Substrate Requirements for a Novel Archaeal Endonuclease That Cleaves Within the 5′ External Transcribed Spacer of *Sulfolobus acidocaldarius* Precursor rRNA

Anthony G. Russell, Holger Ebhardt and Patrick P. Dennis

Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada

Manuscript received March 19, 1999
Accepted for publication May 5, 1999

ABSTRACT

During ribosome biogenesis in the hyperthermophilic archaeon *Sulfolobus acidocaldarius*, at least three separate precursor endonucleolytic cleavages occur within the 144-nucleotide-long 5′ external transcribed spacer (5′ ETS) region of the rRNA operon primary transcript. The 5′ ETS sequence contains three regions of very stable helical structure. One cleavage (5′ to position −98) is in the single-stranded region between the 5′ and the central helical domains; a second cleavage (5′ to position −31) is in the single-stranded region between the central and the 3′ helical domains; and a third cleavage is at the 5′ ETS-16S junction (5′ to position +1). The three sites share a common consensus sequence around the position of cleavage. We have used an *in vitro* pre-RNA processing assay to define some of the sequence and structural recognition elements necessary for the two precursor cleavages 5′ to positions −98 and −31. Surprisingly, none of the three predominant helical domains are required for recognition or targeting of the cleavages, although their removal reduces the rate of cleavage site utilization. We show that the sequence AAG ↓ (CA)UU encompassing each site contains at least some of the essential features for recognition and efficient targeting of the cleavages. Cleavage depends on the presence of a purine 5′ and a uracil two nucleotides 3′ to the scissile phosphodiester bond. Mutations to other bases at these critical positions are either not cleaved or cleaved very poorly. Finally, on the basis of intermediates that are produced during a processing reaction, we can conclude that the cleavages at positions 98 and 31 are not ordered *in vitro*.

THE processing, folding, and maturation of precursor (pre) rRNA and the assembly of small and large ribosomal subunits are complex and highly regulated processes. In almost all organisms, small and large subunit rRNAs are cotranscribed and generated by complex processing of a long primary transcript. The bacterial and archaeal pre-rRNA transcripts generally contain inverted repeats surrounding the 16S and 23S RNA sequences that form extended helical structures and contain the sites for the initial endonucleolytic cleavage and excision of pre-16S and pre 23S from the primary transcript (Dunn and Studier 1973; Young and Steitz 1978; Gegenheimer and Apirion 1981; Chant and Dennis 1986; Dennis 1991; Garrett *et al.* 1991; Durovic and Dennis 1994). In *Escherichia coli* (and presumably other eubacteria), the normal excision endonuclease is the helix-specific RNase III (Court 1993; Nicholson 1996). This enzyme cleaves ~11 bp from the base of a helix, generally cleaves on both strands in a coordinate manner, and exhibits only weak sequence and structural specificity at or near the cleavage sites. Maturation at the 5′ and 3′ ends of 16S and 23S rRNA has not been studied in detail, although polynucleotide phosphorylase and RNase PH have been indirectly implicated in 3′ end trimming (Zhou and Deutscher 1997).

To date, only two endoribonucleases have been described from Archaea. The first is RNase P, the RNA-containing enzyme that carries out maturation at the 5′ end of tRNA, and the second is the bulge-helix-bulge (BHB) endonuclease, the enzyme that excises introns from the transcripts of intron-containing rRNA and tRNA genes (Thompson and Daniels 1988; Kjems *et al.* 1989; Thompson *et al.* 1989; La Grandeur *et al.* 1993). The BHB enzyme from *Haloferax volcanii* has been extensively characterized and shown to have a stringent substrate requirement consisting of two 3-base bulges on opposite strands of an extended helix and separated by 4 bp (Thompson *et al.* 1989; Kleman-Leyer *et al.* 1997). Both a crystal structure of the homologous enzyme from *Methanococcus jannaschii* and an NMR structure of a BHB substrate have been determined (Diener and Moore 1998; Li *et al.* 1998). The BHB motif has also been observed within most of the repeat sequences that surround 16S and 23S genes in archaeal rRNA operons (Hui and Dennis 1985; Dennis 1991; Garrett *et al.* 1991). *In vivo* analysis of rRNA-processing intermediates indicates that these BHB motifs, when present, are used for the excision of pre-16S or pre-23S rRNA from the rRNA operon primary transcript (Chant and Dennis 1986; Dennis *et al.* 1998).

In eukaryotic organisms pre-rRNA synthesis and pro-

Corresponding author: Patrick P. Dennis, Department of Biochemistry and Molecular Biology, University of British Columbia, 2146 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada.
E-mail: patrickp.dennis@ubc.ca

cessing and ribosomal subunit assembly occur in a specialized compartment, the nucleolus (reviewed by Gerbi 1995; Maxwell and Fournier 1995). The pre-rRNAs lack the extended processing helices found in typical bacterial and archaeal transcripts and utilize ribonucleoprotein (RNP) complexes containing a large collection of small nucleolar (sno) RNAs to cleave, modify, fold, and assemble rRNAs into mature ribosomal subunits (Venema and Tollervey 1995; Kiss-László *et al.* 1996; Ni *et al.* 1997). Many of the sno-RNAs associate with the essential and abundant nucleolar protein fibrillarin (Balakin *et al.* 1996). Interestingly, archaea possess a fibrillarin-like protein (Amiri 1994) but its role in archaeal ribosome biogenesis has not yet been established.

The genome of the hyperthermophilic archaeon *Sulfolobus acidocaldarius* contains a single, eucaryotic-like 16S-23S rRNA operon that lacks spacer tRNAs genes; the 5S rRNA gene is unlinked and separately transcribed (Olsen *et al.* 1985; Durovic and Dennis 1994; Durovic *et al.* 1994). Both the 16S and 23S genes are flanked by long inverted repeats, which, through long-range interactions, are capable of forming helical stems within the primary rRNA transcript. The 23S stem contains the standard BHB motif and appears to be a substrate for the archaeal-specific BHB endonuclease; cleavages on opposite strands within the bulge regions liberate pre-23S rRNA from the primary transcript and allow subsequent maturation and assembly into 50S subunits. The 16S processing stem contains an aberrant BHB motif (Figure 1C) that is not expected to be recognized by the BHB endonuclease (on the basis of the biochemical characterization of the *H. volcanii* enzyme; Thompson *et al.* 1989). In a previous study, we detected two *in vivo* cleavages within the 5′ ETS region at positions −98 (site 1) and −31 (site 2) of the *S. acidocaldarius* rRNA operon primary transcript (Durovic and Dennis 1994; Potter *et al.* 1995). To elucidate the role of these cleavages in 16S rRNA processing, we developed an *in vitro* processing system that utilized cell-free extract to cleave a 5′ ETS-containing substrate at the two positions (Potter *et al.* 1995). Both cleavages appear to be mediated by a novel archaeal endonuclease that has been extensively purified (A. G. Russell and P. P. Dennis, unpublished results). Precursor cleavages are important in ribosome biogenesis. In this study, we have analyzed various deletions and single nucleotide (nt) substitutions within the 5′ ETS substrate RNA in order to define the substrate features that are recognized by this pre-rRNA processing activity. Our results show that the primary sequence around the cleavage site is required for efficient and accurate cleavage; three domains of secondary structure within the 5′ ETS substrate are not required for cleavage but their presence enhances the rate of cleavage.

MATERIALS AND METHODS

Substrate RNAs: The plasmid pPD1157α contains an insert extending from position −149 to +72 in the 16S gene of the single-copy *S. acidocaldarius* rrn operon (Figure 1) in the vector pGEM3Zf(+). *In vivo*, the site of pre-rRNA transcription initiation is residue −144. This construct has an SP6 RNA polymerase promoter followed by polylinker sites for *Hin*dIII and *Sph*I, the remnant of a *Pst*I restriction site, and an insert segment ending with an *Sma*I site that is followed by an *Eco*RI site (see Figure 1). The *Sph*I-*Sma*I fragment from pPD1157α was recloned between the *Sph*I and *Hin*cII sites of pGEM3Zf(+) to create pPD1157β. The 5′ deletion substrate 5′Δ-61 was generated by more extensive unidirectional exonuclease III digestion of the progenitor to pPD1157α (Durovic and Dennis 1994). It differs from pPD1157α only in lacking nucleotides between positions −149 and −62. The 3′ deletions of plasmid pPD1157β (3′Δ-16 and 3′Δ-67) were created by unidirectional exonuclease III digestion from the *Bam*HI site and use of the *Kpn*I site for protection according to the protocol as specified in Promega's Erase-a-Base system (Promega, Madison, WI). All 5′ deletion mutants that remove 5′ ETS sequence to position −101 were generated by excising the *Hin*dIII-*Hin*dIII fragment (−162 to −102) either from pPD1157α or from pPD1157β 3′ deletion derivatives. The 3′ deletion substrates ending at position +6 were generated by SP6 transcription of plasmids linearized with *Msp*I, which cleaves at position 6 within the 16S RNA sequence.

