRNA endonuclease has been shown to recognize a crucial element involving the intron (8). Yeast tRNA introns contain a conserved purine residue three nucleotides upstream of the 3′-splice site. This base must be able to pair with a pyrimidine at position 32 in the anticodon loop in order for the intron to be recognized by either yeast or *Xenopus* endonuclease (8, 32). These experiments suggested complexities in the structure of the pre-tRNA effecting the recognition by the eukaryal endonuclease, which had not been previously appreciated. It has also been demonstrated that there are different requirements for the recognition of the 5′- and 3′-splice sites (32).

Recognition of archaeal tRNA splice sites by the archaeal tRNA endonuclease relies solely on the BHB motif (Fig. 1*B*). Yeast pre-tRNAs are not substrates for the *Haloferax volcanii* endonuclease (33), and unlike the eukaryal endonuclease, the mature domain of the pre-tRNA is not required for intron excision by the archaeal enzyme (12). Despite the differences in both substrate and the mechanism for substrate recognition between the archaeal and

* This minireview will be reprinted in the 1998 Minireview Compendium, which will be available in December, 1998.

‡ To whom correspondence should be addressed. Fax: 626-796-7066; E-mail: abelsonj@cco.caltech.edu.

[1] The abbreviations used are: BHB, bulge-helix-bulge; pre-tRNA, precursor tRNA; TAR, trans-activating response region.

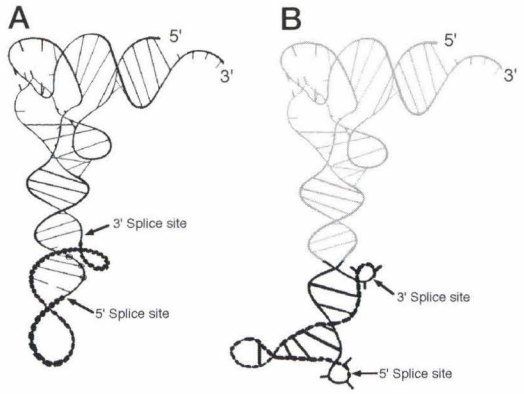


FIG. 1. **Proposed tertiary structure of the tRNA splicing substrates.** Model of the eukaryal tRNA splicing endonuclease (*A*) as proposed by Lee and Knapp (9) based on the crystal structure of $tRNA^{Phe}$ and the substrate for the archaeal tRNA splicing endonuclease (*B*) modeled from the NMR structure of the HIV-1 TAR RNA (42, 43). The intron is shown in *bold, dashed lines*. The proposed tertiary structure of archaeal tRNA is shown in *gray*.

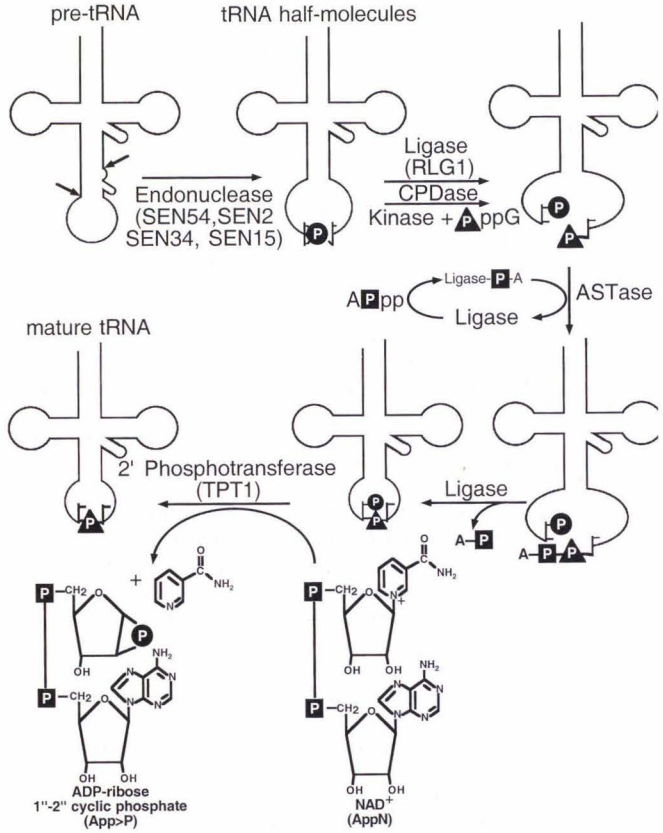


FIG. 2. **The tRNA splicing pathway of yeast.** Gene names are given in parentheses for the proteins of the pathway. See text for details. CPDase, cyclic phosphodiesterase; ASTase, adenylyl synthetase. Adapted with permission from Refs. 24 and 30.

eukaryal systems, as we shall discuss below, the endonuclease that catalyzes the first step in splicing has been conserved between Eukarya and Archaea. Different mechanisms for substrate recognition have evolved since the divergence from their common ancestor.

The Yeast and Archaeal tRNA Splicing Endonucleases: Related Enzymes

The characterization of the yeast tRNA endonuclease was extremely difficult for two reasons. The enzyme was present at very low levels, approximately 150 molecules per cell, and it appeared to be an integral membrane protein. However, after many years of work the enzyme was successfully purified, and the genes for all four of its subunits were cloned (6, 21). Of particular advantage in purifying the enzyme was the construction of a modified gene for the 44-kDa subunit, *SEN2*, containing His^8 and Flag epitope affinity tags. The *SEN2* gene had been found earlier in genetic screens (34, 35), and the *sen2–3* allele was shown to specifically block 5′-splice site cleavage (35).

The enzyme turned out to be an $\alpha\beta\gamma\delta$ heterotetramer whose subunits have molecular masses of 54 (*SEN54*), 44 (*SEN2*), 34 (*SEN34*), and 15 kDa (*SEN15*). Each of these genes proved to be essential for cell viability. The amino acid sequence of each of the four subunits contains a canonical nuclear localization sequence. Sen2 contains the only plausible transmembrane sequence, suggesting that it anchors the endonuclease complex to the nuclear membrane. Two subunits of endonuclease, Sen2 and Sen34, contain a homologous domain approximately 130 amino acids in length, suggesting that they perform a similar function. The excitement came when we learned that apparent homologs of this domain are encoded by the gene for the archaeal tRNA splicing endonuclease of *H. volcanii* cloned by Daniels and co-workers (36) and in endonuclease homologs found in the sequenced genomes of *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, and *Archaeoglobus fulgidis*.

The homology between Sen2, Sen34, and the archaeal endonucleases immediately suggested a model in which the yeast endonuclease contains two distinct active sites, one for each splice site (6, 37). The fact that *sen2–3* is defective in cleavage of the 5′-splice site suggested that Sen2 contains the active site for 5′-splice site cleavage and led to the prediction that Sen34 cleaves the 3′-splice site. This hypothesis was strongly supported by the observation that a Sen34 mutant enzyme is defective in 3′-splice site cleavage (6).

Our biochemical experience with the endonuclease had suggested strong interactions between the subunits. To probe the nature of these interactions a two-hybrid experiment was performed in which all possible pairwise combinations of the four subunits were probed (6). It turned out that strong interactions were only seen between Sen2 and Sen54 and between Sen34 and Sen15. These results together with those described above lead us to a model for the yeast endonuclease in which Sen2 contains the active site for 5′-splice site cleavage and Sen34 the active site for 3′-splice site cleavage (see below; Fig. 4*C*). There is as yet no evidence as to how the ruler mechanism works, although we propose that it could be via the interaction of Sen54, a very basic protein, and Sen2.

The tRNA splicing endonuclease of the archaeon *H. volcanii* was shown to behave as a homodimer in solution (36). Since its substrate, the consensus BHB motif, has pseudo 2-fold symmetry, it seems very likely that the 2-fold symmetric dimer recognizes its substrate such that each splice site is cleaved by a separate active site (see below). Thus we are led to a unified model of tRNA splicing in which the two splice sites are cleaved by separate protein subunits, each containing an active site.

The Three-dimensional Structure of an Archaeal tRNA Splicing Endonuclease

M. jannaschii contains a gene that encodes an endonuclease homologous to the *H. volcanii* enzyme but is about half the size (179 amino acids in the case of *M. jannaschii*). We believed that a high resolution structure of the simpler archaeal enzyme would shed light on the mechanism of the more complicated but related eukaryal endonuclease and thus embarked on a structural study of the *M. jannaschii* endonuclease, obtaining an x-ray structure refined to a resolution of 2.3 Å (13).

The *M. jannaschii* endonuclease is an α_4 tetramer different from the dimeric *H. volcanii* enzyme (13, 36, 38). Fig. 3*A* shows that the *M. jannaschii* endonuclease monomer consists of two distinct domains: the N-terminal domain (residues 9–84) and the C-terminal domain (residues 85–179). The N-terminal domain consists of three α-helices and a mixed antiparallel/parallel β-pleated sheet of four strands. The C-terminal domain contains two α-helices flanking a five-stranded mixed β-sheet. The last strand β9 is partially hydrogen bonded with β8, but its main interactions mediate the isologous pairing seen in the tetramer (see below). Understanding the interactions between monomers has turned out to be crucial to under-

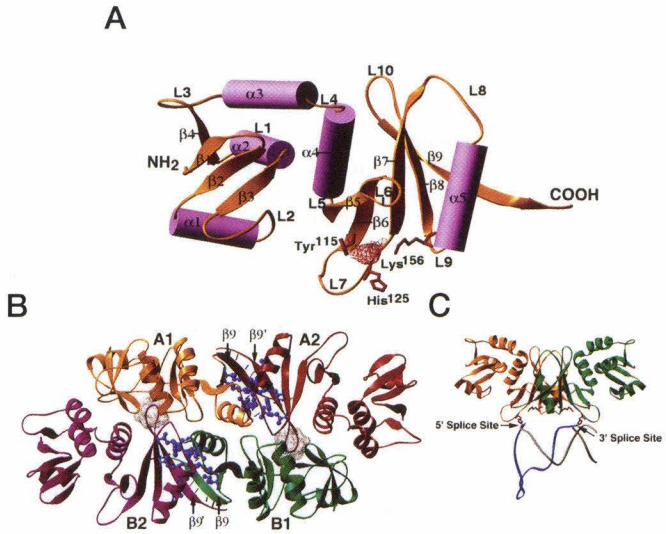


FIG. 3. **Crystal structure of the *M. jannaschii* tRNA splicing endonuclease.** *A*, ribbon representation of the *M. jannaschii* endonuclease monomer. The proposed catalytic triad residues are within 7 Å of each other and are shown in *red* ball-and-stick models (see text). The electron density is drawn near the putative catalytic triad to represent a putative SO_4^{2-} bound in the active site pocket. *B, M. jannaschii* endonuclease tetramer. Each subunit is represented by a distinct color and a label. The tetramer is viewed along the true 2-fold axis relating the A1-A2 and B1-B2 dimers. The main chain hydrogen bonds formed between β9 and β9′ and between loops L8 and L8′ for isologous dimers are drawn as *thin lines*. Side chains of the hydrophobic residues enclosed at the dimer interface are shown as *blue ball-and-stick* models. The heterologous interaction between subunits A1 and B2 (or B1 and A2) through the acidic loops L10 and L8 are highlighted by *dotted surfaces*. *C,* model of the *M. jannaschii* endonuclease active site subunits A1 and B1 docked with a model substrate based on the TAR RNA structure (intron in *blue*) (see also Fig. 1*B*). Adapted with permission from Ref. 13.

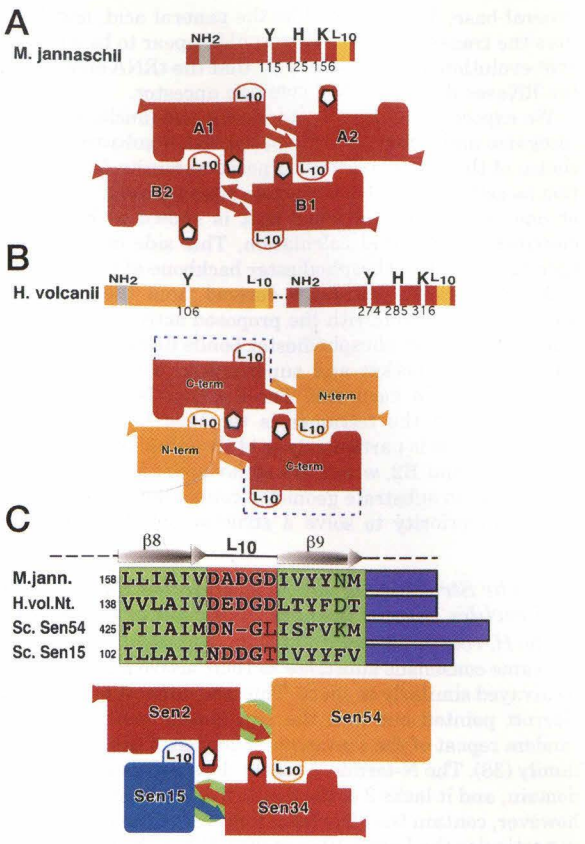


FIG. 4. **Model of the tRNA splicing endonucleases of *M. jannaschii, H. volcanii,* and *S. cerevisiae.*** *A,* model of the *M. jannaschii* homotetramer. Several important structural features discussed in the text are indicated: loop L10, the COOH-terminal β9 strands (*arrows*) and the conserved catalytic residue His^{125} (*pentagon*). *B,* proposed subunit arrangement of the *H. volcanii* endonuclease. The two tandem repeats are more similar to the *M. jannaschii* endonuclease sequence than to each other. The NH_2-terminal repeat lacks two of the three putative active site residues (*white bars*). It does, however, contain many of the features of the COOH domain, which are important for structural arrangement of the enzyme, in particular the L10 sequence (*yellow bars*). The COOH-terminal repeat contains all the sequence features of the *M. jannaschii* enzyme. *Dashed lines* represent the polypeptide chain connecting the COOH- and NH_2-terminal repeats. *C*, a proposed structural model for the yeast endonuclease. Conserved amino acid sequences near the COOH termini of archaeal enzymes, *M. jannaschii* (*M. jann.*), *H. volcanii* NH_2-terminal repeat (*H. vol. Nt.*), and yeast Sen54 (*Sc. Sen54*) and Sen15 (*Sc. Sen15*) subunits are aligned. Important hydrophobic residues that stabilize the isologous COOH terminus interaction between *M. jannaschii* subunits A1 and A2 or B1 and B2 are highlighted in *green* and *circled* on the structural models of heterodimers. The sequences of Sen54 and Sen15 aligned with L10 sequences in *M. jannaschii* and *H. volcanii* are highlighted in *red*. Loops L10 on both Sen54 and Sen15 are labeled on the proposed heterotetramer model of the yeast endonuclease. Reprinted with permission from Ref. 13.

standing the structure and evolution of the members of the tRNA splicing endonuclease family.

Two pairs of subunits (A1 and A2, B1 and B2) associate to form isologous dimers via extensive interactions between their β9 strands (Fig. 3*B*). The carboxyl half of β9 from one subunit forms main chain hydrogen bonds with the symmetry-related residues of β9 from the other subunit (β9′), leading to a two-stranded β-sheet spanning the subunit boundary. The two L8 loops, also hydrogen bonded and related by the same symmetry, form another layer on top of the two-stranded β9 sheet. The β9-β9′ sheet together with L8-L8′ encloses a hydrophobic core at the intersubunit surface. These hydrophobic residues are important for stabilizing the dimer interface and lead to an extremely stable dimeric unit, which we believe has been conserved in evolution (see below).

The tetramer is formed via heterologous interaction between the two dimers. The main interaction between the two dimers is via the insertion of loop L10 from subunits A2 and B2 into a cleft in subunits B1 and A1 between the N- and C-terminal domains of each monomer. The interaction is primarily polar between acidic residues in loop L10 and basic residues in the cleft. This arrangement causes the two isologous dimers to be translated relative to each other by about 20 Å. This brings subunits A1 and B1 much closer together than A2 and B2, which do not interact at all, and results, as we shall see below, in an arrangement of subunits in which only one symmetrically disposed pair of active sites can recognize the substrate. These interactions, though probably less stable, are also likely to have been conserved in evolution because they lead to a distinctive interaction between the dimers, which in turn leads to the required positioning of the two active sites.

The tRNA splicing endonucleases cleave pre-tRNA leaving 5′-OH and 2′–3′-cyclic PO_4 termini. This is the same specificity seen in the ribonucleases, RNase A, T1, etc. The RNase A mechanism has been studied extensively, and a first guess would be that chemically the endonuclease mechanism should be similar (39, 40). The reaction pathway for RNase A is a two-step acid-base-catalyzed reaction. A general base abstracts a proton from the 2′-OH of ribose leading to an in-line attack on the adjacent phosphodiester bond and the formation of a pentacovalent intermediate. The general acid protonates the 5′-leaving group leading to the 2′–3′ cyclic PO_4 product. In a second step a proton is abstracted from H_2O, OH^- attacks, and the 2′–3- cyclic PO_4 is hydrolyzed to give the 3′-PO_4. In RNase A, His^{12} is the general base in the first step, His^{119} is the general acid, and the pentacovalent transition state is stabilized by Lys^{41}.

