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

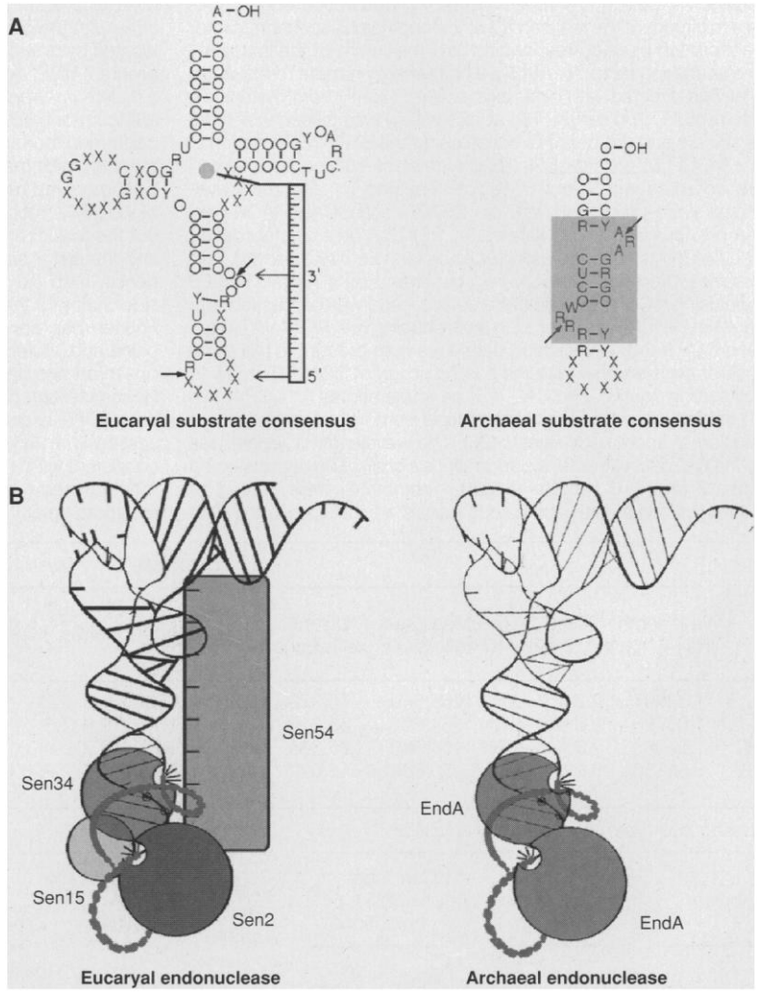

Fig. 1. (**A**) Consensus sequence and secondary structure of precursor tRNA substrate for yeast and archaeal endonucleases. Splice sites are indicated by short arrows. O and X = nonconserved bases in regions of conserved and variable secondary structure, respectively; Y = pyrimidines; R = purines. Yeast 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