Site-specific nucleotide substitutions were generated from construct pPD1157γ. This was created by cloning an *Sph*I-*Eco*RI fragment from pPD1157β into the same restriction sites in pSELECT-1. This maintained identical 5′ and 3′ multiple-cloning-site sequences flanking the insert. Promega's Altered Sites system was employed to create the substitutions using degenerate mutagenic oligonucleotides at positions (−101), (−100), (−99) (−98), (−97), (−96), (−32), and (−32 and −31). The four-nucleotide internal duplication around site 1 was generated by cloning an *Sph*I-*Sac*I fragment from pPD1157β between the same sites in pGEM5Zf(+), which lacks a multiple cloning *Hin*dIII restriction site. This construct was then linearized with *Hin*dIII, filled in with Klenow Enzyme, and religated. An *Sph*I-*Sac*I fragment from this construct was recloned between the same sites in pGEM3Zf(+) to create pPD1157β(ins4).

The *in vitro* transcripts were synthesized from DNA templates linearized with *Eco*RI for pPD1157α and *Bam*HI for pPD1157β and for all site-specific nucleotide substitutions generated in pPD1157γ. To generate 3′ deletion substrates ending at position +6 within the 16S rRNA, the template plasmids were linearized with *Msp*I. All template DNAs containing other 3′ deletions were linearized with *Eco*RI. Run-off transcripts were then made using SP6 RNA polymerase (according to the protocol in Promega's *Protocols and Applications Guide*, Ed. 2) with [α-^{32}P]CTP for uniformly radiolabeled substrates or CTP for nonradiolabeled substrates for Northern hybridization or primer-extension analysis (Mackie 1986; Potter *et al.* 1995). Transcripts were extracted with phenol/chloroform and precipitated with ethanol before use in the pre-rRNA processing assay. Transcripts produced by SP6 RNA polymerase often contain a nontemplate nucleotide at the 3′ end. This can be visualized as a doublet in the 3′ distal products in some of the processing reactions. In some experiments a nonspecific 219-nt-long control RNA was used. The RNA was transcribed with SP6 RNA polymerase from a plasmid linearized with *Bam*HI. The plasmid contains a 178-nt-long insert into the polylinker *Hin*dIII site of pSP72; 168 nt of the insert are derived from the *E. coli* S20 ribosomal protein gene. The transcript, containing two internal *Hin*dIII sites, is antisense with respect to the S20 coding region.

***In vitro* processing of normal and mutant pre-rRNA substrates:** Midlog phase cultures of *S. acidocaldarius* grown at 78°–80°, pH 3.7 (Brock *et al.* 1972), were rapidly cooled to 0° and harvested by centrifugation. Cell pellets were resus-

pended in one-tenth volume of buffer (50 mm TRIS, pH 8.0, 15 mm $MgCl_2$, 1 mm EDTA, 1 mm DTT) and disrupted by two bursts of 30-sec sonication using a Heat Systems-Ultrasonics model W185D Sonifier cell disruptor with an ultramicrotip. Cell debris was removed by centrifugation and ammonium sulfate was added to the supernatant to 35% w/v. The pellet obtained after centrifugation was redissolved in 50% glycerol and used as the source of processing activity. A typical 35% ammonium sulfate fraction contained ~0.06 μg RNA and 0.65 μg protein per μl. In some experiments, a more highly purified preparation of the activity was used, prepared by successive glycerol gradient centrifugations followed by ion exchange chromatography; details of the purification and characterization of the endonuclease activity will be described elsewhere.

Assays were performed by mixing 1 μl of 35% ammonium sulfate fraction (or in some cases more highly purified material), 4 μl of reaction buffer (100 mm Tris acetate, pH 7.5, 100 mm magnesium acetate, 500 mm potassium acetate), and 13 μl of H_2O and equilibrating the mixture at 75° for 5 min. The reaction was initiated by the addition of 2 μl (~50 ng) of radiolabeled substrate RNA. Incubation was continued for various time intervals before the addition of 80 μl of an ice-cold stop solution (0.1 m EDTA, pH 8.0, 2.5 m ammonium acetate; 6.0 μg/μl yeast RNA). Processing products were then extracted with phenol/chloroform, precipitated with ethanol, and separated on an 8% denaturing polyacrylamide-urea gel. The products were visualized by autoradiography. A molecular length size standard generated by end-labeling the *Msp*I fragments of pBR322 with the Klenow fragment of DNA polymerase was routinely used as a length marker when separating processing products on the gels.

Northern hybridization: Both radiolabeled and nonradiolabeled processing intermediates and products generated from pPD1157α transcripts were separated on an 8% denaturing polyacrylamide-urea gel and electrophoretically transferred to a Hybond N hybridization membrane using the Bio-Rad (Richmond, CA) mini-transblot device. The membrane was then subdivided into several parts, each containing a radiolabeled pBR322 *Msp*I size marker, unprocessed substrate RNA, and substrate RNA that was processed for 5 min. The nonlabeled substrate parts were then individually probed, using hybridization mixes as described previously (Mackie 1986), with either of the following: oAR7, 5′ CGGGGCGGGAGGGC TTTTCA 3′ (complementary to positions −20 to −1 of the substrate); oAR8, 5′ GGAATGAGACTTCTGAGGTT 3′ (complementary to −63 to −44); oAR9, 5′ ATCCCCCCGCGCG GTTTTTG 3′ (complementary to −122 to −103); or oSP10, 5′ CTCCCATGGCTTATCCCTACCCC 3′ (complementary to +35 to +57 of the 16S gene). Blots were hybridized overnight at 38° in the presence of 50% v/v formamide. The blots were then washed twice for 1 hr at 63° with 2× SSPE and 0.1% SDS. The bands were visualized using a Molecular Dynamics (Sunnyvale, CA) PhosphorImager and aligned using the size marker present on each blot. This allowed creation of the composite Figure 2 in which only the 5-min processing lane from each blot has been shown.

Primer extension: Primer extensions were carried out as described previously (Potter *et al.* 1995) to detect and map 5′ ends generated by partial endonuclease processing of unlabeled substrate RNAs. Primers oAR7 and oSP10 were used to detect cleavages at sites 1 and 2, and primer oAR9 was used to detect cleavages at the * site. In our experiments, the primer extension assay was of limited usefulness for two reasons. First, hybridization of the oAR7, 8, and 9 primers to the substrate RNA was inefficient, presumably because of higher order structure in the substrate RNA. Second, in limited reactions only the 5′ end closest to the primer binding site is detectable.

RESULTS

The standard substrate used for *in vitro* processing contains the entire 144-nt-long 5′ ETS and the first 69 or 72 nt of 16S RNA sandwiched between short 5′ and 3′ flanking vector sequences (Figure 1B). When mixed with cell-free processing activity, the substrate is cleaved efficiently at sites 1 and 2 and less efficiently at site 4 within the 5′ ETS as determined previously by primer extension and S1 nuclease analysis (Potter *et al.* 1995). A more informative and sensitive *in vitro* assay used a uniformly radiolabeled substrate but the pattern of intermediates and products generated during the reaction was complex and difficult to interpret. To identify and characterize these intermediates and products more thoroughly, Northern blot analysis was used. Two reactions, one utilizing radiolabeled substrate and the other utilizing unlabeled substrate, were performed in parallel. After electrophoretic separation and transfer of the reaction products to a nylon membrane, the nonradioactive lanes were probed with four separate oligonucleotides complementary to the regions between the cleavage sites in the substrate RNA. Bands appearing in the Northern hybridization were then identified and correlated to the seven predominant bands generated in the processing reaction with radiolabeled substrate RNA as follows (Figure 2). Band I, hybridizing to all four oligonucleotide probes, is the unprocessed transcript. Band II of ~180 nt in length was identified as the site 1 to 3′ intermediate because of its ability to hybridize to oligonucleotide probes B, C, and D and its failure to hybridize to probe A. Band III of ~150 nt in length was identified as the 5′ to site 2 product because of its ability to hybridize to probes A and B and its failure to hybridize to probes C and D. Bands IV, V, VI, and VII were similarly identified. The detection of both a 5′-site 2 intermediate (band III) and a site 1-3′ intermediate (band II) clearly indicates that there is no concerted or temporal ordering to the site 1 and site 2 *in vitro* cleavage events.