In the endonuclease family there is a conserved histidine residue at position 125 in the *M. jannaschii* enzyme. There is strong evidence that this histidine is part of the active site. A change to alanine in the equivalent histidine at position 242 in Sen34 was shown to impair 3′-splice site cleavage (6). Daniels[2] has shown that the equivalent histidine mutant in the *H. volcanii* enzyme impairs cleavage as has Garrett (38) for a His^{125} to Ala mutant in the *M. jannaschii* endonuclease.

His^{125} (in L7) is found in a cluster with conserved residues Tyr^{115} (in L7) and Lys^{156} (in α5) on the surface of the monomer and forms a pocket into which the scissile phosphate to be cleaved is proposed to fit (Fig. 3*A*). Significantly these three residues can be spatially superimposed with the catalytic triad of RNase A. In the superposition, His^{125} is equivalent to His^{12} of RNase A and should be the

[2] C. J. Daniels, personal communication.

general base; Tyr^{115} should be the general acid, and Lys^{156} stabilizes the transition state. This would appear to be a case of convergent evolution, because it is clear that the tRNA endonucleases and the RNases do not share a common ancestor.

We expect two of the *M. jannaschii* endonuclease active sites to recognize and cleave the symmetric tRNA substrate. We favor the choice of the symmetrically disposed subunits A1 and B1 to function as active subunits. The active sites on subunits A1 and B1 are at one side of the tetramer that is shown to be basic from an electrostatic potential calculation. This side of the surface could therefore bind the phosphodiester backbone of the tRNA substrate.

A model of the substrate derived from the TAR RNA NMR structure docks well with the proposed active subunits A1 and B1 (Fig. 3*C*). The two phosphodiester bonds fit exactly into the A1 and B1 active site pocket and superimpose with the putative SO_4^{2-} density found in each site. The distance between other pairs of active sites in the tetramer is too long to fit with this model substrate. This is particularly so of the other symmetrically related pair, in A2 and B2, which are so far apart that it is unlikely that any change in substrate geometry could allow for a fit. Obviously it is of high priority to solve a structure of the enzyme-substrate complex.

The Structure of the M. jannaschii Endonuclease Provides Insight into the Evolution of tRNA Splicing

The *H. volcanii* dimer and the *M. jannaschii* tetramer recognize the same consensus substrate so their active sites must ultimately be arrayed similarly in space. This was difficult to understand until Garrett pointed out that the *H. volcanii* monomer is actually a tandem repeat of the consensus sequence of the endonuclease gene family (38). The N-terminal repeat does not contain the N-terminal domain, and it lacks 2 of the 3 putative active site residues. It does, however, contain the structural features of the C-terminal domain, in particular the Loop L10 sequence. Fig. 4 shows a proposed model of the *H. volcanii* enzyme, which is best described as a pseudo-tetramer of two pseudo-dimers. The structure of the pseudo-dimer is predicted to contain a two stranded $\beta9$-$\beta9'$ pleated sheet, an important structural feature of the *M. jannaschii* dimer. The *H. volcanii* enzyme only contains two active sites (found in the C-terminal repeats), and these are proposed to occupy an identical spatial configuration to those in the A1 and B1 subunits of the *M. jannaschii* enzyme. The pseudo-dimers are proposed to interact via the conserved loop L10 sequences in the N-terminal repeats, equivalent to those in the A2 and B2 subunits in the *M. jannaschii* enzyme. The *H. volcanii* enzyme tells us that only two of the active sites are necessary, but to array these in space correctly one must retain important features of both the isologous dimer interactions ($\beta9$-$\beta9'$) and the dimer-dimer interactions mediated by Loop L10.

The yeast endonuclease contains two active site subunits, Sen2 and Sen34. The other two subunits do not appear to belong to the endonuclease gene family; however, Garrett pointed out that both Sen15 and Sen54 contain a stretch of sequence similarity to the endonuclease family near their COOH termini (38). Upon inspection of the crystal structure of the *M. jannaschii* endonuclease, this sequence conservation appears to contain the Loop L10 and hydrophobic core interactions crucial to the association of the monomers to form the tetrameric *M. jannaschii* endonuclease. We have proposed that the strong Sen2-Sen54 and Sen34-Sen15 interactions seen in two-hybrid experiments (13) are mediated by the C-terminal $\beta9$-$\beta9'$-like interactions. These two heterodimers are proposed to interact to form the heterotetramer via the conserved Loop 10 sequences of Sen15 and Sen54.

Thus it is likely that what has been conserved since the divergence of the Eukarya and the Archaea is the endonuclease active site and the means to array two of them in a precise and conserved spatial orientation. Further support for this evolutionary pathway is supported by the results of Tocchini-Valentini and co-workers (41), where it is demonstrated that both the eukaryal and archaeal endonucleases can accurately cleave a universal substrate containing the BHB motif. The eukaryal enzyme seems to dispense with the ruler mechanism for tRNA substrate recognition when cleaving the universal substrate. This leads to the conclusion that the precise positioning of two active sites in endonuclease has been conserved. Thus, subunits A1 and B1 comprise the active site core of all tRNA splicing endonucleases, and subunits A2 and B2 position the two active sites precisely in space. The eukaryal enzyme has evolved a distinct measuring mechanism for splice site recognition via the specialization of the A2 and B2 subunits while retaining the ability to recognize and cleave the primitive consensus substrate.

REFERENCES

1. Biniszkiewicz, D., Cesnaviciene, E., and Shub, D. A. (1994) *EMBO J.* **13,** 4629–4635
2. Reinhold-Hurek, B., and Shub, D. A. (1992) *Nature* **357,** 173–176
3. Kuhsel, M. G., Strickland, R., and Palmer, J. D. (1990) *Science* **250,** 1570–1573
4. Goodman, H. M., Olson, M. V., and Hall, B. D. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5453–5457
5. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R., and Rutter, W. J. (1978) *Proc. Natl. Acad. Sci. U. S. A.* **75,** 190–194
6. Trotta, C. R., Miao, F., Arn, E. A., Stevens, S. W., Ho, C. K., Rauhut, R., and Abelson, J. N. (1997) *Cell* **89,** 849–858
7. Ogden, R. C., Lee, M. C., and Knapp, G. (1984) *Nucleic Acids Res.* **12,** 9367–9382
8. Baldi, M. I., Mattoccia, E., Bufardeci, E., Fabbri, S., and Tocchini-Valentini, G. P. (1992) *Science* **255,** 1404–1408
9. Lee, M. C., and Knapp, G. (1985) *J. Biol. Chem.* **260,** 3108–3115
10. Swerdlow, H., and Guthrie, C. (1984) *J. Biol. Chem.* **259,** 5197–5207
11. Thompson, L. D., Brandon, L. D., Nieuwlandt, D. T., and Daniels, C. J. (1989) *Can. J. Microbiol.* **35,** 36–42
12. Thompson, L. D., and Daniels, C. J. (1988) *J. Biol. Chem.* **263,** 17951–17959
13. Li, H., Trotta, C. R., and Abelson, J. N. (1998) *Science* **280,** 279–284
14. Hopper, A. K., Banks, F., and Evangelidis, V. (1978) *Cell* **14,** 211–219
15. Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A., and Abelson, J. N. (1978) *Cell* **14,** 221–236
16. O'Farrell, P. Z., Cordell, B., Valenzuela, P., Rutter, W. J., and Goodman, H. M. (1978) *Nature* 438–445
17. Greer, C. L., Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32,** 537–546
18. Peebles, C. L., Ogden, R. C., Knapp, G., and Abelson, J. (1979) *Cell* **18,** 27–35
19. Knapp, G., Ogden, R. C., Peebles, C. L., and Abelson, J. (1979) *Cell* **18,** 37–45
20. Peebles, C. L., Gegenheimer, P., and Abelson, J. (1983) *Cell* **32,** 525–536
21. Rauhut, R., Green, P. R., and Abelson, J. (1990) *J. Biol. Chem.* **265,** 18180–18184
22. Phizicky, E. M., Schwartz, R. C., and Abelson, J. (1986) *J. Biol. Chem.* **261,** 2978–2986
23. Westaway, S. K., Phizicky, E. M., and Abelson, J. (1988) *J. Biol. Chem.* **263,** 3171–3176
24. Westaway, S. K., Belford, H. G., Apostol, B. L., Abelson, J., and Greer, C. L. (1993) *J. Biol. Chem.* **268,** 2435–2443
25. Belford, H. G., Westaway, S. K., Abelson, J., and Greer, C. L. (1993) *J. Biol. Chem.* **268,** 2444–2450
26. Xu, Q., Teplow, D., Lee, T. D., and Abelson, J. (1990) *Biochemistry* **29,** 6132–6138
27. Culver, G. M., McCraith, S. M., Consaul, S. A., Stanford, D. R., and Phizicky, E. M. (1997) *J. Biol. Chem.* **272,** 13203–13210
28. McCraith, S. M., and Phizicky, E. M. (1990) *Mol. Cell. Biol.* **10,** 1049–1055
29. McCraith, S. M., and Phizicky, E. M. (1991) *J. Biol. Chem.* **266,** 11986–11992
30. Culver, G. M., McCraith, S. M., Zillmann, M., Kierzek, R., Michaud, N., Lareau, R. D., Turner, D. H., and Phizicky, E. M. (1993) *Science* **261,** 206–208
31. Reyes, V. M., and Abelson, J. (1988) *Cell* **55,** 719–730
32. Di Nicola Negri, E., Fabbri, S., Bufardeci, E., Baldi, M. I., Gandini Attardi, D., Mattoccia, E., and Tocchini-Valentini, G. P. (1997) *Cell* **89,** 859–866
33. Palmer, J. R., Nieuwlandt, D. T., and Daniels, C. J. (1994) *J. Bacteriol.* **176,** 3820–3823
34. Winey, M., and Culbertson, M. R. (1988) *Genetics* **118,** 49–63
35. Ho, C. K., Rauhut, R., Vijayraghavan, U., and Abelson, J. (1990) *EMBO J.* **9,** 1245–1252
36. Kleman-Leyer, K., Armbruster, D. A., and Daniels, C. J. (1997) *Cell* **89,** 839–848
37. Belfort, M., and Weiner, A. (1997) *Cell* **89,** 1003–1006
38. Lykke-Andersen, J., and Garrett, R. A. (1997) *EMBO J.* **16,** 6290–6300
39. Thompson, J. E., and Raines, R. T. (1994) *J. Am. Chem. Soc.* **116,** 5467–5468
40. Walsh, C. (1979) *Enzymatic Reaction Mechanisms*, pp. 202–207, W. H. Freeman and Co., San Francisco
41. Fabbri, S., Fruscoloni, P., Bufardeci, E., Di Nicola Negri, E., Baldi, M. I., Gandini Attardi, D., Mattoccia, E., and Tocchini-Valentini, G. P. (1998) *Science* **280,** 284–286
42. Calnan, B. J., Tidor, B., Biancalana, S., Hudson, D., and Frankel, A. D. (1991) *Science* **252,** 1167–1171
43. Aboul-ela, F., Karn, J., and Varani, G. (1996) *Nucleic Acids Res.* **24,** 3974–3981

Appendix II

Cell, Vol. 89, 849–858, June 13, 1997, Copyright ©1997 by Cell Press

The Yeast tRNA Splicing Endonuclease: A Tetrameric Enzyme with Two Active Site Subunits Homologous to the Archaeal tRNA Endonucleases

Christopher R. Trotta, Feng Miao,*
Eric A. Arn,† Scott W. Stevens,
Calvin K. Ho, Reinhard Rauhut,‡
and John N. Abelson
Division of Biology
California Institute of Technology
Pasadena, California 91125

Summary

The splicing of tRNA precursors is essential for the production of mature tRNA in organisms from all major phyla. In yeast, the tRNA splicing endonuclease is responsible for identification and cleavage of the splice sites in pre-tRNA. We have cloned the genes encoding all four protein subunits of endonuclease. Each gene is essential. Two subunits, Sen2p and Sen34p, contain a homologous domain of approximately 130 amino acids. This domain is found in the gene encoding the archaeal tRNA splicing endonuclease of H. volcanii and in other Archaea. Our results demonstrate that the eucaryal tRNA splicing endonuclease contains two functionally independent active sites for cleavage of the 5′ and 3′ splice sites, encoded by *SEN2* and *SEN34*, respectively. The presence of endonuclease in Eucarya and Archaea suggests an ancient origin for the tRNA splicing reaction.

Introduction

Although RNA splicing is a required step for gene expression in all three taxonomic kingdoms, at least four different mechanisms of splicing have evolved. The most primitive mechanism, autocatalytic self splicing of Group I and Group II introns, involves two phosphotransfer reactions (Cech, 1993). Pre-mRNA splicing requires ATP hydrolysis and more than 100 proteins, but it proceeds via the same chemical mechanism used by Group II introns, suggesting that it is also RNA catalyzed and that it shares a common origin with Group II intron splicing (Moore et al., 1993; Madhani and Guthrie, 1994). In contrast to these RNA-catalyzed reactions, splicing of eucaryotic tRNA introns occurs by a mechanism involving three protein enzymes (Winey and Culbertson, 1989; Abelson, 1991; McCraith and Phizicky, 1991). In the archaeal kingdom, tRNA splicing proceeds by a similar protein-based enzymatic mechanism, but this splicing is distinct by a number of functional criteria (Thompson et al., 1989). Splicing of all known prokaryotic tRNA introns occurs via the primitive autocatalytic mechanism (Reinholdhurek and Shub, 1992; Biniszkiewicz et al., 1994). So while tRNA splicing occurs in all three kingdoms, the mechanism for this process is not conserved. Major questions regarding the origin of tRNA introns and the relationship between the disparate enzymatic machineries required for their removal therefore remain. Do the different splicing mechanisms reflect independent origins or a single origin and later divergence? The results of this work together with those in the accompanying papers of Di Nicola Negri et al. (1997 [this issue of *Cell*]) and Kleman-Leyer et al. (1997 [this issue of *Cell*]) provide new insight into the origin of the tRNA splicing machinery.

Much of our knowledge of the mechanism of eucaryotic tRNA splicing comes from studies in Saccharomyces cerevisiae. The S. cerevisiae genome contains 272 tRNA genes of which 59, encoding ten different tRNAs, contain introns (C. R. T., unpublished data). These introns are 14–60 bases in length, and all interrupt the anticodon loop immediately 3′ to the anticodon (Ogden et al., 1984). Splicing of introns from precursor tRNA (pre-tRNA) is accomplished through the action of a site-specific endonuclease, an RNA ligase, and a phosphotransferase. The tRNA splicing endonuclease (referred to herein as "endonuclease") cleaves pre-tRNA at the 5′ and 3′ splice sites to release the intron. The products of the endonuclease reaction are an intron and two tRNA half-molecules bearing 2′,3′ cyclic phosphate and 5′-OH termini (Peebles et al., 1983). tRNA ligase covalently joins the half-molecules through a complex series of ATP- and GTP- dependent reactions (Westaway et al., 1988; Belford et al., 1993; Westaway et al., 1993). After ligation, a 2′ phosphate remains at the splice junction and is removed by a specific phosphotransferase (McCraith and Phizicky, 1991; Culver et al., 1993).

Figure 1A summarizes the conserved features of the ten yeast pre-tRNAs containing introns (as depicted by Lee and Knapp, 1985). There are no conserved sequences at the splice sites, but the intron is invariably located at the same site in the gene, placing the splice sites an invariant distance from the constant structural features of the tRNA body. This conserved placement implies that recognition of the splice sites may be accomplished via interaction with the mature tRNA domain. The hypothesis that the endonuclease measures the distance from the mature domain to the splice sites to correctly place cleavages (Figure 1B) was confirmed by experiments in which insertions in the anticodon stem of pre-tRNA that changed the distance between the mature domain and the splice sites resulted in a predictable shift of the cleavage sites (Reyes and Abelson, 1988). Further evidence for a strict measuring mechanism was provided by the fact that an artificial tRNA intron composed almost entirely of uridine residues was spliced correctly in the yeast system (Reyes and Abelson, 1988).

