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 archaean *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 Steltz 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 RNA 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 RNA, and the second is the bulge-helix-bulge (BHB) endonuclease, the enzyme that excises introns from the transcripts of intron-containing RNA and tRNA genes (Thompson and Daniels 1988; Kjems et al. 1989; Thompson et al. 1989; La Grandeur et al. 1993). The BHB enzyme from *Haloflex 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; Klemens-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 (Dierer 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 RNA 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 RNA from the rRNA operon primary transcript (Chant and Dennis 1986; Dennis et al. 1998).

In eukaryotic organisms pre-rRNA synthesis and pro-
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 (VeNena and To11er vey 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 degenerate mutagenic oligonucleotides at positions (8-100), (8-100), (8-99), (8-99), (8-97), (8-96), (8-96), and (8-32) and (8-32) and (8-31). The four-nucleotide internal duplication around site 1 was generated by cloning an Sphal-Sacl fragment from pPD1157 between the same sites in pGEM3Zf(+) (see Figure 1). All 5’ deletion mutants that remove 5’ ETS sequences to position –101 were generated by excising the HindIII-HindIII fragment (–162 to –102) either from pPD1157 or from pPD1157 3’ deletion derivatives. The three deletion substrates ending at position +6 were generated by SP6 transcription of plasmids linearized with MspI, which cleaves at position 6 within the 16S rRNA sequence.

Site-specific nucleotide substitutions were generated from construct pPD1157. This was created by cloning an Sphal-EcoRI fragment from pPD1157 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 (8-101), (8-100), (8-99), (8-98), (8-97), (8-96), (8-32), and (8-31). The four-nucleotide internal duplication around site 1 was generated by cloning an Sphal-Sacl fragment from pPD1157 between the same sites in pGEM3Zf(+), which lacks a multiple cloning HindIII restriction site. This construct was then linearized with HindIII, filled in with Klenow Enzyme, and religated. An Sphal-Sacl fragment from this construct was recloned between the same sites in pGEM3Zf(+) to create pPD1157h(inc4).

In vitro transcripts were synthesized from DNA templates linearized with EcoRI for pPD1157 and BamH for pPD1157 and for all site-specific nucleotide substitutions generated in pPD1157. To generate 3’ deletion substrates ending at position +6 within the 16S rRNA, the template plasmids were linearized with MspI. All template DNAs containing other 3’ deletions were linearized with EcoRI. Run-off transcripts were then made using SP6 RNA polymerase (according to the protocol in Promega’s Protocols and Applications Guide, Ed. 2) with [32P]CTP for uniformly radiolabeled substrates or CTP for nonradio-labeled 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 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 BamHI. The plasmid contains a 178-nt-long insert into the polylinker HindIII site of pSP72; 168 nt of the insert are derived from the E. coli 520 ribosomal protein gene. The transcript, containing two internal HindIII sites, is antisense with respect to the 520 coding region.

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

**MATERIALS AND METHODS**

**Substrate RNAs:** The plasmid pPD1157 contains an insert extending from position –149 to +72 in the 16S gene of the single-copy S. acidocaldarius rRNA operon (Figure 1) in the vector pGEM3Z(+) . 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 HindIII and SphI, the remnant of a PsI restriction site, and an insert segment ending with an Smal site that is followed by an EcoRI site (see Figure 1). The SphI-Smal fragment from pPD1157 was recloned between the SphI and HindIII sites of pGEM3Z(+) to create pPD1157. The 5’ deletion substrate 5’-Δ61 was generated by more extensive unidirectional exonuclease III digestion of the progenitor to pPD1157 (Durovic and Dennis 1994). It differs from pPD1157 only in lacking nucleotides between positions –149 and –62. The 3’ deletions of plasmid pPD1157 (3’-Δ61 and 3’-Δ67) were created by unidirectional exonuclease III digestion from the BamHI site and use of the Kpnl 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 sequences to position –101 were generated by excising the HindIII HindIII fragment (–162 to –102) either from pPD1157 or from pPD1157 3’ deletion derivatives. The three deletion substrates ending at position +6 were generated by SP6 transcription of plasmids linearized with MspI, which cleaves at position 6 within the 16S rRNA sequence.