The identity of many of the intermediates and product bands was further substantiated by primer extension analysis using the A, C, and D oligonucleotides as primers and by 5′ end labeling of the substrate RNA. This analysis revealed that a previously uncharacterized cleavage event was occurring in the substrate RNA, 5′ to cleavage site 1. The position of this cleavage within the RNA substrate was mapped by primer extension analysis with primer A (data not shown) to position −159 within the *Hin*dIII site in the 5′ vector sequence and was designated * site. Band IV, 110 nt in length, was clearly identified here as the site 2–3′ product on the basis of Northern hybridization to probes C and D. Previously, this band was misidentified as the site 4-3′ product (Potter *et al.* 1995). The Northern analysis also demonstrated that little or no site 4 cleavage was occurring under the conditions employed because no clear or reproducible hybridization signal of the expected size

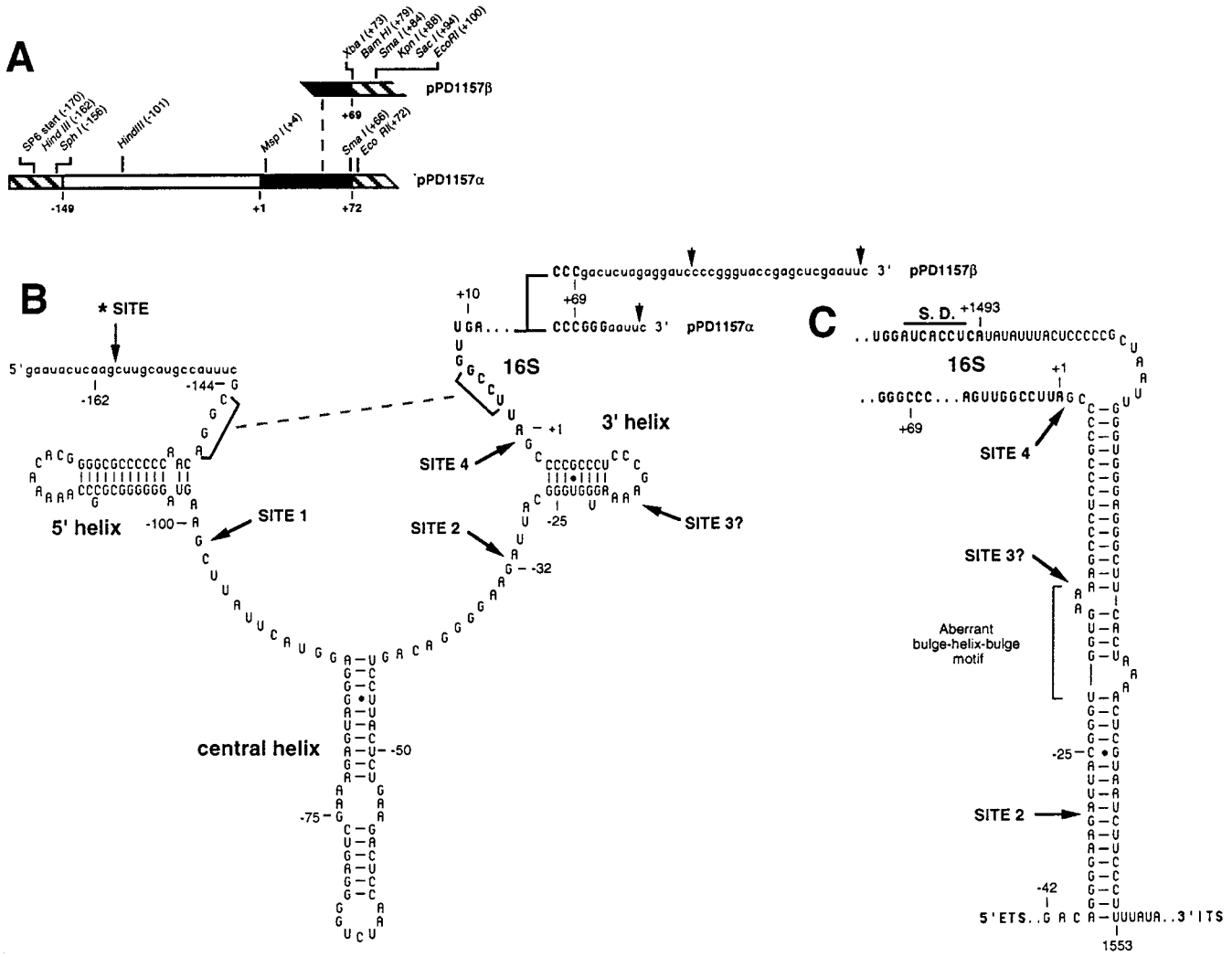

Figure 1.—Plasmid structure and sequences of the 5′ ETS containing substrate RNAs. (A) Plasmid pPD1157α contains 221 nt derived from the single-copy rRNA operon of *S. acidocaldarius* and inserted into the polylinker site of pGEM3Zf(+). The insert consists of 5 nt of 5′ flanking sequence, the entire 144-nt-long 5′ ETS, and the first 72 nt of the 16S rRNA gene. Plasmid pPD1157β is essentially the same except that it contains only the first 69 nt of the 16S gene and has additional restriction sites located distal to the insert. (See materials and methods for details.) (B) The 5′ ETS containing substrates for *in vitro* processing are generated by SP6 transcription of plasmid pPD1157α linearized at the *Eco*RI (GAATTC) site or plasmid pPD1157β or pPD1157γ linearized at the *Bam*HI (GCATCC) or *Eco*RI (GAATTC) sites (↓). Vector polylinker sequences are in lowercase; 5′ ETS and 16S rRNA sequences are in uppercase. Nucleotide numbering is relative to position +1, the 5′ A nucleotide of 16S rRNA. Regions of stable secondary structure predicted by the RNA-fold program are illustrated. The potential sites for endonucleolytic cleavage within the 5′ ETS occur 3′ to G residues located at positions −99, −32, and −1 and are indicated as sites 1, 2, and 4, respectively. An additional cleavage site within the polylinker sequence at the 5′ end of the transcript occurs 3′ to the G residue at position −160 (* site) and removes an 11-nt fragment from the 5′ end of the substrate. (C) *In vivo* pre-rRNA contains an inverted repeat surrounding the 16S rRNA sequence that is predicted to form a stable helical structure containing an aberrant BHB motif. The ascending helical strand derived from the 5′ ETS (nucleotides −42 to −4) sequesters processing site 2; the descending helical strand is derived from the ITS. We recently mapped the 3′ end of the 16S rRNA to position 1493; in an earlier publication, this end was incorrectly mapped to position 1555 at the base of the processing stem (Durovic and Dennis 1994).

with probe D was detected. This maturation cleavage site was therefore not examined further in this study (see discussion).

5′ and 3′ deletions of the substrate RNA: A high-resolution gel illustrating the separation of intermediates and products generated by processing the standard pPD1157α substrate used in Figure 2 is illustrated in Figure 3A (substrate a). Intermediate and product bands are indicated. To examine the importance of sequence and structural features in the 5′ ETS on the recognition and cleavage at sites 1 and 2, 5′ deletions were introduced into the substrate RNA (Figure 3A). The first of these deletions, 5′Δ-102 (substrate b), removed the very stable 5′ hairpin to nucleotide position −102. The second, 5′Δ61 (substrate c), removed the 5′ hairpin, cleavage site 1, and the descending portion of