The intron is not, however, completely passive in the recognition of the splice sites. The Xenopus tRNA endonuclease has been shown to recognize a crucial structural element in tRNA introns (Baldi et al., 1992). Yeast tRNA introns contain a conserved purine residue three

*Present Address: Beckman Research Institute of the City of Hope, 1450 E. Duarte Rd., Duarte, California 91010.
†Present Address: Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115.
‡Present Address: Justus Liebig Universität Giessen, Institut für Mikrobiologie und Molekularbiologie, Frankfurter Strasse 17, 35392 Giessen, Germany.

nucleotides upstream of the 3′ splice site. This base must be able to pair with a pyrimidine at position 32 in the anticodon loop for the intron to be recognized by either yeast or Xenopus endonuclease (Figure 1A) (Baldi et al., 1992; Di Nicola Negri et al., 1997). These data suggested that there could be different requirements for the recognition of the 5′ and 3′ splice sites. More recent evidence in this paper and the accompanying paper by Di Nicola Negri et al. (1997) supports the hypothesis that the endonuclease contains two independent active sites with different substrate recognition mechanisms.

The mechanism of intron removal in tRNA splicing has also been investigated through purification of the endonuclease enzyme from yeast (Rauhut et al., 1990). Endonuclease behaves as a membrane-bound protein present in a low abundance of ~100 molecules per cell (Rauhut et al., 1990). The purified enzyme was seen to contain three subunits of 54, 44, and 34 kDa and, as we report here, a fourth subunit of 15 kDa. The molecular weight of the holoenzyme as judged by gel exclusion chromatography implies that the enzyme is an $\alpha\beta\gamma\delta$ tetramer. In this paper, we report the primary sequences of all four of the endonuclease subunits. Analysis of the peptide sequence of the endonuclease subunits presented here sheds light on the architecture of the enzyme and on its possible interactions with other components of the nucleus.

The pre-tRNAs of Archaea are similar to those of Eucarya in that they contain introns that must be removed by protein enzymes in *trans*. Superficially, these introns are similar to yeast tRNA introns in that they are often (though not always) located in the anticodon loop. Removal of tRNA introns by either eucaryal or archaeal endonucleases yield 2′,3′ cyclic PO_4 and 5′-OH termini. Studies of removal of tRNA introns in Archaea have, however, shown notable differences between the yeast and archaeal tRNA endonucleases. Yeast pre-tRNAs are not substrates for the Haloferax volcanii endonuclease (Palmer et al., 1994), and the mature domain of a Halobacterium volcanii pre-tRNA is not required for correct excision of the intron (Thompson and Daniels, 1988). Rather, archaeal tRNA splice sites must be located in bulged loops separated by a helix of four base pairs (Figure 1C). This evidence would seem to suggest a different origin of the splicing process in these two kingdoms. The results presented here, however, together with those in the accompanying paper by Kleman-Leyer et al. (1997), establish that two of the yeast endonuclease subunits define a gene family that includes the single homodimeric 40 kDa H. volcanii endonuclease as well as related genes from other Archaea. We discuss the evolutionary and mechanistic implications of these findings.

Results

The Gene for the 44 kDa Subunit, *SEN2*

Our primary goal in the characterization of the yeast tRNA splicing endonuclease has been to clone the genes for each of the subunits of this enzyme. We previously reported the isolation and characterization of the cold-sensitive *sen2-3* mutation, an allele of the *SEN2* (splicing endonuclease 2) gene found by Winey and Culbertson (1988) in their search for yeast tRNA mutants defective in tRNA splicing in vitro. At the nonpermissive temperature, *sen2-3* accumulates "2/3" pre-tRNA molecules cleaved at the 3′ splice site only (Ho et al., 1990), strongly suggesting that Sen2p is a component of endonuclease. In this paper, we report the cloning of the *SEN2* gene by complementation of the *sen2-3* mutant and the determination of its nucleotide sequence (GenBank accession #M32336). Analysis of the amino acid sequence of Sen2p reveals the presence of a probable transmembrane sequence between residues 225 and 243 (Figure 2B). During purification, endonuclease behaves as a membrane-bound enzyme (Rauhut et al., 1990). None of the other subunits (see below) contain a convincing transmembrane sequence, so it is possible that Sen2p is the subunit that anchors the endonuclease complex to the nuclear membrane. Sen2p also contains a domain predicted to form a canonical antiparallel coiled-coil structure between residues 122 and 184 (Figure 2B). Sequencing of the mutant *sen2-3* allele reveals a single point mutation causing the change of glycine 292 to glutamate (Figure 2B).

A New Purification Strategy for the tRNA Splicing Endonuclease

We had previously developed a chromatographic procedure for the purification of endonuclease that resulted in homogenous enzyme (Rauhut et al., 1990) and provided sufficient material for the characterization of two of the endonuclease subunits. However, this procedure was

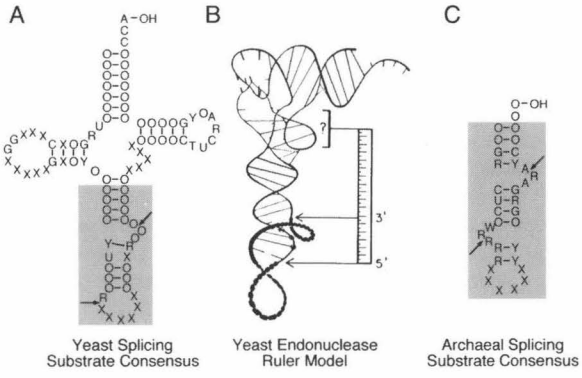


Figure 1. Conserved Sequence and Structure Elements of Yeast Pre-tRNA and Archaeal Minimal Pre-tRNA

(A) A composite, modified from Lee and Knapp (1985), of yeast pre-tRNA sequences (Ogden et al., 1984). (X) = unconserved base in a region of variable length or secondary structure; (O) = an unconserved base in a region of conserved secondary structure. (G), (C), and (A) indicate conserved bases conserved among pre-tRNAs. (Y) and (R) = conserved pyrimidines and purines, respectively. The conserved intron–anticodon tertiary interaction is depicted by a line between the conserved R and Y bases.

(B) Taken from Reyes and Abelson (1988). Presumed tertiary structure of pre-tRNA^{Phe} (from Lee and Knapp, 1985). Endonuclease is proposed to interact with the mature domain and measure the distance to the two splice sites.

(C) A composite of archaeal pre-tRNA sequences, modified from Thompson and Daniels (1988). Position designations are as in (A). This represents the minimal substrate for cleavage by the archaeal endonuclease.

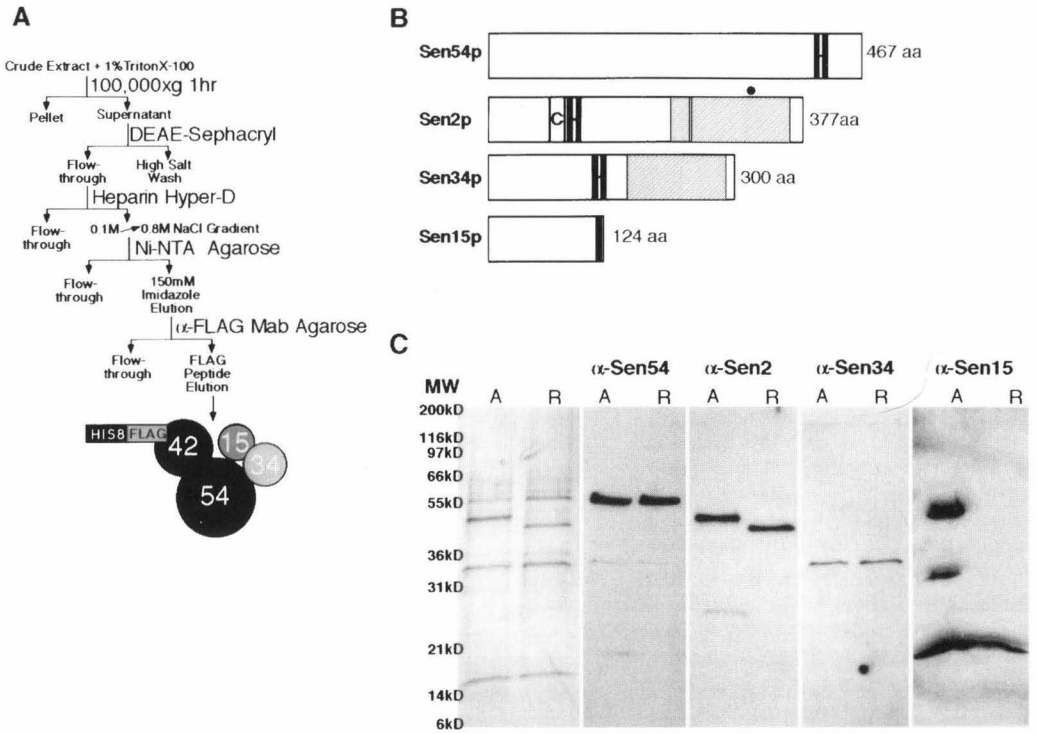


Figure 2. Affinity Purification of Endonuclease from Yeast Allows Isolation of the Genes Encoding the Four Subunits of the Endonuclease

(A) Purification procedure for isolation of affinity-tagged endonuclease from yeast strain CRT44 (see Experimental Procedures for details).

(B) Primary amino acid sequence of each of the endonuclease subunits depicted graphically. Closed boxes represent bipartite or basic nuclear localization sequences. Hatched box in Sen2p and Sen34p represents the homologous region between these two subunits and endonuclease from Archaea (see text for details). Boxed (C) in Sen2p represents putatative coiled-coil as predicted by COILS progam of Lupas et al. (1991). Dot above Sen2p represents G292E mutation. Shaded box in Sen2p represents putative transmembrane domain as predicted using the Kyte-Doolittle algorithm and a window of 19 amino acids.

(C) Left panel: silver stained gel of lane A, 0.5 μg of affinity tagged endonuclease purified as in Figure 2A; lane R, 0.5 μg of endonuclease purified by Rauhut et al. (1990). (MW) represents molecular weight standards (Novex), which are labeled in kilodaltons. Duplicate blots of gel from left panel were probed with 1:500 dilution of antibodies to Sen54p, Sen2p, Sen34p, and Sen15p, respectively.

extremely time consuming and gave a low yield of enzyme. We have developed a highly efficient affinity purification procedure for large-scale preparation of the enzyme based on the cloned *SEN2* gene.

The *SEN2* gene was modified by addition of an amino-terminal peptide tag consisting of 8 histidine residues followed by the 4 residue FLAG epitope. This gene, cloned in a yeast 2μ vector, complemented the haplolethal phenotype of a *SEN2* deletion, proving that the affinity tag did not affect the activity of Sen2p in vivo. The complemented *sen2* deletion strain, CRT44, was used as a source of tagged enzyme for purification by a procedure utilizing standard chromatographic methods and two specific affinity steps, as diagrammed in Figure 2A. This rapid procedure resulted in an efficient (10%–20%) yield of 5 μg of pure enzyme per 150 grams of cells.

A Fourth Subunit in the tRNA Splicing Endonuclease

Pure endonuclease preparations had previously been characterized by SDS–PAGE and silver staining, which detected three subunits with apparent molecular masses of 51 kDa, 42 kDa, and 31 kDa (Rauhut et al., 1990). To compare endonuclease purified by the affinity method to that obtained in our original purification, we employed a high resolution gradient gel procedure. This procedure revealed that both preparations contain proteins of 54, 44, and 34 kDa (cf. lanes A and R, Figure 2C), although the 44 kDa band in lane A has a lower mobility due to its affinity tag (the molecular weights of the proteins are designated here as determined by comparison with markers in gradient gel electrophoresis). Both preparations contained, in addition, a fourth band with an apparent molecular weight of 15 kDa. This band would have run with the dye front in our original procedure. The fact that all four proteins are present in stoichiometric amounts in endonuclease purified by different purification procedures and that all four proteins cosediment in glycerol density gradients (data not shown) strongly suggests that all four polypeptides are subunits of the enzyme. The molecular mass of the enzyme as measured by gel filtration chromatography in our earlier study was 140 kDa (Rauhut et al., 1990). With the fourth subunit, the predicted mass of a 1:1:1:1 endonuclease complex is 146 kDa, very close to the observed molecular mass. It is therefore likely that the subunit constitution of this enzyme is $\alpha\beta\gamma\delta$.

The Genes for the 34 kDa, 54 kDa, and 15 kDa Subunits

Purified endonuclease (50 μg) produced by the original Rauhut procedure was fractionated by SDS–PAGE, and individual bands were excised from the gel and digested

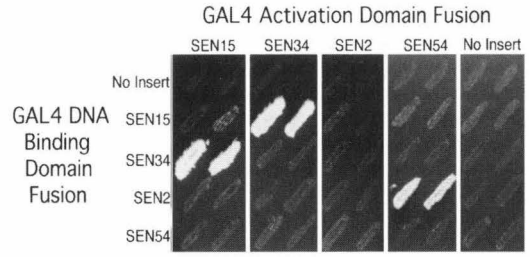


Figure 3. Two-Hybrid Matrix Analysis of Endonuclease Subunits

Activation of histidine reporter gene by various hybrid combinations of Gal4 binding domain and Gal4 activation domain to the subunits of endonuclease. Reporter strain was transformed, and then duplicate colonies from each were picked to a new selective plate. This plate was then replica plated to plates lacking histidine and containing 5 mM 3-aminotriazole (to suppress background growth) to determine activation of the reporter gene.

with lysyl-protease. A peptide resulting from digestion of the 34 kDa subunit yielded the sequence FIAYPGDPL RFX(R?)XLTIQ, uniquely found in the open reading frame (ORF) YAR008w on S. cerevisiae chromosome I. This ORF encodes a protein with a predicted molecular weight of 34 kDa. We have renamed this gene *SEN34*. Further evidence described below confirms that this protein is the 34 kDa subunit of the endonuclease.

Endonuclease (5 μg) purified by the affinity method described above was also fractionated by SDS–PAGE, and the 54 kDa band was digested with lysyl-protease. One unambiguous peptide sequence was obtained, NDDLQHFPTYK, which matches the sequence encoded by ORF YPL083c of S. cerevisiae chromosome XVI. This 467 amino acid protein has a predicted molecular weight of 54 kDa. We have renamed this gene *SEN54*. Sen54p is overall a very basic protein (pI = 9.5) that also contains several highly acidic regions. Further evidence described below confirms that this protein is a subunit of endonuclease.

Using a variation of the affinity purification method (see Experimental Procedures), several micrograms of pure 15 kDa subunit were prepared from SDS–PAGE–fractionated endonuclease and analyzed by trypsin digestion, followed by LC-MS/MS analysis resulting in identification of three peptides, two of which were identical (except for a possible acetylation modification of the methionine) to the peptide sequence MATTDIISLVK, suggesting that this peptide is the amino terminus of the protein (K. Swiderek, personal communication). The two unique peptides are found in the protein encoded by ORF YMR059w on chromosome XIII. This ORF potentially encodes a protein of 17 kDa. However, if the methionine codon at position 20 in this ORF is the translation start site for this protein, as suggested by the MS/MS analysis, the resulting protein would be predicted to have a length of 124 amino acids with a molecular weight of 15 kDa, in agreement with the size determined by SDS–PAGE. We have renamed this gene *SEN15*. Further evidence described below confirms that this protein is a subunit of endonuclease.

Sequence analysis of all four subunits revealed the presence of basic and bipartite nuclear localization sequences of the type described by Kalderon et al. (1984) and Robbins et al. (1991) (Figure 2B, closed boxes). Chromosomal disruptions of each of the four endonuclease subunit genes are lethal in haploid yeast cells, indicating that all of their gene products are essential for vegetative growth, and most likely for the tRNA splicing process (data not shown).

Antibodies to the Subunits of Endonuclease

To unambiguously determine that each of these genes encodes a subunit of endonuclease, we prepared polyclonal antisera specific for each of the four gene products expressed in E. coli. A Western blot experiment was performed in which purified endonuclease was fractionated by SDS–PAGE and individually probed with antisera specific for each of the four gene products. As shown in Figure 2C, probing of endonuclease purified by the affinity procedure (lane A), and endonuclease purified by the Rauhut procedure (lane R), reveals that each of the antisera reacts specifically with the appropriate polypeptide. The antibody to Sen15p was produced to a histidine/FLAG-tagged antigen, and as can be seen, the polyclonal antisera also contains antibodies to the affinity tag present on Sen2p and an apparent degradation product of Sen2p (Figure 2C, lane A; compare α-Sen15 to α-Sen2). Since the two preparations of pure endonuclease were prepared by different purification protocols, these results strongly support the conclusion that each of these four genes encodes a subunit of endonuclease.