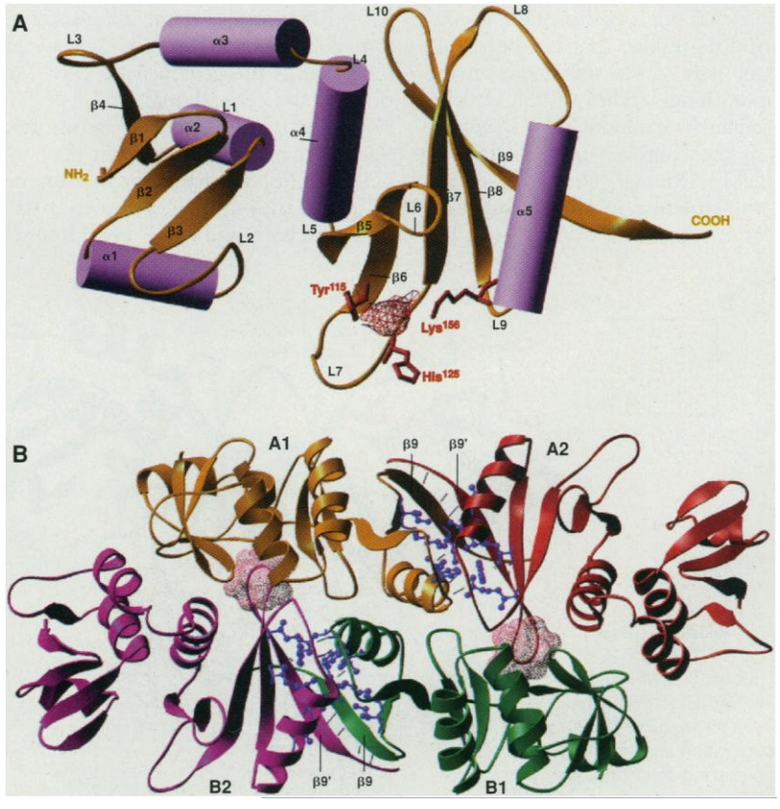

Fig. 2. (**A**) Ribbon representation of one subunit of *M. jannaschii* endonuclease obtained with the RIBBONS program (*20*). The proposed catalytic triad residues are within 7 Å of each other and are shown in red ball-and-stick models (see text). The electron density in the averaged F_o map is contoured 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

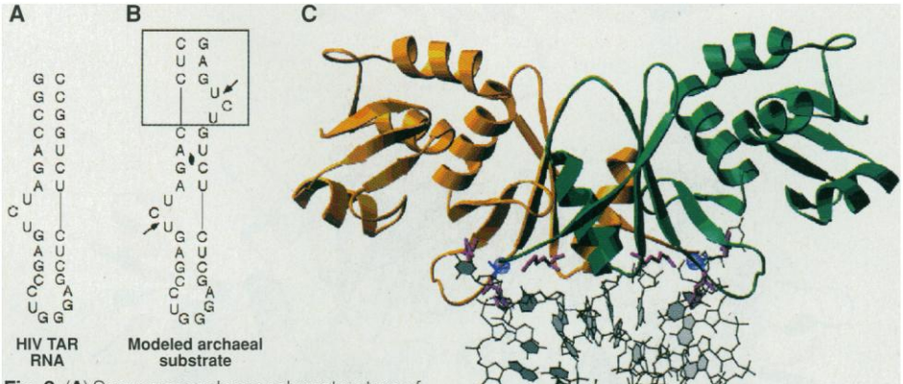

Fig. 3. (**A**) Sequence and secondary structure of HIV TAR RNA (PDB code 1arj) used to model *M. jannaschii* pre-tRNA minimum substrate. (**B**) A symmetry operation around the twofold axis as indicated is applied to coordinates of the TAR RNA nuclear magnetic resonance structure to obtain a structural 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.

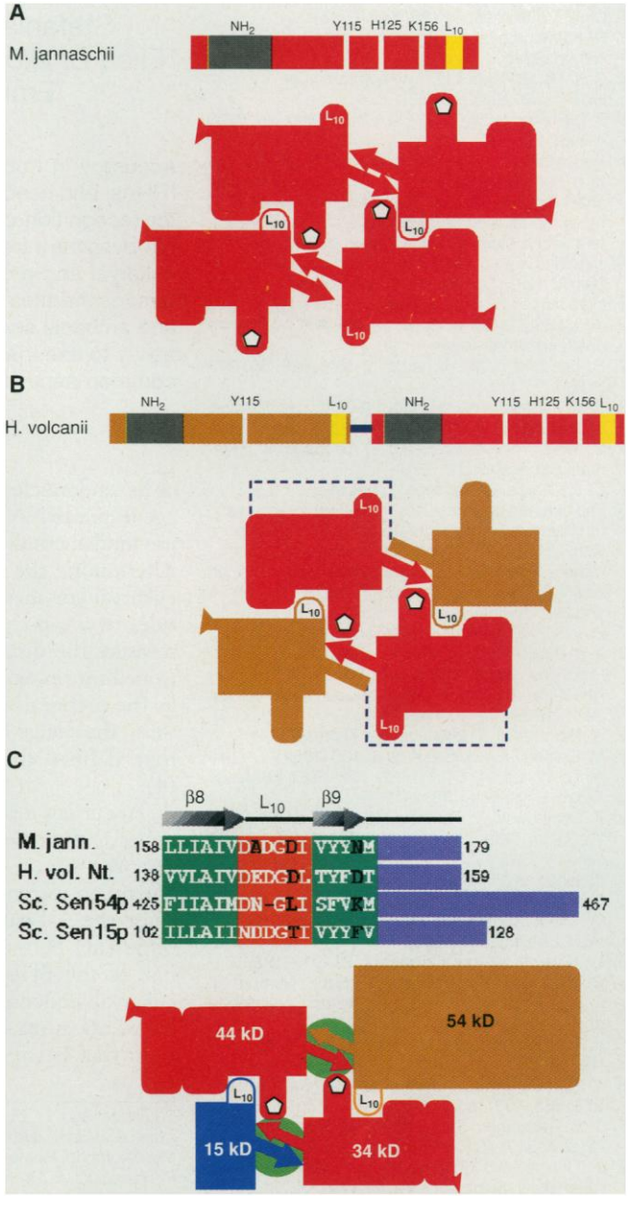

Fig. 4. (**A**) Sequence features and subunit arrangement of the *M. jannaschii* homotetramer. Several important structural features discussed in the text are indicated: loop L10, the COOH-terminal β9 strands (arrows), and the conserved catalytic residue His^{125} (pentagon). (**B**) Sequence features 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