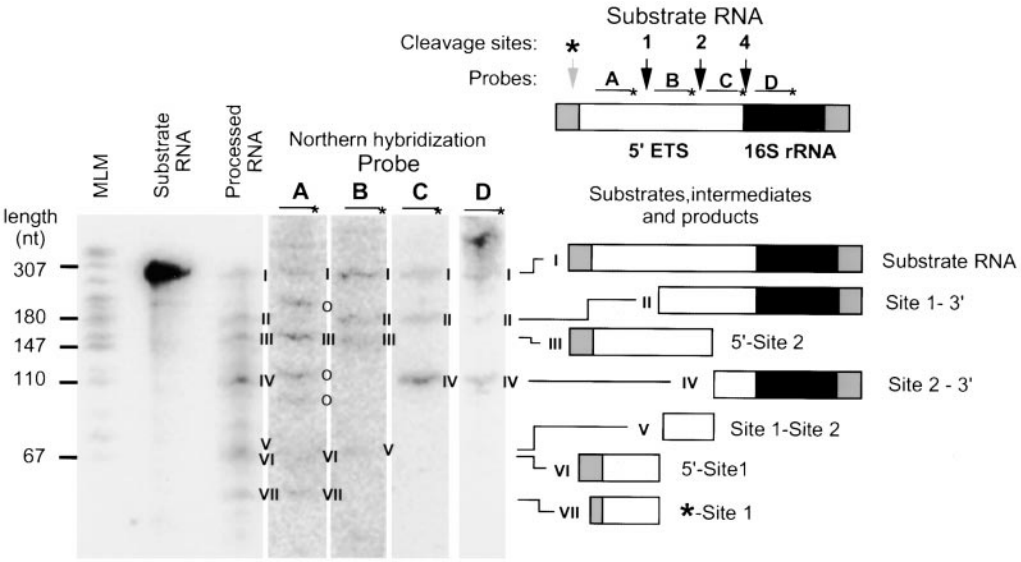

Figure 2.—Characterization of processing intermediates and products by Northern hybridization. Substrate RNA was transcribed from plasmid pPD1157α linearized with *Eco*RI in the presence or absence of [α-^{32}P]-CTP. The first three lanes illustrate a molecular length marker (MLM; only some of the fragment lengths are indicated on the left), unprocessed radiolabeled substrate RNA, and radiolabeled substrate RNA processed for 5 min under standard conditions. Seven bands (I and VII) are indicated. In a parallel reaction, nonradioactive substrate RNA was processed and divided into four parts. All samples were run on an 8% polyacrylamide urea gel and blotted to a nylon membrane. The four nonradioactive lanes were cut out and probed separately with oligonucleotide oAR9(A), oAR8(B), oAR7(C), and oSP10(D). A cartoon of the substrate RNA is depicted at the top right. The stippled, open, and solid regions correspond to vector, 5′ ETS, and 16S sequences, respectively. The positions of cleavage sites *, 1, 2, and 4 and the location of the oligonucleotide hybridization sites are indicated. Various intermediates and products derived from processing of the substrates and their correspondence to radioactive products or Northern hybridization signals are indicated. The precise 5′ end of band III is uncertain; primer extension analysis indicates that it is likely a mixture of authentic and * site 5′ ends. Three bands, designated o, in the A-probe Northern hybridization, have not been identified.

the central hairpin to nucleotide position −61. Surprisingly, 5′ removal of these sequences and the accompanying structural features had no apparent effect on processing at sites 1 and 2 as long as the sites themselves remained intact. Over the course of a 15-min processing reaction using substrate b, all of the expected intermediates and products resulting from cleavage at sites 1 and 2 were observed. For example, the band of ∼70 nt in length and representing the site 1-site 2 product was clearly evident. Because of the 5′ deletion in substrate b, the *-site 2 intermediate observed with wild-type RNA substrate a was replaced by a new 80-nt-long intermediate designated 5′(Δ-102)-site 2. The 180-nt-long site 1-3′ intermediate was clearly evident, whereas the 11-nt-long 5′-site 1 fragment (resulting from cleavage at site 1 in substrate b) was too short to be recovered in the gel system employed. (Note: This 11-nt-long fragment is identical to the 5′-* site fragment generated by cleavage at the * site in the wild-type substrate RNA.) Using substrate c, the site 2-3′ product of ∼110 nt in length was clearly evident, substantiating normal cleavage at site 2 in the absence of site 1. The smaller products originating 5′ to the site 2 cleavage migrate further down the gel and are not shown in Figure 3A.

A series of 3′ deletions in the substrate RNA was also examined (Figure 3B). Deletion 3′Δ+6 (Figure 3B, substrate d) removed all but the first six nucleotides of the 16S sequence. This substrate was processed with normal efficiency at sites 1 and 2 as shown most clearly by the presence of the band of ∼70 nt in length (site 1-site 2 product) and the band of ∼67 nt in length (5′-site 1 product). This indicates that sequences beyond the 5′ ETS-16S junction are not required for 5′ ETS processing. Two additional 3′ deletion substrates were examined. The first, 3′Δ-16 (substrate e), removes all sequences distal to position −16 and thereby prevents formation of the 3′ hairpin. This transcript was processed efficiently at sites 1 and 2 as shown again most clearly by the presence of the 70- and 67-nt-long products. This indicated that the 3′ hairpin (positions −26 to −3) was also not required for recognition and cleavage at sites 1 and 2. The second, 3′Δ-67 (substrate f), removes the 3′ hairpin, processing site 2, and the ascending portion of the central helix (positions −87 to −43). This substrate was cleaved at site 1 as evidenced most clearly by the detection of the 67-nt-long 5′-site 1 product.

Evidence that the lower band of ∼60 nt in length from substrate f contains both the *-site 1 and the site 1-3′ products comes from the following. First, with substrate f the intensity of the band was equal to or greater than the intensity of the 67-nt-long 5′-site 1 band. With other substrates such as a′, d, and e, where this band contains only the single *-site 1 product, its intensity was less than that of the 5′-site 1 band. Second, the 60-nt-long fragment (*-site 1) generated from substrate f comigrates with the identical *-site 1 fragment generated from substrates a′, d, and e (Figure 3B). Third, the unique 60-nt-long product of substrate f (site 1-3′Δ-67 product) was also generated with substrate h (see below); processing of substrate h does not produce the