Two-Hybrid Analysis of the Four Subunits of Endonuclease

To begin to understand the functional interactions of the endonuclease subunits, we have tested the four subunit genes in a two-hybrid matrix experiment. Each of the four genes was translationally linked to either a DNA-binding or transcriptional activation domain for analysis in the Fields two-hybrid system (Fields and Song, 1989; Bartel and Fields, 1995). Figure 3 shows the results of combinatorial expression of subunit gene pairs using rescue of histidine auxotrophy as an assay of interaction. Two interactions were detected in vivo: interaction between Sen2p and Sen54p, and between Sen34p and Sen15p. The reciprocal interaction between Sen2p and Sen54p did not activate expression, indicating that this interaction is sensitive to the orientation of the subunits relative to the binding domain/activation domains. Equivalent results were obtained with a β-galactosidase reporter gene (data not shown). Thus, we have observed interactions between Sen34p and Sen15p and between Sen2p and Sen54p, but the other possible pairwise interactions were not observed.

Sen2p and Sen34p Are Members of a Gene Family of Endonucleases

A search of the GenBank database (version 97) for sequences similar to the four endonuclease subunit genes did not reveal any potential homologs. However, similarity was detected between the Sen2 and Sen34 proteins themselves (see Figure 2B). The similarity occurs in a region of approximately 130 amino acids (see below). A sequence profile of the region of similarity between

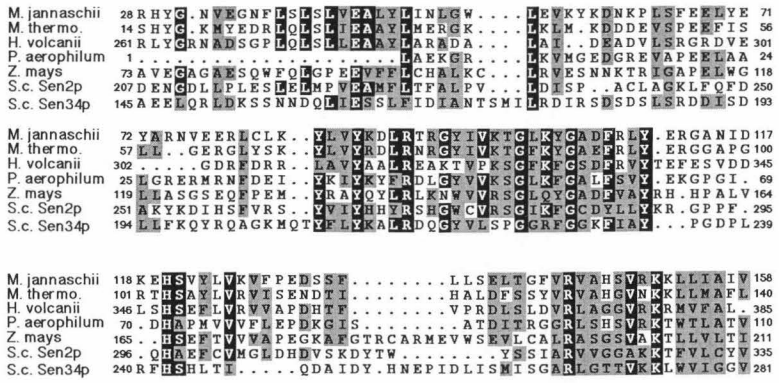


Figure 4. Sequence Alignment between the Archaeal Endonucleases and Two Subunits of the Yeast Endonuclease Identifies a Family of tRNA Splicing Endonucleases

Sequence alignment between endonuclease identified by Daniels and coworkers from H. volcanii and M. thermoautotrophicum, endonuclease from M. jannaschii, putative endonuclease from P. aerophilum, putative endonuclease from Z. mays, and two subunits, Sen34p and Sen2p, of the yeast tRNA splicing endonuclease. Sequence alignment was performed using the GCG program Pileup.

these proteins was used to search the translated GenBank database and was found to match one sequence, a 205 amino acid ORF contained in a presumed intron of an HMG-like gene of Zea mays (Yanagisawa and Izui, 1993).

C. Daniels and colleagues have determined the nucleotide sequence of the gene for the tRNA splicing endonuclease of the Archaea H. volcanii and Methanobacterium thermoautotrophicum (Kleman-Leyer et al., 1997). Homologs of this gene occur in other Archaeon, Pyrobaculum aerophilum and Methanococcus jannaschii (Bult et al., 1996). To determine if the M. jannaschii homolog is a functional homolog of the H. volcanii endonuclease, the gene product was expressed in E. coli, purified, and shown to cleave intron-containing archaeal pre-tRNAs (C. R. T., unpublished data). Interestingly, the endonucleases of the Archaea align with the yeast endonuclease subunits, Sen2p and Sen34p, and the ORF of Z. mays (Figure 4). In all, with sequences for seven members of this gene family available, a convincing homologous core domain of 50 amino acids is now apparent with lesser regions of homology extending over 130 amino acids.

Two Active Sites in Endonuclease

The sequence similarity between Sen2p and Sen34p suggests that these proteins have a similar function. It seems that this function is very likely to be the catalysis of tRNA cleavage, since these proteins belong to a family that includes the homodimeric tRNA splicing endonuclease of H. volcanii. Interestingly, members of the endonuclease family found in the genomes of M. thermoautotrophicum and M. jannaschii are both small proteins made up entirely of the sequence domain found in the yeast and H. volcanii endonucleases. This further supports the notion that the sequence alignment reveals the catalytic core of the endonuclease enzyme. Previous results have indicated that Sen2p carries the active site for 5′ splice site cleavage. Ho et al. (1990) showed that the *sen2-3* mutation results in endonuclease that is defective in cleavage only at the 5′ splice site of tRNA precursors in vivo (Ho et al., 1990). As reported here, the glycine-to-glutamate change in this allele is located at position 292, within the domain conserved between the eucaryal and archaeal endonucleases. This residue itself is not conserved in the endonuclease family, as perhaps expected for a conditional lethal mutation. Given that Sen34p is homologous to Sen2p and to the archaeal endonucleases we predicted that the Sen34p contains the active site for 3′ splice site cleavage. This prediction led to the model shown in Figure 5 in which the 5′ splice site is cleaved by Sen2p and the 3′ splice site by Sen34p.

This model clearly predicts that endonuclease containing a mutant Sen34p subunit should fail to cleave the 3′ splice site of pre-tRNA while retaining full activity in cleavage of the 5′ splice site. To test this prediction, we changed the conserved histidine at position 242 of *SEN34* to an alanine residue (H242A). Extrachromosomal expression of a HIS/FLAG-tagged mutant Sen34p H242A in wild-type haploid cells had a dominant deleterious effect, increasing the doubling time from 2.5 hr to 6 hr in minimal media. Extrachromosomal expression of wild-type Sen34p in the same strain had little or no effect on growth. Endonuclease was purified from each of these strains by the affinity purification method described in the Experimental Procedures to select only endonuclease complexes containing the tagged Sen34p subunit. tRNA splicing assays were performed with endonuclease containing wild-type Sen34p and Sen34p H242A. Endonuclease containing affinity-tagged wild-type Sen34p cleaves pre-tRNA^{Phe} normally, indicating

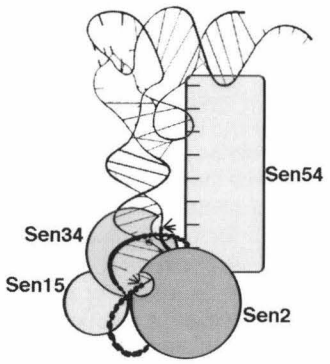


Figure 5. Two Active Site Model for Cleavage of Pre-tRNA by Yeast tRNA Splicing Endonuclease

Yeast pre-tRNA^{Phe} tertiary structure is depicted with a schematic diagram of the yeast endonuclease. In this model, Sen2p and Sen34p each contain an active site for cleavage of the 5′ and 3′ splice sites, respectively. The Sen2p–Sen54p and Sen34p–Sen15p interactions, detected in the two-hybrid analysis, are depicted as well as the potential tRNA recognition mediated by Sen54p (see text for details).

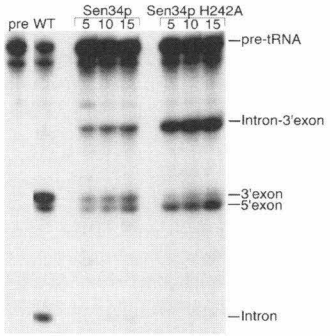


Figure 6. tRNA Splicing Activity of Sen34p H242A-Containing Endonuclease

Yeast pre-tRNA^{Phe} was incubated without endonuclease (pre) or with endonuclease purified from HIS/FLAG-tagged Sen2p (WT), HIS/FLAG-tagged wild-type Sen34p (Sen34p), or HIS/FLAG-tagged Sen34p H242A (Sen34p H242A) expressing yeast. Pre-tRNA was incubated for 5, 10, and 15 minutes, and the resulting cleavage products (shown to the right) were fractionated on a 10% polyacrylamide gel. The amount of endonuclease used in each time course was similar based on silver staining (wild-type enzyme was used at 100× concentration to allow for complete digestion of pre-tRNA).

that the affinity-tagged enzyme functions as wild-type endonuclease (Figure 6). At the low enzyme concentrations used in this experiment, a slight accumulation of a 2/3 molecule consisting of the intron joined to the 3′ exon, derived from cleavage at the 5′ splice site only, is normally seen (Miao and Abelson, 1993). In contrast, endonuclease containing Sen34p H242A shows a marked accumulation of 5′ exon and the intron-3′ exon 2/3 molecule. There is a nominal amount of 3′ exon and intron product, but we cannot determine if this is derived from retention of some 3′ splice site cleavage activity in Sen34p H242A or from the presence of some wild-type endonuclease in the purified mutant extract. However, the assay clearly demonstrates that endonuclease containing Sen34p H242A is, as predicted by the model, impaired in 3′ splice site cleavage.

Discussion

Yeast tRNA Endonuclease Contains Two Active Sites

The sequence similarity between Sen2p and Sen34p and their homology to the archaeal tRNA splicing endonucleases implied that each of these two subunits contains an active site for tRNA cleavage. As described above, endonuclease containing Sen34p H242A appears to be significantly impaired for 3′ splice site cleavage, but 5′ splice site cleavage occurs normally. A complete blockage of 5′ splice site cleavage was seen in endonuclease containing *sen2-3* (Ho et. al., 1990). These results taken together provide strong support for the model in Figure 5, indicating that the yeast tRNA splicing endonuclease contains two distinct, functionally independent active sites. Sen2p contains the active site for cleavage at the 5′ splice site and Sen34p contains the active site for cleavage at the 3′ splice site. We are currently undertaking a more extensive mutagenesis of *SEN34* and *SEN2* to generalize these results.

Additional support for the model shown in Figure 5 is derived from previous work on the Xenopus endonuclease, in which it has been shown that the two cleavage sites in yeast tRNA precursors have different properties. The Xenopus tRNA splicing endonuclease recognition of the 3′ splice site requires a base pair between a purine located 3 bases upstream of the 3′ splice site and the conserved pyrimidine at position 31 in the anticodon loop (the A-I base pair) (Baldi et al., 1992). Cleavage at the 5′ splice site requires only a purine as the adjacent 5′ nucleotide. It is conceivable that different specificities in cleavage could be obtained by a conformational change in a single active site occurring between cleavage reactions. For this to occur, one might expect an obligatory reaction pathway, i.e., cleavage at a primary site followed by cleavage at the second site. However, we have shown that the yeast endonuclease can and does cleave the 5′ and 3′ splice sites in a random order (Miao and Abelson, 1993).

More recently, the accompanying paper by Tocchini-Valentini's laboratory has shown that accurate 3′ splice site cleavage does not require the mature tRNA domain of the pre-tRNA (Di Nicola Negri et al., 1997). A small substrate containing only an intron and anticodon stem is cleaved by the Xenopus endonuclease at the 3′ splice site but not at the 5′ splice site. While both splice sites in an intact pre-tRNA are subject to the measuring mechanism, it appears that only the 5′ splice site absolutely requires this mechanism.

The H. volcanii tRNA splicing endonuclease cleaves with the same general mechanism as the eucaryal endonuclease, producing 2′,3′ cyclic PO_4 and 5′-OH termini, but has a different splice site recognition mechanism and fails to cleave yeast pre-tRNAs (Thompson and Daniels, 1988, 1990). The mature tRNA domain is not a requirement for the archaeal endonuclease. Archaeal tRNA splice sites are located in bulged regions separated by four helical base pairs (Figure 1C). In the accompanying paper by Kleman-Leyer et al. (1997), the H. volcanii endonuclease is shown to exist as a homodimer, and thus contains two active sites. The view that each of these active sites is functional is strengthened by the evidence presented here for two distinct active sites in the yeast endonuclease. However, because of the near symmetry of the archaeal splice site structure, we cannot rule out that a single active site could act interchangeably. Interestingly, most archaeal pre-tRNAs appear to contain the conserved A-I base pair of eucaryotic pre-tRNAs in the context of the four base pair stem separating the bulged loops of the splice sites. In this regard, the requirements for 3′ splice cleavage are more similar between Archaea and Eucarya than are those for 5′ splice site cleavage. In both kingdoms the splice site must be in a single-strand bulge near a base-paired element (in the eucaryotic pre-tRNAs, it is the A-I base pair).

Subunit Interactions in Endonuclease

In a complete matrix of two-hybrid interactions between the four subunits of endonuclease, strong interactions were seen only between the Sen2p and Sen54p pair and the Sen34p and Sen15p pair. Since all four subunits

remain tightly associated through two different purification regimens, it seems likely that there are also interactions between the two presumed heterodimers. These interactions, however, would not be detected in a two-hybrid experiment if they exist only within the context of the intact enzyme, or if three of the subunits are required for the interaction. The involvement of each of the active site subunits in a specific heterodimeric interaction suggests that the different specificities of the two active sites may be determined by these interactions. For example, the measuring mechanism, seen primarily in the recognition of the 5′ splice site, may be determined through the interaction between Sen2p and Sen54p, as indicated in Figure 5. Sen54p is a very basic protein and as such might be expected to interact with the phosphate backbone of the pre-tRNA substrate, presumably in the mature tRNA domain. Sen15p could be primarily responsible for the recognition of the A-I basepair.

Evolutionary Implications of the tRNA Splicing Endonuclease Gene Family

The existence of homologous tRNA splicing endonuclease genes in Eucarya and Archaea suggest that this enzyme was present in their common ancestor. Since endonuclease participates in tRNA splicing in both Eucarya and Archaea, it seems likely that the original function of this enzyme was the same. The discussion of the time of appearance of introns in evolution has been a lively one, and it is safe to say that a consensus has not emerged. The simplest hypothesis consistent with the data presented here and in the accompanying papers is that the function of endonuclease in the common eucaryal-archaeal ancestor was the removal of introns from tRNA. A less parsimonious hypothesis is that the tRNA splicing function of this enzyme evolved independently in both kingdoms from an earlier, more generalized RNA processing function. In some present-day Archaea, the tRNA splicing endonuclease can apparently also function to remove introns from ribosomal RNA (Kjems and Garrett, 1985; Kjems et al., 1989; Burggraf et al., 1993). Recently, Walter and coworkers have detected an intron in a yeast transcription factor gene that is not processed by the normal mRNA spliceosomal machinery (Cox and Walter, 1996; Sidrauski et al., 1996). They have demonstrated that tRNA ligase is involved in the splicing of this intron, but the tRNA splicing endonuclease has not yet been implicated. Clearly, members of this family of proteins have a number of functions, but it is not yet possible to determine which are original and which have been co-opted later in evolution.

Although the endonuclease active site is apparently a very primitive protein domain, it is clear that its substrate specificity has been altered in evolution because the recognition of the splice sites is different in Eucarya and Archaea. We speculate that the peculiarities of the eucaryal system are in part explained by the heterodimeric interaction of the divergent subunits—Sen2p with Sen54p and Sen34p with Sen15p. Many fascinating questions remain concerning this evolutionary specialization. Since Sen54p and Sen15p have no known homologs, where did these proteins come from? Do Sen54p and Sen15p play a role in splice site selection? Can the active site subunits function in conjunction with other protein subunits, thereby modifying endonuclease specificity for other RNA processing reactions? We are now able to begin to address these questions.

A further complexity of the evolution of the tRNA splicing process is indicated by the lack of a yeast tRNA ligase (or T4 RNA ligase) homolog in the entire M. jannaschii genome (Bult et al., 1996; C. R. T. and E. A. A., unpublished data). Interestingly, several archaeal genomes contain a clear homolog of a bacterial RNA ligase that can form a 2′-5′ linkage when joining tRNA half-molecules derived from endonucleolytic cleavage of yeast pre-tRNAs (Greer et al., 1983; Arn and Abelson, 1996). The complete ligation mechanism for archaeal tRNA splicing has not yet been determined, but it has been demonstrated that an RNA splicing ligase activity present in H. volcanii extracts does not require ATP or exogenous phosphate (R. Gupta, personal communication). The E. coli 2′-5′ RNA ligase is the only known RNA ligase activity that does not require a nucleotide triphosphate cofactor (Arn and Abelson, 1996). We are presently attempting to ascertain whether the archaeal 2′-5′ RNA ligase homolog functions in tRNA splicing in vivo.

Experimental Procedures

Yeast Methods

Yeast were transformed by the lithium acetate procedure (Gietz and Schiestl, 1995). Yeast minimal media (SD) contained Bacto yeast nitrogen base without amino acids (YNB − aa, 6.7g/l), dextrose (2% w/v), and a supplemental amino acid mix (Bio101) lacking various amino acids for auxotroph selection. All other yeast media were prepared according to Sherman et al. (1986). For fermentation growth of CRT44, SD-histidine-tryptophan, prepared as described above, was used and supplemented with 25 μg/ml Kanamycin. The fermentation was fed using SD + 10×-histidine-tryptophan amino acid mix with constant flow for 24 hr.