Site-specific nucleotide substitutions were generated from construct pPD1157. This was created by cloning an Sphal-EcoRI fragment from pPD1157 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 Sphal-Sacl fragment from pPD1157 between the same sites in pGEM3Zf(+), which lacks a multiple cloning HindIII restriction site. This construct was then linearized with HindIII, filled in with Klenow Enzyme, and religated. An Sphal-Sacl fragment from this construct was recloned between the same sites in pGEM3Zf(+) to create pPD1157h(inc4).

In vitro transcripts were synthesized from DNA templates linearized with EcoRI for pPD1157 and BamHI for pPD1157 and for all site-specific nucleotide substitutions generated in pPD1157. To generate 3’ deletion substrates ending at position +6 within the 16S rRNA, the template plasmids were linearized with MspI. All template DNAs containing other 3’ deletions were linearized with EcoRI. Run-off transcripts were then made using SP6 RNA polymerase (according to the protocol in Promega’s Protocols and Applications Guide, Ed. 2) with [32P]CTP for uniformly radiolabeled substrates or CTP for nonradio-labeled 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 BamHI. The plasmid contains a 178-nt-long insert into the polylinker HindIII site of pSP72; 168 nt of the insert are derived from the E. coli 520 ribosomal protein gene. The transcript, containing two internal HindIII sites, is antisense with respect to the 520 coding region.

In vitro processing of normal and mutant pre-rRNA substrates: Midlog phase cultures of S. acidocaldarius grown at 78–80°C, pH 3.7 (Brock et al. 1972), were rapidly cooled to 0°C and harvested by centrifugation. Cell pellets were resus-
puriﬁed preparation of the activity was used, prepared by extension and S1 nuclease analysis of pBR322 with the Klenow fragment of DNA polymerase. Bands generated in the processing reaction with radiolabeled substrate were used to detect cleavages at the * site. In our experiments, the primer 5’-CTCCCATGGCTTATCCCTACCCC 3’ was used to detect cleavages at sites 1 and 2, and primer oAR9 was used for Northern hybridization to probes C and D. Previously, the primer A (data not shown) was used to position 38 of the 16S gene). Blots were hybridized overnight with primer A (data not shown) to position 20 to 63 of the 16S gene (complementary to the regions be exempted from each blot has been shown. With primer A (data not shown) to position 38 of the 16S gene). Blots were hybridized overnight with primer A (data not shown) to position 20 to 63 of the 16S gene (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 HindIII 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

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 HindIII 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 EcoRI (GAATTC) site or plasmid pPD1157β or pPD1157γ linearized at the BamHI (GCATCC) or EcoRI (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
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′(-Δ102)-site 2. The 180-nt-long site 1-site 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 (Figures 3B and 3C). Deletion 3′ΔS6 (Figures 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′Δ6 (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′ΔE7 (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 d 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′ΔE7 product) was also generated with substrate h (see below); processing of substrate h does not produce the...
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-
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 (AAGj(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, GjAUU, 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, sitespecific mutagenesis was carried out to create nucleotide substitutions in the substrate RNA at positions -101, -100, -99, -98, -97, and -96 around site 1.
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 HindIII 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 pPD1157. The altered templates were transcribed with SP6 RNA polymerase following linearization with BamHI 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 (e.g., 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>
<thead>
<tr>
<th>Nucleotide Position</th>
<th>Nucleotide substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>G</td>
</tr>
<tr>
<td>A-101</td>
<td>++</td>
</tr>
<tr>
<td>A-100</td>
<td>++</td>
</tr>
<tr>
<td>G-99</td>
<td>WT</td>
</tr>
<tr>
<td>C-98</td>
<td>++</td>
</tr>
<tr>
<td>U-97</td>
<td>–</td>
</tr>
<tr>
<td>U-96</td>
<td>NA</td>
</tr>
</tbody>
</table>

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.
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 HindIII 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-
A Novel Archaeal Endonuclease

of the rRNA operon and presumably requires re-arrangements 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 3OS 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. Either 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 resembles 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 RNA.

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 (MT-6340) to P.D.P. and a University of British Columbia Graduate Fellowship to A.G.R.

LITERATURE CITED


Communicating editor: P. Blum