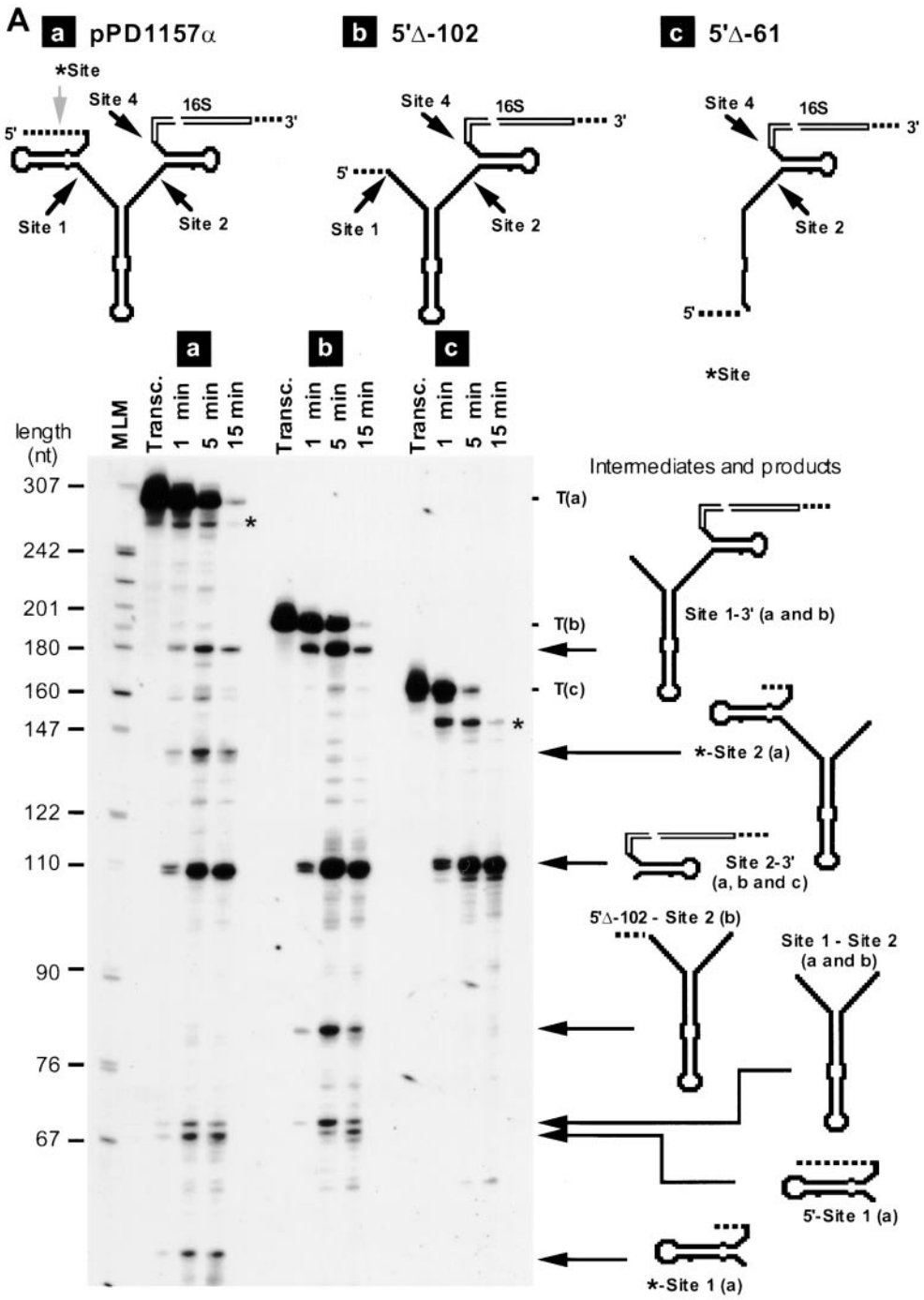

Figure 3.—Processing of substrates containing 5′ deletions in the 5′ ETS. Radiolabeled substrate RNAs were transcribed from plasmids with SP6 RNA polymerase and designated as follows: a, wild type from pPD1157α; b, from pPD1157α5′Δ-102; c, from pPD1157α5′Δ-61; a′, wild type from pPD1157β linearized with *Bam*HI; d, from pPD1157β linearized with *Msp*I at position +6 within the 16S rRNA gene, 3′Δ+6; e, from pPD1157β 3′Δ-16; f, from pPD1157β 3′Δ-67; g, from pPD1157β 5′Δ-102, 3′Δ-16; and h, from pPD1157β 5′Δ-102, 3′Δ-67. Unless otherwise indicated, plasmids were linearized for transcription with *Eco*RI. Transcripts were employed in standard reactions and samples removed for analysis at 1, 5, and 15 min or 1 and 10 min on an 8% polyacrylamide urea gel and visualized by autoradiography. Lanes indicated as MLM are molecular length markers derived by digestion of pBR322 with *Msp*I; not all band lengths are indicated. (A) Substrates containing 5′ deletions. (B) Substrates containing 3′ deletions. (C) Substrates containing both 5′ and 3′ deletions. The structures of the respective transcript substrates are indicated above and their mobility position is indicated by T on the right. The mobility locations and, in most instances, the structures of the various intermediates and products are also illustrated on the right. The position of the band resulting from * site cleavage to shorten the full-length transcripts (a, c, a′, d, e, and f) is indicated (*). The b, h, and g transcripts were generated by deleting sequences between two *Hin*dIII sites (positions −162 and −101; see Figure 1); this removes the * site while preserving the site 1 sequence.

overlapping *-site 1 product. Taken together, these results demonstrate that the central helix (positions −87 to −43) is not required for processing as site 1.

Finally, the processing of two substrates containing both 5′ and 3′ deletions was examined (Figure 3C). The deletions present in substrate g (5′Δ-102, 3′Δ-16) prevent the formation of both the 5′ and 3′ helices. Nonetheless, this substrate was cleaved at both sites 1 and 2 as evidenced by the appearance of a 70-nt-long band representing the site 1-site 2 product. Finally, substrate h is incapable of forming any obvious, stable secondary structure and contains only processing site 1. It was cleaved at site 1 as evidenced by the accumulation of the expected site1-3′ end product of ~60 nt in length.

Although none of the three stable helical structures present in the wild-type 5′ ETS appears to be required for cleavage at sites 1 and 2, the absence of these structures affects the efficiency of cleavage. This was particu-

Figure 3.—*Continued.*

larly evident with substrates f and h and to a lesser extent with substrate g. In these instances, a substantial portion of the input transcript remained uncleaved after 10 min of incubation with processing activity whereas wild-type transcript was fully consumed after 10 min (compare a′ in Figure 3B with f, g, and h in Figure 3, B and C).

Nucleotide substitutions in the substrate RNA: Regions surrounding sites 1 and 2 as well as the fortuitous * site in the 5′ polylinker region of the substrate share a high degree of sequence similarity (AAG↓(C/A)UU) with cleavage occurring 3′ to the G residues at positions −160, −99, and −32 in the substrate RNA. (The 5′ ETS-16S rRNA junction shares a similar sequence, G↓AUU, but is not efficiently cleaved under the reaction conditions employed for site 1 and 2 cleavage; see discussion.) To determine the importance of flanking residues in cleavage site recognition and utilization, site-specific mutagenesis was carried out to create nucleotide substitutions in the substrate RNA at positions −101, −100, −99, −98, −97, and −96 around site 1

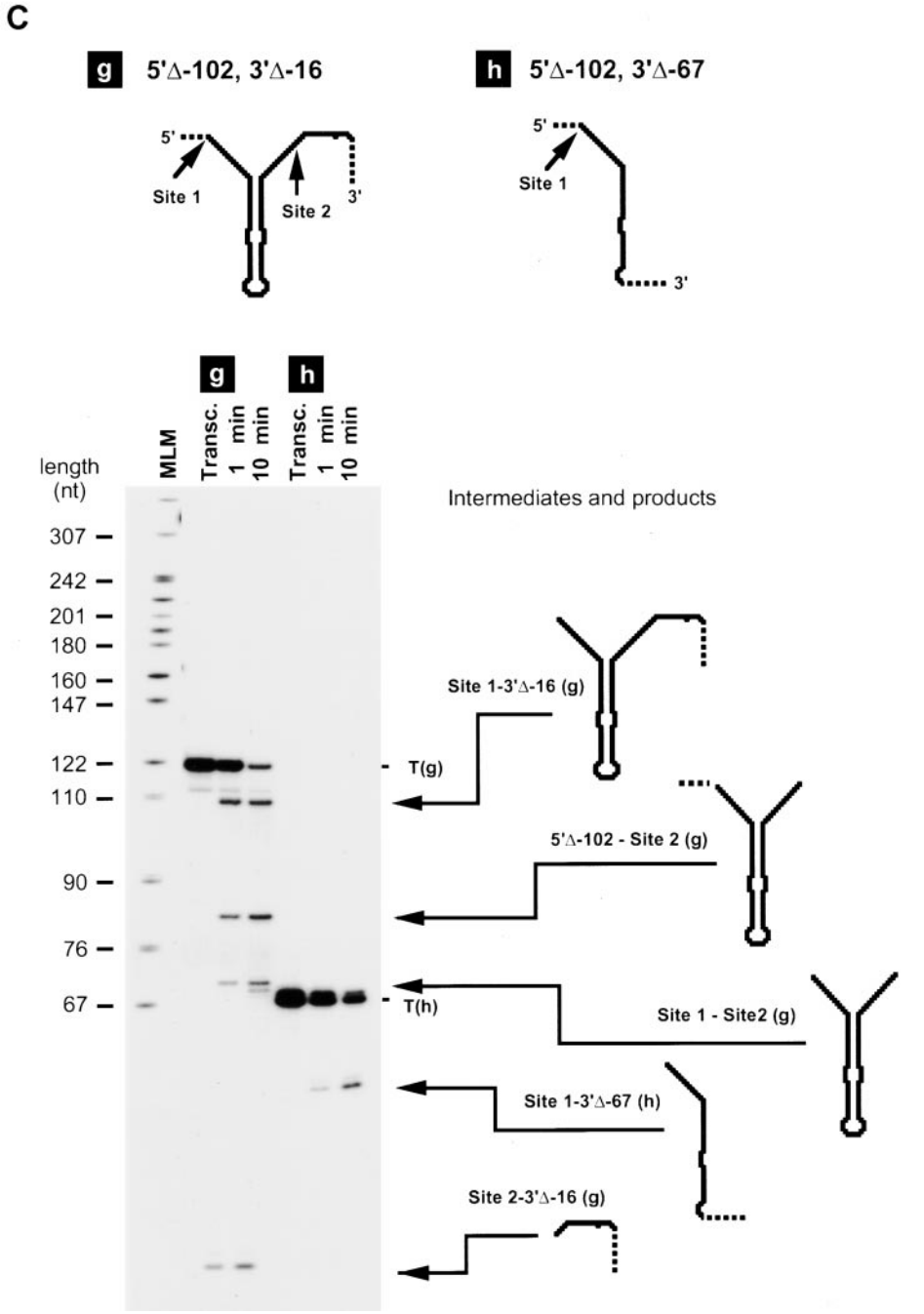

Figure 3.—*Continued.*

and at positions −32 and −31 around site 2. Mutant and control wild-type RNAs were used in standard processing reactions and the accumulation of cleavage intermediates and products was compared.