Two-hybrid strains were obtained by a generous gift from P. Bartel of S. Fields laboratory and are described in Table 1. All two-hybrid techniques used are described by Hannon et al. (1995). Plates used for analysis contained 5 mM 3-aminotriazole that suppressed background growth associated with histidine reporter assay (Hannon and Bartel, 1995).

Strains

E. coli strains utilized include HB101 and XL1-Blue (Novagen) and DH5 (GIBCO-BRL). S. cerevisiae strains are described in Table 1. Knockout strains YPH274Δ*SEN2* and CRT54 were created utilizing methods described in Guthrie and Fink (1991). Knockout strains of *SEN15* (CRT15) and *SEN34* (CRT34) were created utilizing methods and plasmids generously obtained from and described by Wach (1996). All knockouts were confirmed by PCR analysis, and each knockout was complemented with the corresponding ORF only.

Strain CRT44 was prepared by transforming YPH274Δ*SEN2*pC10 with YpH8/FLAG *SEN2*, plating on selective media (SD-His-Trp-Ura), and streaking out isolates on selective media (SD-His-Trp) + 5 FOA to select for loss of pC10 (URA3 CEN/ARS).

Plasmid Preparation

All restriction enzymes used were obtained from New England BioLabs and Boehringer Mannheim and used in accordance with the manufacturers' instructions. Restriction digests, DNA ligations, and electrophoresis of DNA were performed by standard methods of Sambrook et al. (1990).

pC10 was prepared by subcloning an EcoRV/EcoRI fragment of DNA derived from the YcP50 *sen2-3* complementing clone, into the vector pRS416 (URA3)(Sikorski and Hieter, 1989). This fragment

Table 1. Yeast Strains

Strain	Genotype	Source/Reference
JD51	*Mat a/α; ura3-52/ura3-52; leu2-3, 112/leu2-3, 112; trp1Δ63/trp1Δ63; lys2-801/lys2-801; his3Δ200/his3Δ200*	A. Varshavsky
YPH274	*Mat a/α; ura3-52/ura3-52; leu2Δ1/leu2Δ1; lys2-801/lys2-801trp1Δ1/trp1Δ1;; his3Δ200/his3Δ200; ade2-101°/ade2-101°*	P. Heiter
YPH274ΔSEN2	*Mat a; ura3-52; leu2Δ1; lys2-801; trp1Δ1; his3Δ200; ade2-101°; SEN2::HIS3* and YpH8/FLAG SEN2	This study
CRT44	Same as YPH274ΔSEN2 except contains YpH8/FLAG SEN2	This study
CRT54	Same as JD51 except *SEN54::HIS3/SEN54*	This study
CRT34	Same as JD51 except *SEN34::kanMX6/SEN34*	This study
CRT15	Same as JD51 except *SEN15::S. pombe HIS3/SEN15*	This study
HF7c	*Mat a; ura3-52; leu2-3, 112; lys2-801; trp1-901; his3Δ200; ade2-101°*; gal4-9542; gal80-538; LYS2::$GAL1_{UAS}$-GAL_{TATA}-*HIS3*; *URA3::*$GAL4_{17merx3}$-$CYC1_{TATA}$-*lacZ*	Feilotter et al., 1994
Y527	*Mat a; ura3-52; leu2-3, 112; lys2-801; trp1-901; his3Δ200; ade2-101°*; gal4-9542; gal80-538; URA3::$GAL4_{17merx3}$-$CYC1_{TATA}$-*lacZ*	Bartel et al., 1993

contains an open reading frame coding for the 377 amino acid *SEN2* gene plus 250 base pairs in the promoter region and 345 base pairs of downstream region. pC10 was capable of complementing a knockout of the *SEN2* ORF.

YpH8/FLAG was prepared by PCR of the *CUP1* promoter, from JD51, to allow cloning of the *CUP1* promoter into the HindIII/EcoRI site of pBluescript (Stratgene). This plasmid was then digested with XbaI, filled in, and cut with PstI, and the PstI–NaeI fragment from pGBT9 (from S. Fields), containing the *ADH1* terminator sequence, was cloned into this site, yielding pBSCUP/TERM. This plasmid was then cut with HindIII/SacI, and the resulting fragment was subcloned into pRS424 (Sikorski and Hieter, 1989), to yield YpCUP/TERM. This plasmid was cut with NdeI and EcoRI, and the HIS/FLAG linker was cloned into this site, generating YpH8/FLAG. YpH8/FLAG *SEN2* was prepared by digestion of pGBT9-*SEN2* with EcoRI and PstI, gel isolation of the fragment containing *SEN2*, and subcloning into EcoRI/PstI-digested YpH8/FLAG. YpH8FLAG *SEN34* was prepared by digestion of pGBT9-*SEN34* with EcoRI and PstI, gel isolation of the fragment containing *SEN34*, and subcloning into EcoRI/PstI-digested YpH8/FLAG. Oligonucleotide-directed mutagenesis was performed on single-stranded YpH8/FLAG-*SEN34* to effect an H242A mutation according to the protocols provided by the manufacturer (Amersham).

Two-hybrid vectors were prepared by cloning PCR products of *SEN15* (oligos *SEN15*-N 5′-GGAATTCGCAACGACAGATATCATATC-3′ and -C 5′-TTCTGCAGATCTTCAATTTCTTTTCGGTTTTCG-3′), *SEN54* (oligos *SEN54*-N 5′-GGAATTCCAATTCGCTGGGAAG-3′ and -C 5′-TTC TGCAGTTAATGCTTCTTCCATCTTTG-3′), *SEN2* (oligos *SEN2*-N 5′-GGA ATTCATGTCTAAAGGGAGGG and -C 5′-CGGGATCCCAATGAAA AGTTTAGAGC) and *SEN34* (oligos *SEN34*-N 5′-CGGAATTCGGG ATCCATATGCCACCGCTAGTATTTGAC-3′ and -C 5′-CGGGATCCC GGGAAGCTTAACCAAATCCAGCCC-3′) digested with EcoRI and BamHI and cloned into EcoRI/BamHI-digested pGBT9 and pGAD424 (Hannon and Bartel, 1995) (these constructs were sequenced).

Antibodies and Western Blots

Polyclonal antibodies to Sen34p were produced by injection of rabbit (CoCalico Biosciences) with peptides NH_2-DKTTAEELQRLDK SSC-COOH and NH_2-DIRSDSDSLSRDDIC-COOH. Polyclonal antibodies to Sen2p were produced by injection of rabbits with peptide derived from the C terminus of Sen2p (NH_2-MSKGRVNQKRYPLC-COOH). All peptides were produced by the Caltech Peptide Facility. The antibodies were used directly in rabbit serum at the indicated dilution. Polyclonal antibodies to both Sen54p and Sen15p were prepared in the following manner: antigen was derived from purification of HIS8/FLAG-tagged protein from E. coli. Each gene was cloned into the EcoRI/PstI site of YpH8/FLAG by subcloning from pGBT9 *SEN54* and pGBT9 *SEN15*. Each construct was then cut with NdeI/PstI and subcloned into Pet11a. Purification was carried out under denaturing conditions by Ni-NTA chromotography as suggested by Qiagen. Protein eluted from the Ni-NTA resin was gel purified by Prep-Cell Chromatography (Biorad) per the manufacturer's suggested protocol with elution into buffer containing 0.05% SDS. The eluate was directly injected into Swiss Webster mice per the protocol of Susan Ou of the Caltech Monoclonal Antibody Facility.

SDS–PAGE was performed on 10%–20% gradient gels (Biorad). Proteins were transferred to PVDF (Immobilon) by semidry transfer (Biorad). Western blots were performed as described by Harlow and Lane (1988) and were visualized by enhanced chemiluminescence (ECL, Amersham) (Harlow and Lane, 1988).

Endonuclease Assay

Endonuclease was assayed by incubation of 2 μl of extract for a specified amount of time at 30°C in a 10 μl reaction containing 20 mM Na-HEPES (pH 7.5), 5 mM $MgCl_2$, 2.4 mM spermidine-HCl (pH 7.5), 0.1 mM DTT, 0.4% Triton X-100, and 10 fmol pre-$tRNA^{Phe}$ body-labeled with [α-^{32}P]UTP (specific activity ~500 cpm per fmol tRNA) by T7 polymerase transcription according to Sampson and Saks (1993). RNA products were phenol/chloroform extracted, precipitated, and fractionated on a 10% polyacrylamide gel containing 8M urea. Gels were exposed to phosphorimager for 8 hr.

Purification of tRNA Endonuclease

Purification of endonuclease was carried out utilizing three separate purification strategies. The first is described by Rauhut et al. (1990). This protein was used to obtain the gene for the 34 kDa subunit.

Affinity purification was carried out utilizing yeast strain CRT44 grown by fermentation at 30°C. The cells were harvested and stored at −70°C. All subsequent steps were performed at 4°C. Yeast cells (150 g) resuspended in 1.5× breaking buffer (37.5 mM Tris [pH 8.0], 15% (v/v) Glycerol, 5 mM B-ME, 1 mM EDTA, 0.1% (v/v) Triton X-100, 405 mM NH_4SO_4) to 300 ml was ground in a Bead-Beater (BioSpec Corp) for 20×1 minute intervals, and Triton X-100 (Boehringer-Mannheim) was added to 1%. The extract was incubated for 1 hr and then centrifuged (40,000 rpm × 1.5 hr). The supernatant was loaded onto a 300 ml DEAE–Sepharose CL-6B equilibrated in 1× breaking buffer + 1% Triton X-100. DEAE effluent was diluted approximately 3-fold and loaded onto a 100 ml Heparin-HyperD (BioSepra) column equilibrated in buffer A (25 mM MOPS [pH 7.1], 10% Glycerol, 1 mM EDTA, 5 mM BME, 0.9% Triton X-100, 100 mM NaCl), at 1 ml/minute. The column was washed with 500 ml buffer B (25mM Tris [pH 8.0], 10% Glycerol, 5mM BME, 0.3% Triton X-100, 150 mM NaCl). A 2 × 200 ml gradient of 0.1–0.8 mM NaCl in buffer B was applied. Active fractions were pooled and bound in batch to 5 ml of Ni-NTA agarose (Quiagen, equilibrated in buffer B) for 1.5 hr. The suspension was poured into a 2 × 10 mm column and washed extensively with 5 ml of each of the following buffers: buffer N (25mM Tris–HCl [pH 7.5], 150 mM NaCl, 10% Glycerol, 5 mM B-mercaptoethanol, 0.3% Triton X-100) + 650 mM NaCl + 15 mM imidazole; buffer N + 650 mM NaCl + 20mM imidazole; buffer N + 20 mM imidazole; buffer N + 40 mM imidazole. Endonuclease was eluted in Buffer N + 150 mM imidazole. The eluate was loaded onto a 1 ml M2 monoclonal anti-FLAG antibody column (Kodak) equilibrated in Buffer N. The column was washed in buffer N and

eluted with Buffer N + 150 μg/ml FLAG peptide. This endonuclease was used to prepare the 54 kDa subunit.

A modified purification protocol was developed to prepare the 15 kDa subunit and purification of mutant containing endonuclease. Yeast cells (strain CRT44 or as indicated in the text) were resuspended and broken as above. The extract was then spun at 40,000 rpm × 2 hr to pellet the nuclear membranes. The pellet was resuspended in 50 ml of Buffer R (50 mM Tris [pH 7.5], 150 mM NaCl, 5% Glycerol, 1 mM EDTA, 5 mM BME, 0.9% Triton X-100) and incubated for 1 hr. The extract was subjected to centrifugation as before and the supernatant was loaded directly onto a 2.5 ml M2 monoclonal anti-FLAG antibody column. The column was washed with buffer R, and endonuclease was eluted with Buffer S (30 mM Tris [pH 8.0], 500 mM NaCl, 10% Glycerol, 5 mM BME, 0.3% Triton X-100, 5 mM Imidazole) containing 150 μg/ml FLAG peptide. The elution was incubated with 1 ml of Ni-NTA agarose, equilibrated in Buffer S, for 1.5 hr and poured into a 3 ml column. The column was washed with Buffer S + 40 mM imidazole, and endonuclease was eluted in Buffer S + 150 mM imidazole.

Protein Sequencing

Three separate procedures were utilized for sequencing the subunits of endonuclease. In all three cases, protein was submitted as a gel slice with various amounts of protein in each. For the 34 kDa subunit, 25 μg of protein was submitted to the protein sequencing facility at Cold Spring Harbor and was digested in-gel with lysyl-endoprotease, followed by extraction and HPLC purification of the resulting peptides following the protocols of the facility (Dan Marshak, personal communication). Fractions containing single peaks were sequenced by Edman degradation.

For the 54 kDa subunit, 5 μg of protein was digested in-gel with trypsin as described by Hellman et al. (1995) (Hellman et al., 1995), except that 0.3 μg of Promega modified pig trypsin and speedvac drying of the gel slice was used. Peptides were HPLC purified, and fractions containing single mass peaks, as determined by mass spectrometry on a Voyager Mass Spec, were submitted for Edman Degradation sequencing on an Applied Biosystems sequencer (Gary Hathaway, personal communication).

For the 15 kDa subunit, 5 μg of protein was submitted to the protein sequencing facility at the City of Hope (Duarte, CA) for in-gel trypsin digestion, followed by peptide sequencing utilizing a Beckman LC/MS-MS protein sequencer (K. Swiderek, personal communication). Data obtained was used to search the yeast genome database.

Acknowledgments

Correspondence regarding this paper should be addressed to J. N. A. (abelsonj@starbase1.caltech.edu). We wish to acknowledge first and foremost the protein sequencing work done by Gary Hathaway and Dirk Krapf at Caltech Protein Analysis Laboratory, Dr. Terry Lee and Dr. Kristine Swiderek at the City of Hope Protein Sequencing Facility, and Dr. Dan Marshak at the Cold Spring Harbor Protein Sequencing Facility. Special thanks to Paul Bartel and Stan Fields for two-hybrid strains and plasmids. We also want to thank Chuck Daniels for communicating the endonuclease sequences of H. volcanii and M. thermoautotrophicum prior to publication and Drs. Mel Simon and Ung-Jin Kim for sequence of the P. aerophilum homolog. We wish to thank Drs. Christine Guthrie, Eric Phizicky, Chris Greer, and Anita Hopper for their insight and critical review of the manuscript and members of the Abelson laboratory for their advice and support. This work was supported by generous grants from the American Cancer Society and National Research Service Award (to C. R. T.).

Received February 10, 1997; revised May 5, 1997.

References

Abelson, J. (1991). RNA splicing in yeast. Harvey Lect. *85*, 1–42.

Arn, E.A., and Abelson, J.N. (1996). The 2′-5′ ligase from Escherichia coli. J. Biol. Chem. *271*, 31145–31153.

Baldi, M.I., Mattoccia, E., Bufardeci, E., Fabbri, S., and Tocchini-Valentini, G.P. (1992). Participation of the intron in the reaction catalyzed by the Xenopus tRNA splicing endonuclease. Science *255*, 1404–1408.

Bartel, P.L., and Fields, S. (1995). Analyzing protein–protein interactions using 2-hybrid system. Meth. Enzymol. *254*, 241–263.

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.

Biniszkiewicz, D., Cesnaviciene, E., and Shub, D.A. (1994). Self-splicing group-I intron in cyanobacterial initiator methionine tRNA: evidence for lateral transfer of introns in bacteria. EMBO J. *13*, 4629–4635.

Bult, C.J., White, O., Olsen, G.J., Zhou, L.X., Fleischmann, R.D., Sutton, G.G., Blake, J.A., Fitzgerald, L.M., Clayton, R.A., Gocayne, J.D., 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 rRNA gene of the archaeon Pyrobaculum aerophilum. Proc. Natl. Acad. Sci. USA *90*, 2547–2550.

Cech, T. (1993). Structure and mechanism of the large catalytic RNAs: group I and group II introns and ribonuclease P. In RNA World, R.F. Gesteland and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 239–270.

Cox, J.S., and Walter, P. (1996). A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. Cell *86*, 391–404.

Culver, G.M., McCraith, S.M., Zillmann, M., Kierzek, R., Michaud, N., Lareau, R.D., Turner, D.H., and Phizicky, E.M. (1993). An NAD derivative produced during tRNA splicing: ADP-ribose 1′′-2′′ cyclic phosphate. Science *261*, 206–208.

Di Nicola Negri, E., Fabbri, S., Bufardeci, E., Baldi, M.I., Gandini Attardi, D., Mattoccia, E., and Tocchini-Valentini, G.P. (1997). The eucaryal tRNA splicing endonuclease recognizes a tripartite set of RNA elements. Cell, this issue.