Conservation of Substrate Recognition Mechanisms by tRNA Splicing Endonucleases

Stefania Fabbri, Paolo Fruscoloni, Emanuela Bufardeci, Elisa Di Nicola Negri, Maria I. Baldi, Domenica Gandini Attardi, Emilio Mattoccia, Glauco P. Tocchini-Valentini*

Accuracy in transfer RNA (tRNA) splicing is essential for the formation of functional tRNAs, and hence for gene expression, in both Eukaryotes and Archaea. The specificity for recognition of the tRNA precursor (pre-tRNA) resides in the endonuclease, which removes the intron by making two independent endonucleolytic cleavages. Although the eukaryal and archaeal enzymes appear to use different features of pre-tRNAs to determine the sites of cleavage, analysis of hybrid pre-tRNA substrates containing eukaryal and archaeal sequences, described here, reveals that the eukaryal enzyme retains the ability to use the archaeal recognition signals. This result indicates that there may be a common ancestral mechanism for recognition of pre-tRNA by proteins.

The endonucleases responsible for cleavage of pre-tRNAs during splicing appear to use fundamentally different mechanisms for determining the sites of cleavage (*1*). The eukaryal enzymes bind to invariant nucleotides in the body of the mature RNA and measure the distance to equivalently positioned intron-exon junctions (*2*, *3*), whereas the archaeal enzymes recognize a structure, the bulge-helix-bulge (BHB) motif, that defines the intron-exon boundaries (*4*).

A curious feature of the eukaryal enzymes is their dependency on base-pairing near the site of 3′ cleavage, the so-called anticodon-intron pair (A-I pair) (*5*). Because the A-I pair generates a three-nucleotide (nt) bulged structure that resembles half of the BHB, we studied whether the eukaryal endonucleases retained elements of an ancestral RNA recognition mechanism that is still used by the archaeal enzymes. In general, archaeal pre-tRNAs are not recognized by the eukaryal endonuclease, and vice versa (*4*).

We designed and constructed a hybrid pre-tRNA molecule, pre-$tRNA^{Archeuka}$, that is a substrate for both the eukaryal and archaeal endonucleases. This molecule consists of two regions derived from yeast pre-$tRNA^{Phe}$ (nt 1 to 31 and nt 38 to 76) joined by a 25-nt insert that corresponds to the BHB motif of the archaeal pre-tRNAs (Fig. 1). It has a typical eukaryal mature domain with putative cleavage sites located at the prescribed distance from the reference elements as well as a correctly positioned A-I base pair, all of which should ensure correct recognition by the eukaryal endonuclease. In addition, the presence of the BHB motif should confer substrate characteristics that are recognizable by the archaeal enzyme. Indeed, we found that precise excision of the intron was catalyzed by the eukaryal *Xenopus* and yeast endonucleases and by the archaeal *Sulfolobulus sulfataricus* endonuclease (Fig. 1, lanes 3, 7, and 11). Thus, pre-$tRNA^{Archeuka}$ is a universal substrate.

Because the reference sites for eukaryal endonucleases are in the mature domain of the tRNA (*2*, *3*, *5*), insertion of two base pairs into the anticodon stem, to generate pre-$tRNA^{Phe}_{v}$, increases the size of the

S. Fabbri and E. Bufardeci, EniChem, Istituto Guido Donegani SpA, Laboratori di Biotecnologie, 00015 Monterotondo, Rome, Italy.
P. Fruscoloni, E. Di Nicola Negri, M. I. Baldi, D. Gandini Attardi, E. Mattoccia, Istituto di Biologia Cellulare, CNR, Viale Marx 43, 00137 Rome, Italy.
G. P. Tocchini-Valentini, Istituto di Biologia Cellulare, CNR, Viale Marx 43, 00137 Rome, Italy, and Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: glauco@biocell.irmkant.rm.cnr.it