Of the six positions examined around site 1, only substitutions at positions G-99 and U-97 affect site 1 utilization (Figure 4). Substitutions with C or U at position −99 in place of G completely blocked cleavage as evidenced by the failure to detect any of the anticipated products resulting from site 1 cleavage, whereas substitution with A at this position exhibited the normal pattern of cleavage. These results indicate that the processing activity at site 1 requires a G or A nucleotide 5′ to the cleavage site. Substitution with A, C, or G at position 97 in place of U also blocks cleavage at site 1. The properties of all the site 1 substitutions that have been examined are summarized in Table 1.

To further characterize the sequence requirements for cleavage at site 1, we constructed a duplication of four nucleotides by Klenow filling of the *Hin*dIII restriction site (AAGCUU→AAGCUAGCUU). This construct now contains a tandem duplication of the core of the site 1 recognition element. We predict that the modified substrate should be cut at two positions within the duplication to produce site 1 intermediates or products that are both identical in size and four nucleotides longer

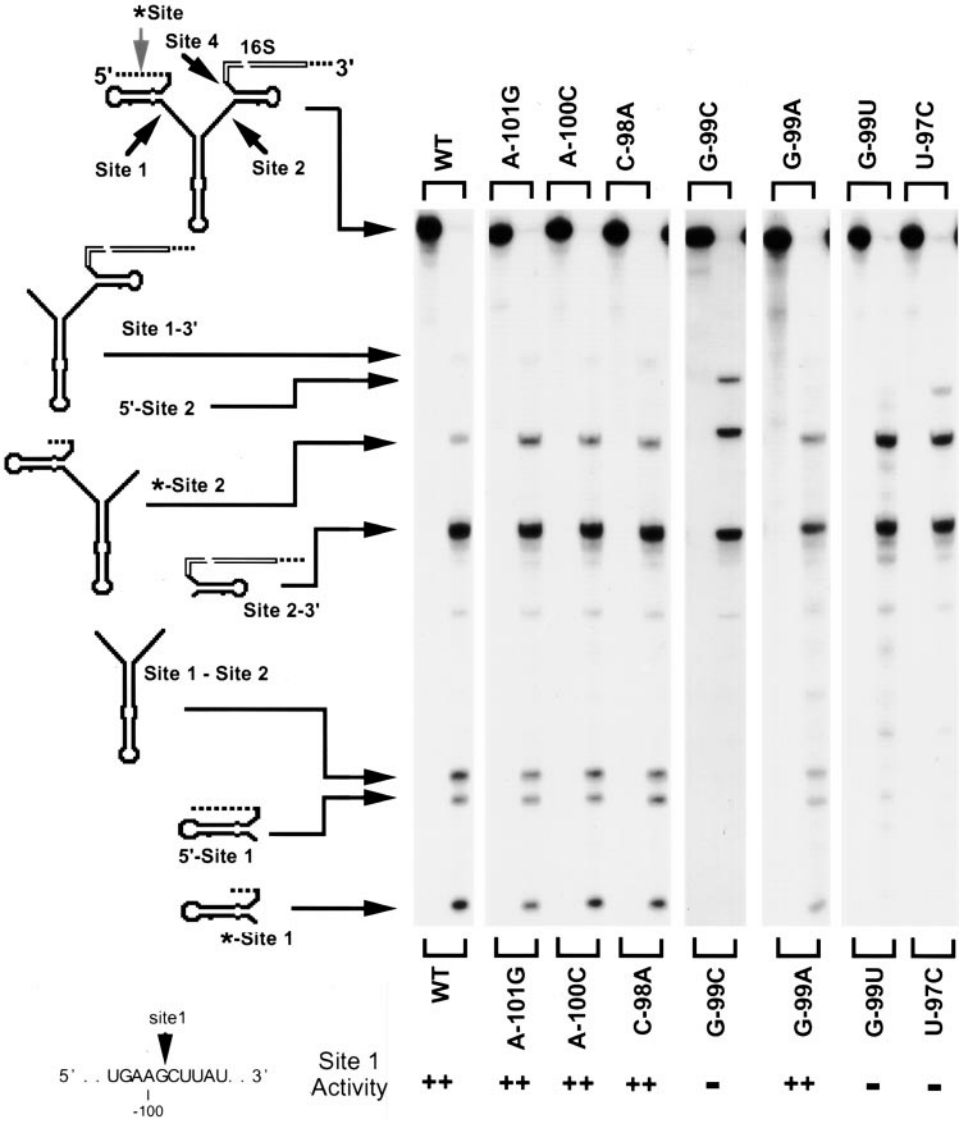

Figure 4.—Analysis of 5′ ETS substrates containing site-specific nucleotide substitutions around site 1. Site-directed mutagenesis was used to obtain nucleotide substitutions at positions −101, −100, −99, −98, and −97 around site 1 within the 5′ ETS region on plasmid pPD1157γ. The altered templates were transcribed with SP6 RNA polymerase following linearization with *Bam*HI and utilized in a standard processing reaction. For each transcript, the 0- and 5-min samples (left and right lanes in each pair, respectively) were loaded onto an 8% polyacrylamide urea gel and visualized by autoradiography. The wild-type transcript is designated WT. Mutant transcripts are designated by the wild-type base, the position of substitutions, and the replacement base (*i.e.*, A −101 G). Intermediates and products produced during the processing reactions are identified and illustrated on the left. The cleavage activity at site 1 was deduced from the relative intensities of intermediates and products for the wild-type and mutant substrates and is summarized at the bottom: ++, normal cleavage; +, reduced cleavage; −, little or no cleavage. The sequence around the wild-type site 1 is shown at the bottom left.

than those generated with the wild-type substrate. This prediction was confirmed (Figure 5). In the processing reaction using this mutant, site 1 products appeared as doublets, with one band migrating at or near the same position as the wild-type product and the other band migrating about four nucleotides longer than the wild-type product. For the site 1 to 3′ product, this interpretation was independently confirmed by primer extension analysis (data not shown). It is also evident that the rate of cleavage at site 1 in the duplication substrate was reduced by severalfold compared to the rate of cleavage of site 1 in the wild-type substrate. This suggests that efficient recognition of the cleavage site must require features beyond the AGCU core tetranucleotide.

A number of other substitutions were made at positions −32 and −31 surrounding cleavage site 2 (Figure 6). Deletion or substitution with U or C at position −32 in place of G (either alone or in combination with changes at position −31) blocked cleavage as evidenced by the failure to detect appreciable amounts of the anticipated intermediate and products. Substitution with A at position −32 reduced the rate of utilization about three- to fourfold as evidenced by the lower level of

TABLE 1

Endonuclease activity on substrates containing single nucleotide substitutions around site 1

	Nucleotide substitutions			
Nucleotide Position	G	A	U	C
A-101	++	WT	NA	NA
A-100	++	WT	++	++
G-99	WT	++	−	−
C-98	++	++	++	WT
U-97	−	−	WT	−
U-96	NA	NA	WT	++

Cleavage efficiency at site 1 is indicated as ++, wild-type; −, aberrant cleavage, little or no cleavage; NA, not analyzed; WT, the wild-type nucleotide at this position.

Figure 5.—Processing of a 5′ ETS transcript containing a four-nucleotide duplication at processing site 1. Plasmid pPD1157β(ins4) contains a four-nucleotide duplication of the AGCU sequence overlapping processing site 1. Radiolabeled substrates were transcribed for *in vitro* processing from the parent plasmids pPD1157β and the insertion plasmid pPD1157β(ins4) linearized with *Bam*HI. The (A) wild-type and (B) mutant transcripts were processed in a standard reaction and samples were removed at 1 and 5 min for analysis. The intermediate resulting from * site cleavage of the primary transcript is identified (*) on the autoradiogram. Other intermediates and products accumulating in the two reactions are also identified. Four doublet bands that differ in length by four nucleotides because of the duplication within site 1 are indicated on the right by double-headed arrows; the structures of these and other intermediates and products are indicated on the left.

accumulation of the site 2 intermediate and products. This contrasts with the G-to-A substitution at position −99, which had no apparent effect on site 1 recognition and cleavage by the endonuclease (see Figure 4). The detrimental effect of the A-to-G substitution at position −32 was suppressed by a second A-to-U substitution at position −31. These results confirm the importance of a purine 5′ to the scissile phosphodiester bond and further highlight the complexity of site recognition and utilization by the endonuclease.

The fortuitous * site cleavage in the RNA substrate occurs within the sequence derived from the polylinker *Hin*dIII site of the vector used to transcribe substrate RNAs. The sequence, AAGCUU, is exactly the same as that surrounding the normal cleavage site 1. The ability to cleave at the * site was further investigated in RNAs that lacked *S. acidocaldarius* 5′ ETS and 16S rRNA sequences. The site was recognized and cleaved between the G and C residues in an SP6 transcript of the polylinker region from pGEM7Zf(+) but not recognized or cleaved in the complementary T7 transcript from the same plasmid (data not shown). These results confirm the presence of additional features outside of the hexanucleotide sequence that contribute to recognition and cleavage by the endonuclease.