Fields, S., and Song, O.K. (1989). A novel genetic system to detect protein–protein interactions. Nature *340*, 245–246.

Gietz, R.D., and Schiestl, R.H. (1995). Transforming yeast with DNA. Meth. Mol. Cell. Biol. *5*, 255–269.

Greer, C., Javor, G., and Abelson, J.N. (1983). RNA ligase in bacteria: formation of a 2′,5′ linkage by an E. coli extract. Cell *33*, 899–906.

Guthrie, C., amd Fink, G.R. (1991). Methods in Enzymology. Guide to Yeast Genetics and Molecular Biology. (San Diego, CA: Academic Press, Inc.).

Hannon, G., and Bartel, P. (1995). Identification of interacting proteins using the 2-hybrid system. Meth. Mol. Cell. Biol. *5*, 289–297.

Harlow, E., and Lane, D. (1988). Antibodies: A Laboratory Manual (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Hellman, U., Wernstedt, C., Gonez, J., and Heldin, C.H. (1995). Improvement of an in-gel digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. Anal. Biochem. *224*, 451–455.

Ho, C.K., Rauhut, R., Vijayraghavan, U., and Abelson, J. (1990). Accumulation of pre-tRNA splicing 2/3 intermediates in a Saccharomyces cerevisiae mutant. EMBO J. *9*, 1245–1252.

Kalderon, D., Roberts, B.L., Richardson, W.D., and Smith, A.E. (1984). A short amino-acid sequence able to specify nuclear location. Cell *39*, 499–509.

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

Kjems, J., Jensen, J., Olesen, T., and Garrett, R.A. (1989). Comparison Of tRNA and rRNA intron splicing in the extreme thermophile and Archaebacterium Desulfurococcus mobilis. Can. J. Microbiol. *35*, 210–214.

Kleman-Leyer, K., Armbruster, D.A., and Daniels, C.J. (1997). Properties of the H. volcanii tRNA intron endonuclease reveal a relationship between the archaeal and eucaryal tRNA intron processing systems. Cell, this issue.

Lee, M.C., and Knapp, G. (1985). tRNA splicing in Saccharomyces cerevisiae: secondary and tertiary structures of the substrates. J. Biol. Chem. *260*, 3108–3115.

Lupas, A., Vandyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162–1164.

Madhani, H.D., and Guthrie, C. (1994). Dynamic RNA–RNA interactions in the spliceosome. Annu. Rev. Genet. *28*, 1–26.

McCraith, S.M., and Phizicky, E.M. (1991). An enzyme from Saccharomyces cerevisiae uses NAD+ to transfer the splice junction 2′-phosphate from ligated tRNA to an acceptor molecule. J. Biol. Chem. *266*, 11986–11992.

Miao, F., and Abelson, J. (1993). Yeast tRNA-splicing endonuclease cleaves precursor tRNA in a random pathway. J. Biol. Chem. *268*, 672–677.

Moore, M., Query, C.C., and Sharp, P. (1993). Splicing of precursors to mRNAs by the spliceosome. In RNA World, R.F. Gesteland and J.F. Atkins, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press), pp. 303–349.

Ogden, R.C., Lee, M.C., and Knapp, G. (1984). Transfer RNA splicing in Saccharomyces cerevisiae: defining the substrates. Nucl. Acids Res. *12*, 9367–9382.

Palmer, J.R., Nieuwlandt, D.T., and Daniels, C.J. (1994). Expression of a yeast intron-containing tRNA in the Archaeon Haloferax volcanii. J. Bacteriol. *176*, 3820–3823.

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

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

Reinholdhurek, B., and Shub, D.A. (1992). Self-splicing introns in tRNA genes of widely divergent bacteria. Nature *357*, 173–176.

Reyes, V.M., and Abelson, J. (1988). Substrate recognition and splice site determination in yeast tRNA splicing. Cell *55*, 719–730.

Robbins, J., Dilworth, S.M., Laskey, R.A., and Dingwall, C. (1991). 2 interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell *64*, 615–623.

Sherman, F., Fink, G.R., and Hicks, J.B. (1986). Methods in Yeast Genetics. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sidrauski, C., Cox, J.S., and Walter, P. (1996). tRNA ligase is required for regulated mRNA splicing in the unfolded protein response. Cell *86*, 405–413.

Sikorski, R.S., and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics *122*, 19–27.

Thompson, L.D., Brandon, L.D., Nieuwlandt, D.T., and Daniels, C.J. (1989). tRNA intron processing in the halophilic Archaebacteria. Can. J. Microbiol. *35*, 36–42.

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.

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

Wach, A. (1996). PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in Saccharomyces cerevisiae. Yeast *12*, 259–265.

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.

Westaway, S.K., Phizicky, E.M., and Abelson, J. (1988). Structure and function of the yeast tRNA ligase gene. J. Biol. Chem. *263*, 3171–3176.

Winey, M., and Culbertson, M.R. (1988). Mutations affecting the tRNA-splicing endonuclease activity of Saccharomyces cerevisiae. Genetics *118*, 609–617.

Winey, M., and Culbertson, M.R. (1989). Split tRNA genes and their products: a paradigm for the study of cell-function and evolution. Yeast *5*, 405–427.

Yanagisawa, S., and Izui, K. (1993). Molecular cloning of 2 DNA-binding proteins of maize that are structurally different but interact with the same sequence motif. J. Biol. Chem. *268*, 16028–16036.

GenBank Accession Number

The accession number for the *SEN2* nucleotide sequence reported in this paper is M32336.

Appendix III

Reprint Series
10 April 1998, Volume 280, pp. 279–284

SCIENCE

Crystal Structure and Evolution of a Transfer RNA Splicing Enzyme

Hong Li, Christopher R. Trotta, and John Abelson*

Crystal Structure and Evolution of a Transfer RNA Splicing Enzyme

Hong Li, Christopher R. Trotta, John Abelson*

The splicing of transfer RNA precursors is similar in Eucarya and Archaea. In both kingdoms an endonuclease recognizes the splice sites and releases the intron, but the mechanism of splice site recognition is different in each kingdom. The crystal structure of the endonuclease from the archaeon *Methanococcus jannaschii* was determined to a resolution of 2.3 angstroms. The structure indicates that the cleavage reaction is similar to that of ribonuclease A and the arrangement of the active sites is conserved between the archaeal and eucaryal enzymes. These results suggest an evolutionary pathway for splice site recognition.

Introns are found in the tRNA genes of organisms in all three of the great lines of descent: the Eucarya, the Archaea, and the Bacteria. In Bacteria, tRNA introns are self-splicing group I introns and the splicing mechanism is autocatalytic (*1*). In Eucarya, tRNA introns are small and invariably interrupt the anticodon loop 1 base 3′ to the anticodon. They are removed by the stepwise action of an endonuclease, a ligase, and a phosphotransferase (*2*). In Archaea, the introns are also small and often reside in the same location as eucaryal tRNA introns (*3*). Splicing in Archaea is catalyzed by an endonuclease, but the mechanism of ligation is likely different from that in Eucarya as there is no homolog of the eucaryal tRNA splicing ligase in the complete genome sequence of several members of the Archaea (*4*).

The tRNA splicing endonucleases of both the Eucarya and the Archaea cleave the pre-tRNA substrate leaving 5′-hydroxyl and 2′,3′ cyclic phosphate termini, but these enzymes recognize their substrates differently. The eucaryal enzyme uses a measuring mechanism to determine the position of the universally positioned splice sites relative to the conserved domain of pre-tRNA (Fig. 1A) (*5*). In Archaea, the enzyme recognizes a pseudosymmetric substrate in which two bulged loops of 3 bases are separated by a stem of 4 base pairs (bp) [bulge-helix-bulge (BHB)] (Fig. 1A) (*6*). These observations suggested that the tRNA splicing mechanisms of the three major kingdoms evolved independently.

Recently, however, characterization of eucaryal and archaeal endonucleases has revealed a possible common origin for these enzymes. The yeast endonuclease contains four distinct subunits of 15, 34, 44, and 54 kD (*7*). The endonuclease from the archaeon *Haloferax volcanii* is a dimer of identical 37-kD subunits (*8*). The *H. volcanii* enzyme is homologous to both the 34- and 44-kD subunits of the yeast enzyme, which suggests that these distinct subunits contain active sites for tRNA splicing. This hypothesis was strongly supported by the finding that *sen2-3*, a mutant in the 44-kD subunit, is selectively defective in 5′ splice site cleavage (*9*), whereas a mutant in the 34-kD subunit, H242A, is selectively defective in 3′ splice site cleavage (*7*). Thus the 44-kD subunit cleaves the 5′ splice site and the 34-kD subunit is proposed to cleave the 3′ splice site (*7*). The function of the two nonhomologous subunits (the 15- and the 54-kD subunits) remains unclear, although the basic 54-kD subunit was suggested to embody the molecular ruler (*7*). In *H. volcanii*, the homodimeric enzyme is proposed to interact with the pseudosymmetric substrate so that each splice site is cleaved by a symmetrically

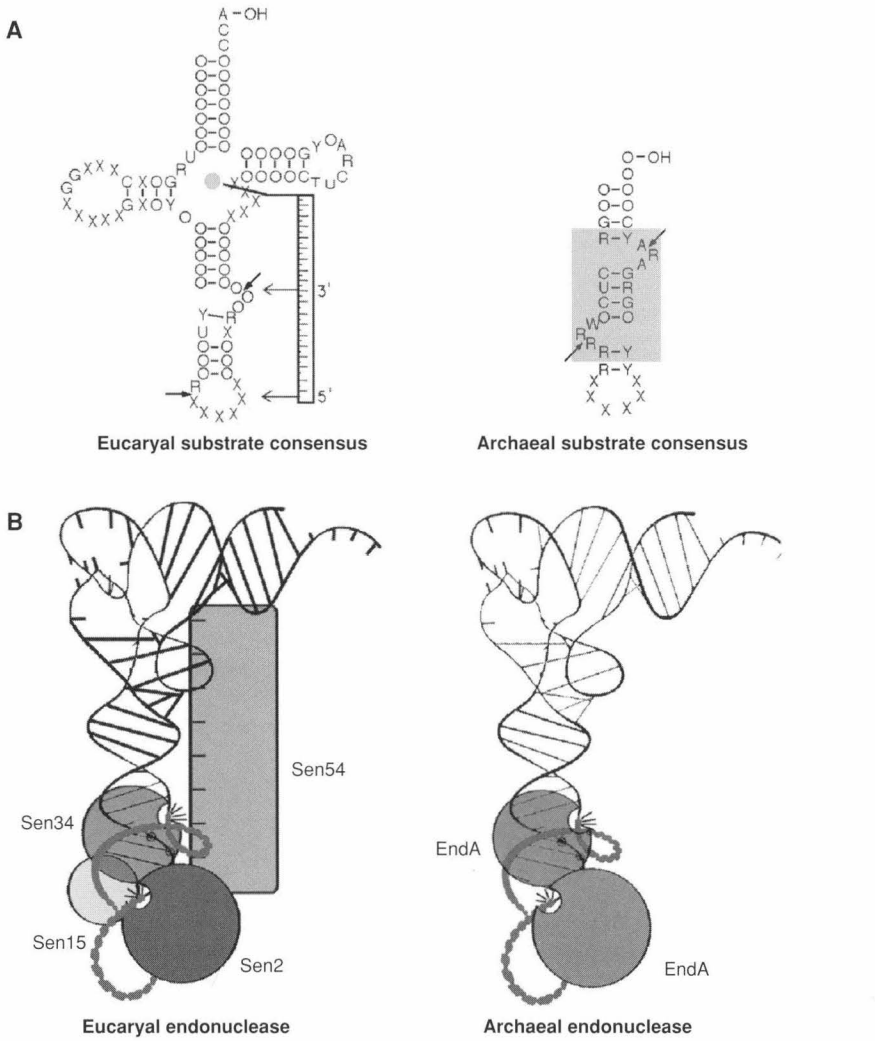


Fig. 1. (**A**) Consensus sequence and secondary structure of precursor tRNA substrate for yeast and archaeal endonucleases. Splice sites are indicated by short arrows. O and X = nonconserved bases in regions of conserved and variable secondary structure, respectively; Y = pyrimidines; R = purines. Yeast endonuclease is proposed to interact with the mature domain of the tRNA and measure the distance to the splice sites. In contrast, the archaeal enzyme recognizes two 3-nucleotide loops that are separated by a helix of 4 bp (shaded area). (**B**) Comparison of endonuclease models in Eucarya (yeast) (*7*) and Archaea (*H. volcanii*) (*8*) (see text).

Division of Biology, Mail Code 147-75, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: abelsonj@cco.caltech.edu

disposed active site monomer (Fig. 1B). Thus it is likely that the spatial disposition of the active subunits has been conserved throughout evolution.

We believed that a high-resolution structure of the archaeal tRNA endonuclease would shed light on the mechanism of the more complicated but related eucaryal endonuclease. Therefore, we determined the crystal structure of the endonuclease from the archaeon *Methanococcus jannaschii*, which is a homotetramer of 21 kD (179 amino acids) (*10*).

The *M. jannaschii* endonuclease gene with His_8 and FLAG epitope tag sequences (*7*) at its 5′ end was cloned by polymerase chain reaction from genomic DNA into the pET11a vector. Soluble protein was obtained by overexpression in *Escherichia coli* (*11*). The structure was determined by the multiwavelength anomalous diffraction (MAD) method on crystals derivatized by $Au(CN)_2$. The final model, which includes the tetrameric endonuclease of residues 9 to 179, 53 water oxygen atoms, 2 partially occupied SO_4^- ions, and 4 gold atoms, was refined to 2.3 Å with good stereochemistry to an *R* factor of 0.204 and a free *R* value of 0.268 (Table 1).

The endonuclease monomer folds into two distinct domains, the NH_2-terminal domain (residues 9 to 84) and the COOH-terminal domain (residues 85 to 179) (Fig. 2A). The NH_2-domain consists of a mixed antiparallel/parallel β sheet and three α helices. The first three β strands (β1 to β3) are antiparallel and, together with β4, are packed against two nearly perpendicular α helices, α1 and α2. The third α helix, α3, is associated with α2 via an antiparallel coiled-coil. The connection between β4 and α3 (residues 64 to 67) is disordered, which implies that this loop is flexible. This feature is consistent with the proteolytic sensitivity of this region (*10*). A short loop

Table 1. Statistics for MAD data collection and phase determination from a $Au(CN)_2$-derivatized endonuclease crystal flash cooled to 100 K (*11*). Crystals belong to the space group $P2_12_12_1$, with unit cell dimensions $a = 61.8$ Å, $b = 79.8$ Å, and $c = 192.8$ Å. Three wavelengths near the L_{III} absorption edge of Au were collected from a single crystal at National Synchrotron Light Source beamline X4A using Fuji image plates: λ_1, absorption edge corresponding to a minimum of the real part (*f′*) of the anomalous scattering factor for Au; λ_2, absorption peak corresponding to a maximum of the imaginary part (*f″*) of the scattering factor for Au; λ_3, a high-energy remote to maximize *f′*. The crystal was oriented with its *c** axis inclined slightly from the spindle axis (~16°) to avoid a blind region. The *a** axis initially was placed within the plane of the spindle and the beam direction using the STRATEGY option in the program MOSFLM (version 5.4) (*21*). A total of 180 consecutive 2° images were collected with every 8° region repeated for all three wavelengths. All data were processed with the DENZO and SCALEPACK programs (*22*). A previously collected data set at 1.5418 Å on a crystal soaked with the same $Au(CN)_2$ solutions was included as the fourth wavelength. This data set was measured with CuKα x-rays generated by a Rigaku RU200 rotating anode using a SIEMENS multiwire detector and was processed with the program XDS (*23*) followed by scaling and reducing with ROTAVATA and AGROVATA in CCP4 (*24*). All reflection data were then brought to the same scale as those of wavelength 1 with the scaling program SCALEIT followed by the Kraut scaling routine FHSCAL (*25*) as implemented in CCP4. We treated MAD data as a special case of the multiple isomorphous replacement with the inclusion of anomalous signals (*26*). The wavelength 1 (edge) was taken as the "native" data with intrinsic anomalous signals. The data collected at wavelengths 2 (peak), 3 (remote 1), and 4 (remote 2) were treated as individual "derivative" data sets. Anomalous signals are included in the data set of wavelength 2 where the x-ray absorption of Au is expected to be maximum. Because of the absence of a sharp white line at the Au L_{III} edge, the observed diffraction ratio at wavelength 2 is small, 0.061, compared with the error signal ratio of 0.036 computed from centric reflections. Two gold atoms in each asymmetric unit could be clearly located by both dispersive difference (λ_3 or λ_4 with respect to λ_1) and anomalous (λ_2) Patterson functions with the program HASSP (*27*). Two partially occupied sites were later located by difference Fourier. Initial phases were computed at 3.0 Å with the program MLPHARE as implemented in the CCP4 program suite, which uses a maximum-likelihood algorithm to refine heavy atom parameters (*28*). There are four noncrystallographically related monomers per asymmetric unit. We performed fourfold averaging in parallel with solvent flattening and histogram mapping with the program DM (*29*). The subsequent electron density map was improved markedly and the polypeptide chain could be traced unambiguously. A model representing the monomeric endonuclease containing all but the first 25 amino acids (the 8 His, the 9 FLAG epitope tag amino acids, and the first 8 amino acids in the endonuclease) was built into the electron density map with the interactive graphics program O (*30*). The assignment of side chains for the *M. jannaschii* endonuclease was confirmed by a difference Fourier map between the $Au(CN)_2$-derivatized protein and a selenomethionine-substituted protein. The final model was refined to 2.3 Å against the observed anomalous data of wavelength 3 with X-PLOR version 3.8 (*31*). Noncrystallographic restraints were applied to atoms between dimers A1-A2 and B1-B2 excluding those from loops L3 and L7. The structure was assessed with PROCHECK (*32*) and VERIFY3D (*33*) and was found to be consistent with a correct structure. More than 90% of the residues are in the most-favored regions, and no residues are in disallowed regions of a Ramachandran plot (*32*).