DISCUSSION

In most bacteria and archaea, the catalytic activities responsible for the initial cleavages in pre-rRNA transcripts are the bacterial RNaseIII and the archaeal bulge-helix-bulge endonucleases. Both activities cleave within long duplex structures that surround 16S and 23S rRNA sequences and release pre-16S and pre-23S rRNAs from the primary transcript (Dennis 1991; Garrett *et al.* 1991; Court 1993; Nicholson 1996). As in most other archaea, the *S. acidocaldarius* 16S and 23S sequences within the pre-rRNA transcript are surrounded by heli-

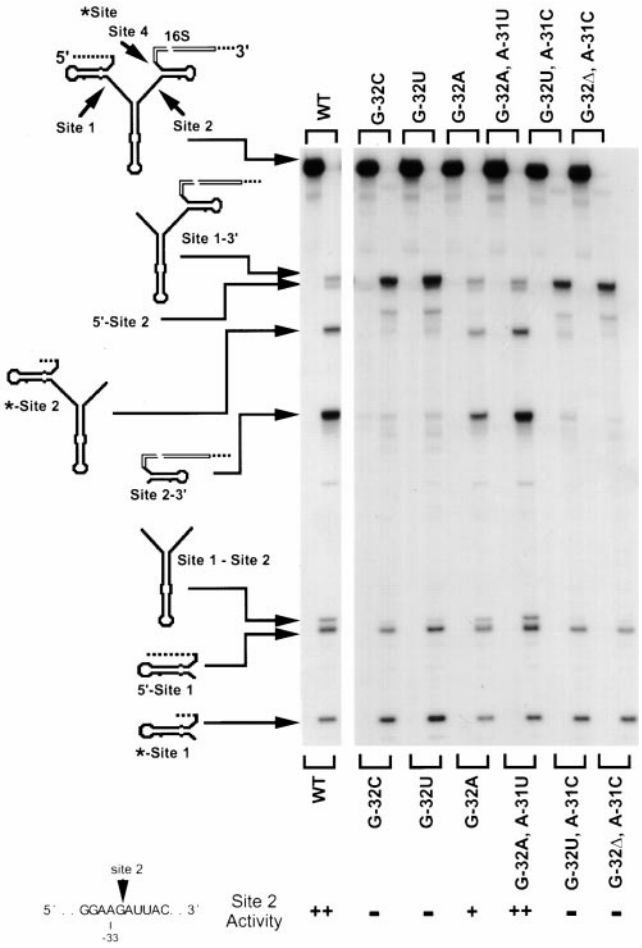

Figure 6.—Analysis of 5′ ETS substrates containing site-specific nucleotide substitutions around site 2. Site-directed mutagenesis was used to obtain single and double substitutions at positions −32 and −31. Other details are as described in the legend to Figure 4.

cal structures. The 23S helix contains the canonical BHB motif and excision of pre-23S rRNA has been shown to occur by cleavage within this motif (Durovic and Dennis 1994). However, the 16S helix contains an aberrant motif that lacks a critical nucleotide from the loop on the ascending strand (Figure 1C) and the mechanism for pre-16S excision remains unclear. If the *S. acidocaldarius* BHB endonuclease exhibits the same substrate specificity as the well-characterized halophile enzyme (Thompson *et al.* 1989; Kleman-Leyer *et al.* 1997), the aberrant 16S motif should not be a substrate for cleavage (but see below).

In our studies, we have detected two novel cleavages that occur in the 5′ ETS upstream of the BHB motif. These novel cleavages have been reproduced *in vitro*. However, the role of these cleavages and their relationship to formation of the 16S helix and the excision and maturation of 16S rRNA remain to be established. The pathway for processing and maturation of 16S rRNA is necessarily constrained by the directional transcription of the rRNA operon and presumably requires rearrangements in RNA secondary structure. As the RNA polymerase transcribes through the 5′ ETS and into the 16S gene, the 5′ ETS is predicted to form three regions of localized secondary structure as depicted in Figure 1B. At this point, the novel cleavage sites 1 and 2 are exposed and accessible to endonuclease cleavage. As the polymerase exits the 16S gene and transcribes through the 16S-23S intergenic space, the 5′ ETS, whether still intact or already cleaved at sites 1 and/or 2, is predicted to rearrange to form the very stable 16S processing stem. If cleavage has occurred at site 2, the stem will be shortened by only 8 bp (Figure 1C). The presence of the 16S-23S intergenic spacer sequence in the *in vitro* reaction strongly favors formation of the 16S processing helix and effectively blocks endonucleolytic cleavage at site 2 (A. G. Russell and P. P. Dennis, unpublished results). Moreover, the endonuclease activity that we have purified and characterized is unable to cleave at site 3 in either the presence or absence of the complementary 16S-23S intergenic sequence.

The archaeal BHB enzyme is homologous to the two excision subunits of the tetrameric eukaryotic tRNA intron endonuclease (Kleman-Leyer *et al.* 1997; Trotta *et al.* 1997). The eukaryotic enzyme, unlike its halophile counterpart, identifies cleavage sites by a ruler mechanism and is relatively insensitive to cleavage site geometry. If the *S. acidocaldarius* enzyme shares this eucaryotic feature, cleavage could occur at the aberrant site within the 16S processing stem of pre-rRNA and release the pre-16S from the primary transcript. There is some indication that endonucleolytic cleavage may occur *in vivo* at site 3 (*i.e.*, the BHB motif) within the 5′ ETS but so far no indication of cleavage on the opposite strand within the intergenic spacer. This may be a technical problem of detection by S1 nuclease analysis or may indicate that the intermediates are short-lived and rapidly trimmed to the mature 3′ end of 16S rRNA. The retention of the 16S processing stem in the *S. acidocaldarius* rRNA operon implies that it continues to play an important or essential role in 16S rRNA processing or 30S subunit assembly.

Cleavage site consensus sequence: The novel endonuclease that cleaves in the 5′ ETS region of pre-rRNA has been characterized *in vitro* using a radiolabeled substrate RNA. The substrate was cleaved efficiently at site 1 (between positions −99 and −98) and site 2 (between positions −32 and −31) within the 5′ ETS and at a third fortuitous * site (between positions −160 and 159) within the 5′ polylinker sequence. Comparison of the sequences surrounding these sites revealed a hexanucleotide consensus of AAG(A/G)UU. The G residue 5′ and the U residue two nucleotides 3′ to the scissile phosphodiester bond have been shown to be important in recognition and cleavage by the endonuclease. Several experiments indicate that more than the consensus sequence is required for efficient recognition and cleav-

age. First, removal of secondary structural elements from the 5′ ETS, while not blocking the cleavage reactions, significantly reduces the rate of the reactions. Second, an SP6 RNA transcript derived from the polylinker region of plasmid pGEM7Z$^+$ was cleaved at the single AAGCUU sequence, whereas the complementary T7 transcript also containing the AAGCUU sequence was not cleaved. Third, an antisense *E. coli* S20 transcript used as a control in several experiments contains two copies of the AAGCUU hexanucleotide; neither appeared to be efficiently recognized and cleaved by the endonuclease activity (data not shown). Fourth, examination of the substrate RNA revealed a number of other sequences that show a high degree of similarity to the hexanucleotide consensus and conserve a G at position 3. The best match, AAGCUG, occurs at position −77 and a second match, AAGUCU, occurs at position −55. Neither of these sites appeared to be cleaved. Both of these match sites may be partially obstructed by RNA secondary structure in the wild-type substrate but are respectively exposed in the two deletion substrates 5′Δ-61 and 3′Δ-67. Neither of the deletion substrates appears to be cleaved at these positions. Moreover, the second match at position −55 has a C residue at position +2 relative to the expected site of cleavage; at site 1, a U-to-C substitution at the corresponding position abolishes cleavage by the endonuclease. Finally, a 4-base duplication at site 1 to create two hexamers (AAGCUA and UAGCUU) that overlap by 2 nt was generated. Both of these targets were cleaved as evidenced by the doublet patterns of site 1 cleavage products (see Figure 5), but the rate of cleavage is reduced compared to the wild-type substrate. Collectively, these experiments showed that, although the hexanucleotide contains important determinants for recognition and cleavage (including the G and U residues 5′ and 2 nt 3′ to the cleavage site), other determinants beyond this sequence appear to exist.