*Data collection statistics**

	Wavelength (Å)	Resolution (Å)	Measured reflections	Unique reflections	R_{merge}†	$\langle I\rangle/\langle\sigma(I)\rangle$	Completeness (%)	*f′* (e^-)	*f″* (e^-)	Dispersive ratio (30 to 3 Å)‡	Anomalous ratio (30 to 3 Å)‡
λ_1 (edge)	1.03997	2.3 (2.4–2.3)	165,549	75,778	0.052 (0.153)	11.3 (2.6)	89.2 (62.2)	−20.21	10.16	—	0.061 [0.041]
λ_2 (peak)	1.03795	2.3 (2.4–2.3)	161,691	75,310	0.051 (0.151)	10.6 (2.6)	89.0 (62.0)	−15.21	10.20	0.059	0.061 [0.036]
λ_3 (remote 1)	1.02088	2.3 (2.4–2.3)	164,948	76,385	0.054 (0.167)	10.3 (2.4)	90.4 (65.4)	−10.76	9.87	0.060	0.062 [0.036]
λ_4 (remote 2)	1.54180	3.0 (3.1–3.0)	50,259	16,679	0.085 (0.359)	8.7 (2.2)	83.3 (42.6)	−5.10	7.30	0.165	—

Phasing and symmetry averaging statistics | *Refinement statistics*

Phasing power			Figure of merit		Corr. coef. of NCS§ operations		Resolution (Å)	*R* value (all data)	R_{free} value (10% data)	rms deviations‖		No. of atoms in refined model			
λ_2	λ_3	λ_4	Initial	After DM	Initial	After DM				Bond (Å)	Angle (°)	Protein	H_2O	SO_4^-	Au
0.11	0.24	1.02	0.22	0.78	0.37	0.86	50.0–23	0.204	0.268	0.011	1.586	5584	53	9	4

*Numbers in parentheses correspond to those in the last resolution shell. †R_{merge} indicates agreement of individual reflection measurements over the set of unique averaged reflections. $R_{merge} = \Sigma_{h,i} |I(h)_i - \langle I(h)\rangle| / \Sigma_{h,i} I(h)_i$ where *h* is the Miller index, *i* indicates individually observed reflections, and $\langle I(h)\rangle$ is the mean of all reflections of the Miller index *h*. Bijvoet mates are treated as independent reflections when computing R_{merge}. ‡Diffraction ratios are defined as $\langle \Delta|F|^2\rangle^{1/2}/\langle |F|^2\rangle^{1/2}$ where $\Delta|F|^2$ are taken between the current and the reference wavelength (λ_1) for dispersive ratios and between matched Bijvoet pairs for anomalous ratios. Ratios for centric reflections at each wavelength are shown in brackets. §Noncrystallographic symmetry. ‖Geometry values of the final protein model were compared with ideal values from Engh and Huber (*34*).

consisting of three charged residues (Glu^{81}, Glu^{82}, Arg^{83}) links the NH_2- to the COOH-terminal domain. The COOH-terminal domain is of the α/β type and consists of a central four-stranded mixed β sheet β5-β6-$\overline{\beta 7}$-$\overline{\beta 8}$ (overbars denote parallel strands) cradled between helices α4 and α5. Two basic folding motifs lead to this fold: the rare αββ(α4-β5-β6) and the more common βαβ(β7-α5-β8). The last β strand, β9, is partially hydrogen-bonded with β8 and has little involvement in the COOH-terminal domain folding; rather, it participates in dimerization (see below).

The quaternary structure of the endonuclease is unusual in that it does not assume the D_2 point group symmetry observed in most homotetrameric proteins. When viewed along a noncrystallographic twofold axis, the homotetrameric *M. jannaschii* endonuclease resembles a parallelogram with the four subunits occupying the four corners (Fig. 2B). Subunits A1 and B1, related by a true twofold axis, are closer in space than subunits A2 and B2, which are related by the same twofold axis. Subunits A1 and A2 or B1 and B2 are related by two nonorthogonal pseudo-twofold axes plus a 1.6-Å translation along the rotation axes. The resulting dimers, A1-A2 and B1-B2, lie on two opposite sides of the parallelogram. Each of the pseudo-twofold axes relating monomers within the A1-A2 and B1-B2 dimers is at an angle of 32° with respect to the twofold axis that relates the two dimers. In addition, the three twofold axes do not intersect. Each pseudo-twofold axis is shifted about 10 Å in the opposite direction relative to the twofold axis that relates the two dimers. It is therefore more precise to describe the endonuclease as a dimer of dimers. The α carbons of the A1-A2 dimer can be superimposed with those of the B1-B2 dimer with a root-mean-square difference of 0.47 Å. The α carbons of the subunits within each dimer have a root mean square of 1.2 Å. The largest deviation is observed on loop L7, which harbors the conserved His^{125} residue. The asymmetrical organization of the four identical subunits suggests that the two subunits in the dimers have nonequivalent roles.

The endonuclease dimers A1-A2 and B1-B2 associate isologously in a tail-to-tail fashion (Fig. 2B). The dimerization interface buries extensive surface area (2422 $Å^2$) (*12*) and is formed by the interaction of the central COOH-terminal β sheets from each monomer by main chain hydrogen bonding and hydrophobic interactions. The pseudo-twofold axis that relates the two monomers lies between the two β9 strands and between the two L8 loops in each subunit. The NH_2 portion of β9 (residues169 to 172) is hydrogen-bonded to β8 by the main chain atoms. The COOH-terminal half of β9 (residues 172 to 177) bends away from the plane of the pleated central β sheet to form main chain hydrogen bonds with the symmetry-related residues of β9 in the isologous subunit (β9′); as a result, a two-stranded β sheet is produced that spans the subunit boundary. The symmetrically related L8 loops form another layer on top of the two-stranded β9 sheet and, together with the β9 sheet, enclose a hydrophobic core at the intersubunit surface. Prealbumin and enterotoxin have similar modes of subunit association (*13*). The hydrophobic core includes Phe^{139}, Leu^{141}, Leu^{144}, Gly^{146}, Val^{148}, Leu^{158}, Ile^{160}, Met^{174}, and Tyr^{176} (Fig. 2B, blue ball-and-stick model). These hydrophobic residues are important for stabilizing the dimer interface.

The interface between the two dimers is between subunits A1 and B2 and subunits B1 and A2, with each pair contributing 40% of 5420 $Å^2$ of total buried surface area (*12*). The remaining 20% of buried surface area can be accounted for by the interface between subunits A1 and B1. Subunits A2 and B2 are not in contact. The principal interaction between dimers is mediated by close contacts between polar residues of opposite charges in the two heterologously associated subunits. A positively charged cleft (9 Å wide, 12 Å long, 14 Å deep) is formed between the two domains of subunit A1 (or B1). It accommodates a protruding negative surface on the subunit B2 (or A2) formed primarily by L10 and partially by L8 (Fig. 2B, dotted surfaces). Residues from L10 and L8 that are deeply buried by this interdomain cleft are mostly acidic (Asp^{164}-Ala^{165}-Asp^{166}-Gly^{167}-Asp^{168} and Glu^{135}-Asp^{136}). Several positively charged residues (Lys^{88}, Lys^{103}, Lys^{107}, Arg^{113}) line the cleft and interact electrostatically with the acidic loop residues. The interactions between L10 and L8 and the interdomain cleft contribute nearly all the buried surface area at the tetramerization interface. In addition, three completely buried water molecules at

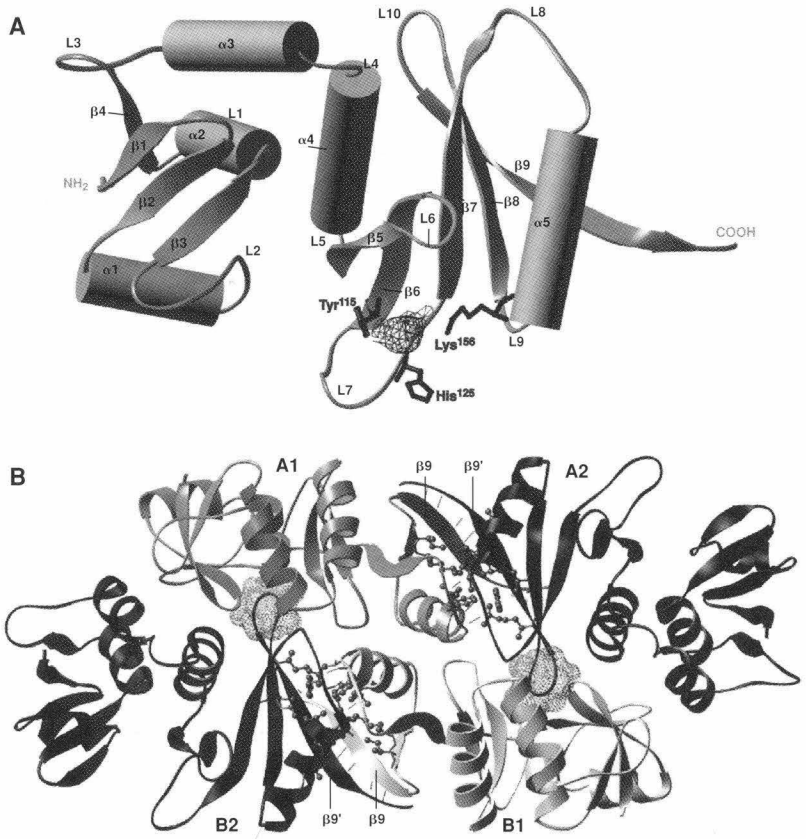


Fig. 2. (**A**) Ribbon representation of one subunit of *M. jannaschii* endonuclease obtained with the RIBBONS program (*20*). The proposed catalytic triad residues are within 7 Å of each other and are shown in red ball-and-stick models (see text). The electron density in the averaged F_o map is contoured at 5σ and is drawn only near the putative catalytic triad. (**B**) Subunit arrangement and interactions in the *M. jannaschii* endonuclease. Each subunit is represented by a distinct color and a label. The tetramer is viewed along the true twofold axis relating the A1-A2 and B1-B2 dimers. The main chain hydrogen bonds formed between β9 and β9′ and between loops L8 and L8′ for isologous dimers are drawn as thin lines. Side chains of the hydrophobic residues enclosed at the dimer interface are shown as blue ball-and-stick models. The heterologous interaction between subunits A1 and B2 (or B1 and A2) through the acidic loops L10 and L8 are highlighted by dotted surfaces.

the base of the cleft form hydrogen bonds with polar residues from both subunits, indicating that this cleft is solvent accessible before tetramerization.

The isologous dimer interaction formed by the COOH-terminal interface is expected to be more stable than the heterologous polar interaction mediated by L10 and L8, which suggests that the tetramer assembles through dimer intermediates. This interpretation is in agreement with protein cross-linking data (*10*). It is likely that the dimers observed in solution are the isologous dimers A1-A2 or B1-B2 in the crystal structure.

The tRNA splicing endonucleases generate 5′-hydroxyl and 2′,3′ cyclic phosphate termini, as do other ribonucleases (RNases) such as RNase A and T1. Cleavage by RNase A is a two-step reaction that is acid-base catalyzed (*14*). Three residues in RNase A have been identified as the catalytic triad for this reaction: His^{12} is the general base in the first step, His^{119} is the general acid, and Lys^{41} stabilizes the pentacovalent transition state.

Splicing endonucleases contain a conserved histidine residue that has been suggested to be part of the active site on the basis of mutational studies (*7, 10, 15*). In the *M. jannaschii* endonuclease structure, His^{125} is at the boundary between L7 and β7 (Fig. 2A). Although His^{125} is well ordered in the structure in all four subunits, the rest of L7 (residues 119 to 124) has weak electron density, which is suggestive of flexibility. The average B factor of L7 is 66 $Å^2$ compared with the average value of 44 $Å^2$ for the protein as a whole. Two striking structural features are observed in this region. First, there are primarily positively charged or polar residues in the vicinity of His^{125}, including two other strictly conserved residues, Tyr^{115} and Lys^{156}. The side chains of His^{125}, Tyr^{115}, and Lys^{156} are within 4 to 7 Å of each other and form a His-Tyr-Lys triad (5.5 Å between His^{125} and Lys^{156}, 6.7 Å between Lys^{156} and Tyr^{115}, 6.1 Å between Tyr^{156} and His^{125}) (Fig. 2A). Second, a distinct electron density feature (>5σ in the averaged F_o map) is found in the middle of the triad. This density is not part of the polypeptide chain and is compatible with the size of a sulfate ion that is present in crystallization and stabilization solutions.

We propose that the side chains of the His-Tyr-Lys triad in *M. jannaschii* endonuclease form the catalytic triad for the cleavage reaction. This His-Tyr-Lys triad is spatially superimposable with the catalytic triad in RNase A (*16*) and is consistent with a model in which the endonuclease His^{125} functions as the general base, Tyr^{115} functions as the general acid, and Lys^{156} functions as a stabilizer of the transition state. Although mutagenesis studies do not provide conclusive support for the essentiality of Tyr^{115}, they do indicate an essential role for both His^{125} and Lys^{156} (*4*). More studies are necessary to establish the exact catalytic mechanism and the function of these residues in catalysis. Interestingly, the inactivating Gly^{292} to Glu mutation in the yeast enzyme occurs within the aligned sequences of the flexible loop L7, which contains both His^{125} and Tyr^{115} (*7*).

Although the *M. jannaschii* endonuclease tetramer contains four active sites, it is likely that only two of these participate in cleavage. The yeast endonuclease tetramer contains two functionally independent active sites for cleavage. The *H. volcanii* endonuclease is a homodimer and thus also contains two functional active sites. The two intron-exon junctions of the archaeal tRNA substrate are related by a pseudo-twofold symmetry axis in the consensus BHB substrate. The *M. jannaschii* endonuclease is therefore predicted to contain only two functional active sites related by twofold symmetry. Consistent with this notion, upon substrate binding the endonuclease is protected from proteolytic cleavage to a maximum of 50% (*10*). Four pairs of enzyme active sites are related by twofold symmetry; these occur in subunits A1 and B1, A2 and B2, A1 and A2, and B1 and B2 (Fig. 2B). We postulate that the active sites reside in A1 and B1; their His-Tyr-Lys triads are separated by about 27 Å. The residues protected upon RNA binding cluster on A1 and B1 around the two proposed catalytic sites, falling on a path between the two sites. This region of the enzyme, formed by L9 and the carboxyl portions of α5 and β9 in A1 and B1, has a positive electrostatic potential suitable for binding the phosphodiester backbone of the pre-tRNA substrate. By contrast, the distance between the His-Tyr-Lys triads of A2 and B2 is 75 Å and there is a more scattered distribution of substrate protected residues between this pair of active sites (*17*).

We attempted to dock a model of the substrate with the tetramer. We created the model from a related structure solved by the nuclear magnetic resonance method, that of the human immunodeficiency virus trans-activation response element (TAR) RNA complexed with arginamide (*18*). TAR RNA is a stem-loop structure with a bulge of three bases in the stem (Fig. 3A). We obtained the substrate model by rotating a section of TAR RNA by 180° to create the pseudosymmetric BHB structure conforming to the archaeal endonuclease consensus substrate (Fig. 3B). This substrate model derived from the arginamide complex of TAR RNA docks well with subunits A1 and B1 in the *M. jannaschii* endonuclease tetramer (Fig. 3C). The two phosphodiester bonds that are cleaved fit precisely into the active

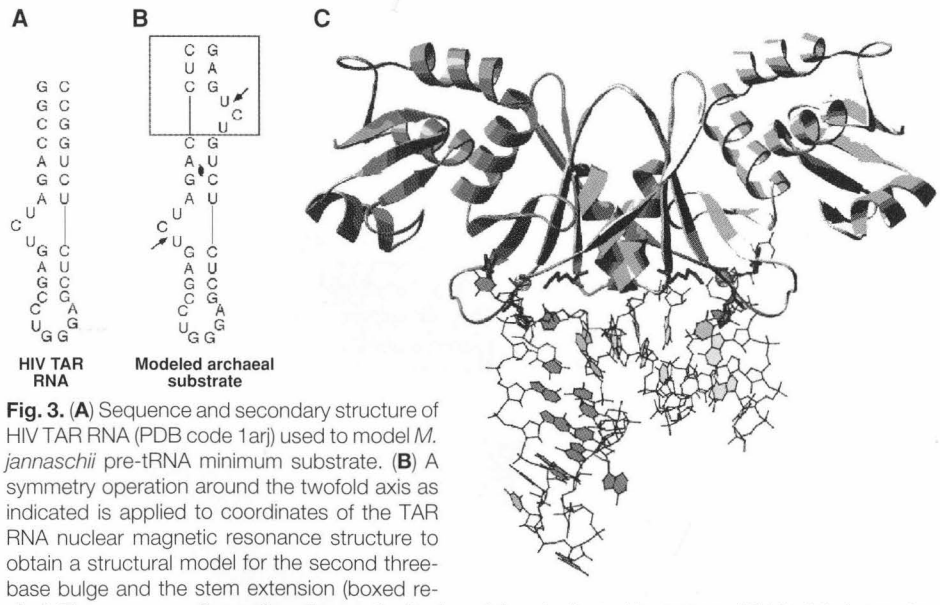


Fig. 3. (**A**) Sequence and secondary structure of HIV TAR RNA (PDB code 1arj) used to model *M. jannaschii* pre-tRNA minimum substrate. (**B**) A symmetry operation around the twofold axis as indicated is applied to coordinates of the TAR RNA nuclear magnetic resonance structure to obtain a structural model for the second three-base bulge and the stem extension (boxed region). The corresponding splice sites on the final model are indicated by arrows. (**C**) Modeled complex of *M. jannaschii* endonuclease with RNA substrate obtained by manually aligning the phosphate backbone of the 4-bp helix with the positively charged surface between subunits A1 and B1. This view is perpendicular to that of Fig. 2B. Only subunits A1 and B1, which are proposed to participate in the cleavage reaction, are shown. The proposed catalytic triad residues are shown in magenta. The phosphodiester bonds at the splice sites on RNA are depicted in blue and fit nearly perfectly with the putative sulfate density shown by blue dots. The distance from the center of the catalytic triad to the surface where RNA binds is about 7 Å.

sites of A1 and B1 and superimpose with the putative sulfate ion. The 4-bp helix separating the two bulges interacts with the basic face of the enzyme and could contact a number of residues that are protected from protease cleavage upon RNA binding (*10*).

As the *H. volcanii* endonuclease dimer and the M. *jannaschii* tetramer recognize the same consensus substrate, their active sites must be arrayed similarly in space. This proposal is strongly supported by the fact that the *H. volcanii* endonuclease monomer is in fact a tandem repeat of the consensus sequence of the endonuclease gene family (*10*) (Fig. 4B). We propose that the *H. volcanii* enzyme forms a pseudotetramer of two pseudodimers (Fig. 4B). The structure of the pseudodimer is predicted to contain a two-stranded β9-β9′ pleated sheet, an important structural feature of the M. *jannaschii* endonuclease dimer (Fig. 4A). The tandem repeat must be connected by a stretch of polypeptide (Fig. 4B, dotted line). Within each dimer of the M. *jannaschii* endonuclease structure, the NH_2-terminus of subunit A2 (or B2) and the COOH-terminus of subunit A1 (or B1) is separated by 28 Å. This requires a span of at least 10 amino acids of an extended polypeptide chain to connect the region between the NH_2-terminal and the COOH-terminal tandem repeats in the *H. volcanii* endonuclease. The *H. volcanii* enzyme model contains only two of the proposed active site triads in the COOH-terminal repeat of each pseudodimer, and these triads are proposed to occupy a spatial configuration identical to those in the A1 and B1 subunits of the M. *jannaschii* endonuclease. The pseudodimers are proposed to interact via the conserved L10 sequences. This *H. volcanii* endonuclease structural model reveals that only two of the active sites are necessary, but to array these in space correctly one must retain features of both the isologous dimer interactions (β9-β9′) and the dimer-dimer interactions mediated by L10 in the M. *jannaschii* enzyme.

The heterotetrameric yeast endonuclease is proposed to contain two active subunits, Sen2 and Sen34. The other two subunits, Sen15 and Sen54, have a sequence to the M. *jannaschii* endonuclease at the COOH-terminus (*10*). This conserved region contains the L10 residues and the hydrophobic residues at the COOH-terminus as observed in β8 and β9 in the crystal structure of the M. *jannaschii* endonuclease. We propose that the strong Sen2-Sen54 and Sen34-Sen15 interactions observed in the two-hybrid experiment (*7*) are mediated by COOH-terminal β9-β9′–like interactions (Fig. 4C) and that these dimers are associated to form the tetramer via the L10 sequences of Sen15 and Sen54 (Fig. 4C).

We conclude that the Archaea and the Eucarya have inherited from a common ancestor an endonuclease active site and the means to array two sites in a precise and conserved spatial orientation. This conclusion is supported by the results of Fabbri *et al.* (*19*), which demonstrate that both the eucaryal and the archaeal endonucleases can accurately cleave a universal substrate containing the BHB motif. The eucaryal enzyme appears to dispense with the more complex ruler mechanism for tRNA substrate recognition when it cleaves the universal substrate. Thus the precise positioning of two active sites in endonuclease appears to have been conserved in evolution. Subunits A1 and B1 comprise the active site core of all tRNA splicing endonucleases, and subunits A2 and B2 position the two active sites precisely in space. The eucaryal enzyme has evolved a distinct measuring mechanism for splice site recognition by specialization of the A2 and B2 subunits, and it has retained the ability to recognize and cleave the primitive consensus substrate.

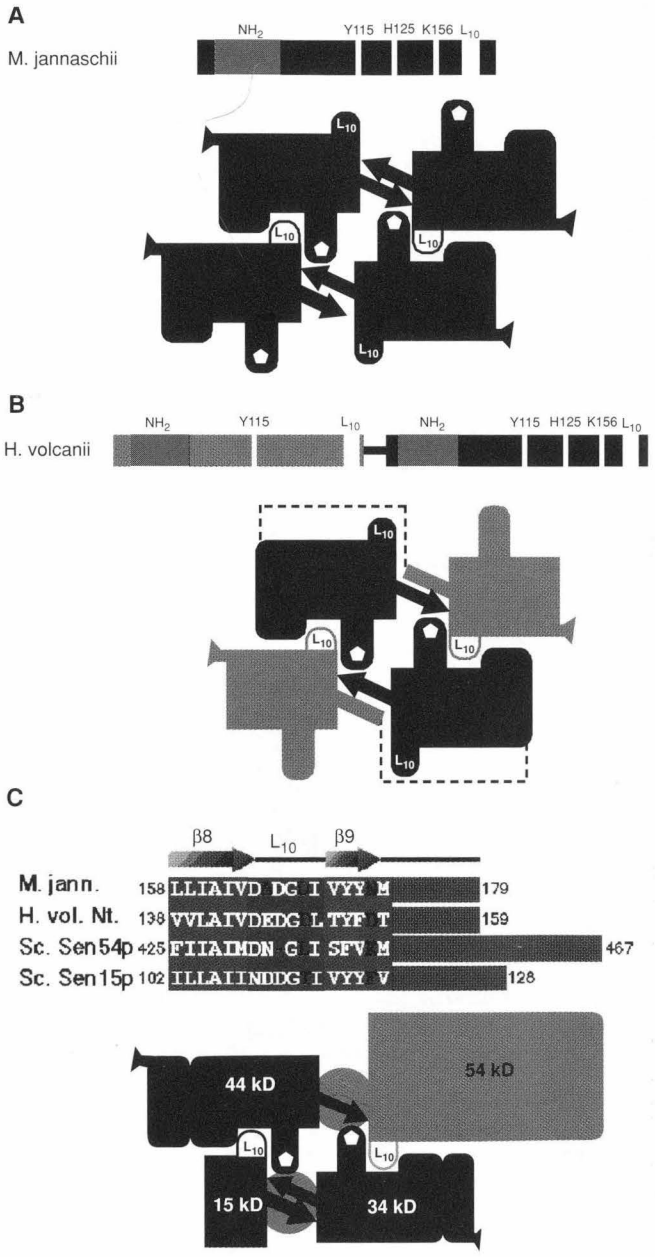


Fig. 4. (**A**) Sequence features and subunit arrangement of the *M. jannaschii* homotetramer. Several important structural features discussed in the text are indicated: loop L10, the COOH-terminal β9 strands (arrows), and the conserved catalytic residue His^{125} (pentagon). (**B**) Sequence features conserved between *H. volcanii* and *M. jannaschii* endonucleases and the proposed subunit arrangement of the former. The two tandem repeats are more similar to the *M. jannaschii* endonuclease sequence than to each other. The NH_2-terminal repeat lacks two of the three putative active site residues (white bars). It does, however, contain many of the features of the COOH-terminal domain that are important for structural arrangement of the enzyme, in particular the L10 sequence (yellow bars). The COOH-terminal repeat contains all the sequence features of the *M. jannaschii* enzyme. Dashed lines represent the polypeptide chain connecting the COOH- and NH_2-terminal repeats. (**C**) Proposed structural model for the yeast endonuclease. Conserved amino acid sequences near the COOH-termini of archaeal enzymes *M. jannaschii* (M. jann.), *H. volcanii* NH_2-terminal repeat (H. vol. Nt), and yeast Sen54 (Sc. Sen 54p) and Sen15 (Sc. Sen15p) subunits are aligned. Important hydrophobic residues that stabilize the isologous COOH-terminus interaction between *M. jannaschii* subunits A1 and A2 (or B1 and B2) are highlighted in green and circled on the structural models of heterodimers derived from the two-hybrid matrix analysis (*7*). The sequences of Sen54 and Sen15 aligned with L10 sequences in *M. jannaschii* and *H. volcanii* are highlighted in red. Loops L10 on both Sen54 and Sen15 are labeled on the proposed heterotetramer model of the yeast endonuclease. Abbreviation for amino acids are as follows: L, Leu; I, Ile; A, Ala; V, Val; D, Asp; G, Gly; Y, Tyr; N, Asn; M, Met; E, Glu; T, Thr; F, Phe; W, Trp; S, Ser; K, Lys.

REFERENCES AND NOTES

1. D. Biniszkiewicz, E. Cesnaviciene, D. A. Shub, *EMBO J.* **13**, 4629 (1994); B. Reinholdhurek and D. A. Shub, *Nature* **357**, 173 (1992).
2. S. K. Westway and J. Abelson, in *tRNA: Structure, Biosynthesis, and Function*, D. Söll and U. Raj Bhandary, Eds. (American Society for Microbiology, Washington, DC, 1995), pp. 79–92; M. R. Culbertson and M. Winey, *Yeast* **5**, 405 (1989); E. M. Phizicky and C. L. Greer, *Trends Biochem. Sci.* **18**, 31 (1993).
3. J. R. Palmer, T. Baltrus, J. N. Reeve, C. J. Daniels, *Biochim. Biophys. Acta* **1132**, 315 (1992); J. Kjems, J. Jensen, T. Olesen, R. A. Garrett, *Can. J. Microbiol.* **35**, 210 (1989).
4. C. R. Trotta, unpublished results.
5. V. M. Reyes and J. N. Abelson, *Cell* **55**, 719 (1988).
6. L. D. Thompson and C. J. Daniels, *J. Biol. Chem.* **265**, 18104 (1990).
7. C. R. Trotta *et al.*, *Cell* **89**, 849 (1997).
8. K. Kleman-Leyer, D. A. Armbruster, C. J. Daniels, *ibid.*, p. 839.
9. C. K. Ho, R. Rauhut, U. Vijayraghavan, J. N. Abelson, *EMBO J.* **9**, 1245 (1990).
10. J. Lykke-Andersen and R. A. Garrett, *ibid.* **16**, 6290 (1997).
11. Recombinant endonuclease was purified from *E. coli* by a Ni-NTA (Qiagen) affinity chromatography step after heat denaturation. Crystals were grown at 22°C by vapor diffusion using KCl as a salting-in precipitant with a starting concentration of 1 M. The well solutions were at pH 5.5 to 6.0 and contained 20 to 80 mM $(NH_4)_2SO_4$. Crystals were soaked for 3 days in 20 mM sodium cacodylate (pH 5.5 to 6.0), 20 mM $(NH_4)_2SO_4$, 25% glucose, 2 mM $Au(CN)_2$ before data collection.
12. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971).
13. C. C. F. Blake *et al.*, *ibid.* **88**, 1 (1974); T. K. Sixma *et al.*, *Nature* **351**, 371 (1991).
14. C. Walsh, *Enzymatic Reaction Mechanisms* (Freeman, San Francisco, 1979).
15. C. J. Daniels, personal communication.
16. The catalytic triad (residues His^{12}, His^{119}, and Lys^{41}) of RNase A was brought manually to superimpose with Tyr^{115}, His^{125}, Lys^{156} of *M. jannaschii* endonuclease. The best superimposition was obtained when Tyr^{115} aligns with His^{119}, His^{125} with His^{12}, and Lys^{156} with Lys^{41}.
17. H. Li, C. R. Trotta, J. N. Abelson, data not shown.
18. J. D. Puglisi, L. Chen, A. D. Frankel, J. R. Williamson, *Proc. Nat. Acad. Sci. U.S.A.* **90**, 3680 (1993); F. Aboul-ela, J. Karn, G. Varani, *Nucleic Acids Res.* **24**, 3974 (1996).
19. S. Fabbri *et al.*, (1998) *Science*, in press.
20. M. Carson, *J. Mol. Graphics* **5**, 103 (1987).
21. A. G. W. Leslie, *CCP4 AND ESF-EACMB Newsl. Protein Crystallogr.* **32**, 2 (1996).
22. Z. Otwinowski, in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing, 29–30 January 1993*, L. I. Sawyer, N. Isaac, S. Bailey, Eds. (Science and Engineering Research Council, Daresbury Laboratory, Warrington, UK, 1993), pp. 56–62; W. Minor, *XDISPLAYF Program* (Purdue University, West Lafayette, IN, 1993).
23. W. Kabsch, *J. Appl. Crystallogr.* **21**, 916 (1988).
24. Collaborative Computational Project, Number 4, "The CCP4 Suite: Programs for Protein Crystallography," *Acta Crystallogr. D* **50**, 760 (1994).
25. J. Kraut *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **48**, 1417 (1962).
26. V. Ramakrishnan and V. Biou, *Methods Enzymol.* **276**, 538 (1997).
27. T. C. Terwilliger, S. H. Kim, D. E. Eisenberg, *Acta Crystallogr. A* **43**, 34 (1987).
28. Z. Otwinowski, in *Isomorphous Replacement and Anomalous Scattering*, W. E. Wolf, P. R. Leslie, A. G. W. Leslie, Eds. (Science and Engineering Research Council, Daresbury Laboratory, Warrington, UK, 1991), p. 80.
29. K. Cowtan, *Joint CCP4 ESF-EACBM Newsl. Protein Crystallogr.* **31**, 34 (1994).
30. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr. A* **47**, 110 (1991).
31. A. T. Brünger, *X-PLOR Version 3.1, A System for Crystallography and NMR* (Yale Univ. Press, Department of Molecular Biophysics and Biochemistry, New Haven, CT, 1996).
32. R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
33. R. Luthy, J. U. Bowie, D. Eisenberg, *Nature* **356**, 83 (1992).
34. R. Engh and R. Huber, *Acta Crystallogr. A* **47**, 392 (1991).
35. We thank D. Graham and C. Woese for *M. jannaschii* genomic DNA, P. Bjorkman and D. Rees for their helpful discussions and generosity in sharing their equipment, C. M. Ogata for x-ray beam time allocation, F. T. Burling and S. Diana for assistance in data collection, R. Story and M. Saks for critical review of the manuscript, and other members of the Abelson laboratory for their advice and support. This work was supported by American Cancer Society grant NP802 and NIH Individual National Research Service Award F32 GM188930-01. PDB code for the coordinates is 1a79.

10 December 1997; accepted 13 February 1998