The cleavage at sites 1 and 2 within the 5′ ETS of *S. acidocaldarius* pre-rRNA resemble in some respects the cleavages mediated by *E. coli* RNase E within mRNAs and the 5S region of pre-rRNA (Belasco 1993). The RNase E target is single-stranded and often adjacent to a region of secondary structure and is defined by a weak consensus with a G residue 5′ to the cleavage site. The adjacent duplex is believed to stabilize the structure of the RNA such that the RNase E target site is maintained in a recognizable single-strand conformation. The regions of secondary structure in the 5′ ETS of *S. acidocaldarius* pre-mRNA may play a similar role in maintaining cleavage sites 1 and 2 in single-stranded conformations. An RNase E-like activity has been identified in *Haloarcula marismortui* and may be used to excise a pre-5S RNA from the primary rRNA transcript (Franzetti *et al.* 1997; Dennis *et al.* 1998). In *S. acidocaldarius*, the 5S gene is not part of the 16S-23S rRNA operon and transcription apparently results in direct production of mature 5S rRNA (Durovic *et al.* 1994).

The sequence at the 5′ ETS-16S junction (site 4) shares sequence similarity to sites 1 and 2 (*i.e.*, GAUU *vs.* AAG(A/C)UU) and suggests that the endonuclease that we are characterizing may be responsible for maturation at the 5′ end of 16S rRNA. *In vitro*, we have detected a low level of cleavage at this position with our purified activity (A G. Russell and P. P. Dennis, unpublished results). The significance of this is unclear and efforts to enhance the endonuclease cleavage at site 4 have not yet been successful. We suspect that the 5′ end maturation of 16S rRNA is complex and that efficient maturation may require (i) more extensive 16S sequence, (ii) the presence of ribosomal proteins, (iii) a rearrangement in the structure of the substrate RNA in order to expose the cleavage site, and (iv) involvement of the 16S processing helix. On the basis of substrate specificity, the novel endonuclease characterized here remains a candidate for the activity that generates the mature 5′ end of 16S rRNA.

We thank Peter Durovic for constructing some of the plasmids used in this study and George Mackie for his advice and encouragement. This work was supported by a grant from the Medical Research Council of Canada (MT6340) to P.P.D. and a University of British Columbia Graduate Fellowship to A.G.R.

LITERATURE CITED

Amiri, K. A., 1994 Fibrillarin-like proteins occur in the domain archaea. J. Bacteriol. **176:** 2124–2127.

Balakin, A. G., L. Smith and M. J. Fournier, 1996 The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related function. Cell **86:** 823–834.

Belasco, J., 1993 mRNA degradation in prokaryotic cells: an overview, pp. 3–12 in *Control of Messenger RNA Stability*, edited by J. Belasco and G. Brawerman. Academic Press, San Diego.

Brock, T. O., K. M. Brock, R. T. Belly and R. L. Weiss, 1972 Sulfolobus: a new genus of sulfur oxidizing bacteria living at low pH and high temperature. Arch. Microbiol. **84:** 54–68.

Chant, J., and P. P. Dennis, 1986 Archaebacteria: transcription and processing of ribosomal RNA sequences in the *Halobacterium cutirubrum*. EMBO J. **5:** 1091–1097.

Court, D., 1993 RNaseIII: a double stranded RNA processing enzyme, pp. 70–116 in *Control of mRNA Stability*, edited by G. Brawerman and J. Belasco. Academic Press, New York.

Dennis, P. P., 1991 The ribosomal RNA operons of halophilic archaebacteria, pp. 251–257 in *General and Applied Aspects of Halophilic Microorganisms*, edited by F. Rodriguez-Valera. Plenum Press, New York.

Dennis, P. P., S. Ziesche and S. Mylvaganam, 1998 Transcription analysis of two disparate rRNA operons in the halophilic archaeon *Haloarcula marismortui*. J. Bacteriol. **180:** 4804–4813.

Diener, J. L., and P. B. Moore, 1998 Solution structure of a substrate for the archaeal pre-tRNA splicing endonucleases: the bulge-helix-bulge motif. Mol. Cell **2:** 883–894.

Dunn, J. J., and F. W. Studier, 1973 T7 early RNAs and *Escherichia coli* ribosomal RNAs are cut from large precursor RNAs *in vitro* by ribonuclease III. Proc. Natl. Acad. Sci. USA **70:** 3296–3300.

Durovic, P., and P. P. Dennis, 1994 Separate pathways for excision and processing of 16S and 23S rRNA from the primary rRNA operon transcript from the hyperthermophilic archaebacterium *Sulfolobus acidocaldarius*: similarities to eucaryotic rRNA processing. Mol. Microbiol. **13:** 229–242.

Durovic, P., U. Kutay, C. Schleper and P. P. Dennis, 1994 Strain identification and 5S rRNA gene characterization of the hyperthermophilic archaebacterium *Sulfolobus acidocaldarius*. J. Bacteriol. **176:** 514–517.

Franzetti, B., G. Sohlberg, G. Zaccai and A. von Gabain, 1997

Biochemical and serological evidence for an RNase E-like activity in halophilic archaea. J. Bacteriol. **179:** 1180–1185.

Garrett, R. A., J. Dalgaard, N. Larsen, J. Kjems and A. S. Mankin, 1991 Archaeal rRNA operons. Trends Biochem. Sci. **16:** 22–26.

Gegenheimer, P., and D. Apirion, 1981 Microbiol. Rev. **45:** 502–541.

Gerbi, S. A., 1995 Small nucleolar RNA. Biochem. Cell Biol. **73:** 845–858.

Hui, I., and P. Dennis, 1985 Characterization of the ribosomal RNA gene cluster in *Halobacterium cutirubrum*. J. Biol. Chem. **260:** 899–906.

Kiss-László, Z., Y. Henry, J.-P. Bachellerie, M. Caizergues-Ferrer and T. Kiss, 1996 Site specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. Cell **85:** 1077–1088.

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

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

La Grandeur, T. E., S. C. Darr, E. S. Haas and N. R. Pace, 1993 Characterization of the RNase P RNA of *sulfolobus acidocaldarius*. J. Bacteriol. **175:** 5043–5048.

Li, H., C. R. Trotta and J. Abelson, 1998 Crystal structure and evolution of a transfer RNA splicing enzyme. Science **280:** 279–284.

Mackie, G. A., 1986 Structure of the DNA distal to the gene for ribosomal protein S20 in *Escherichia coli* K12: presence of a strong terminator and an IS1 element. Nucleic Acids Res. **14:** 6965–6981.

Maxwell, E. S., and M. J. Fournier, 1995 The small nucleolar RNAs. Annu. Rev. Biochem. **35:** 897–934.

Ni, J., A. L. Tien and M. J. Fournier, 1997 Small nucleolar RNAs direct site-specific synthesis of pseudouridine in ribosomal RNA. Cell **89:** 565–573.

Nicholson, A. W., 1996 Structure, reactivity and biology of double stranded RNA. Prog. Nucleic Acid Res. Mol. Biol. **52:** 1–65.

Olsen, G., N. Pace, M. Nuell, B. Klaine, R. Gupta *et al.*, 1985 Nucleotide sequence of the 16S rRNA gene from the thermophilic archaebacterium *Sulfolobus solfataricus*. J. Mol. Evol. **22:** 301–307.

Potter, S., P. Durovic and P. P. Dennis, 1995 Ribosomal RNA precursor processing by a eucaryotic U3 small nucleolar RNA-like molecule in an archaeon. Science **268:** 1056–1060.

Thompson, L. D., and C. J. Daniels, 1988 A tRNA trp intron endonuclease from *H. volcanii*: unique substrate recognition properties. J. Biol. Chem. **263:** 17951–17959.

Thompson, L. D., L. D. Brandon, D. T. Nieuwlandt and C. J. Daniels, 1989 Transfer RNA intron processing in the halophilic archaebacteria. Can. J. Microbiol. **35:** 36–42.

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

Venema, J., and D. Tollervey, 1995 Processing of pre ribosomal RNA in *Saccharomyces cerevisiae*. Yeast **11:** 1629–1650.

Young, R. A., and J. A. Steitz, 1978 Complementary sequences 1700 nucleotides apart form a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. Proc. Natl. Acad. Sci. USA **75:** 3593–3597.

Zhou, Z., and M. P. Deutscher, 1997 An essential function for the phosphate dependent exoribonuclease RNase PH and polynucleotide phosphorylase. J. Bacteriol. **179:** 4391–4395.

Communicating editor: P. Blum