CHARACTERISATION OF A NOVEL PATHWAY FOR RIBOSOMAL RNA MATURATION IN *SULFOLOBUS ACIDOCALDARIUS*

by

PETER VINCENT DUROVIC

B.Sc., The University of British Columbia, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

GENETICS GRADUATE PROGRAM

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

March 1993

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of Biochemistry (Genetics Program)

The University of British Columbia
Vancouver, Canada

Date April 30, 1993

ABSTRACT

Since the initial proposition that the archaebacteria form a primary kingdom as distinct as that of the eubacteria or the eukaryotes, sequence data generated from the ribosomal RNA genes have flooded the databases and periodicals. Phylogenetic trees based on these sequences have been constructed to map the finest details of topology and branching order within the archaebacteria. Yet, despite the plethora of sequence data, relatively little was discovered regarding rRNA gene regulation, transcript processing and requirements for mature ribosome function. The aim of this study is to analyze possible novel regulatory mechanisms in the rRNA genes of the extremely thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*.

The three ribosomal RNA genes were cloned and sequenced. The gene organization was confirmed to differ from that of the halophilic archaebacteria and the eubacteria: the 5S gene was not linked to the 16S and 23S operon, and the operon lacked recognizable tRNA sequences. Southern hybridization unveiled, and sequence data confirmed a long-standing confusion regarding species identity. The previously published *Sulfolobus acidocaldarius* 5S sequence was shown to have been attributed to the wrong species.

Mapping experiments showed that both transcripts initiated downstream of a previously defined archaebacterial promoter sequence. While sequence data showed the 5S transcript start site and end site to be coincidental with the mature 5S termini, the 16S-23S transcript was shown to contain a 143 nucleotide transcribed leader sequence, a 138 nucleotide intergenic sequence, and a trailer sequence of at least 105 nucleotides. Inverted repeat sequences within these transcribed non-coding regions allow for the formation of numerous stem-loops conforming to a semi-conserved archaebacterial structure. While no processing took place within the 5S transcript, extensive processing of the 16S-23S transcript was observed. Of the 12 processing sites mapped, only 6 could be accounted for in the context of precursor processing and maturation events known directly or inferred by analogy from the halophilic archaebacteria and the eubacteria. Alignment of the remaining sites revealed a non-trivial sequence and structural similarity.

If the novel processing indeed took place in the postulated context, it would mark a radical departure from the expected maturation mechanism thought to predate the speciation of archaebacteria and eubacteria. To examine this possibility, *in vitro* transcripts from judiciously selected DNA fragments were subjected to cell-free extract. Analysis of the resultant cleavage products confirmed the presence not only of a novel processing activity mediated by a ribonucleoprotein complex but also of a novel processing pathway. Based on the locations of the novel processing sites within the primary 16S-23S transcript, a model for transcriptional regulation independent of polycistronic linkage is presented.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

1.0 HISTORIC OVERVIEW

Our understanding of the prebiotic origins of life and of biological diversity is biased by the interpretations of previous work manifested through the linguistic constraints of taxonomy, and in turn shapes our own approaches to subsequent research.

Prior to 1977, in the absence of conceptual foundations and technical approaches, the nature of phylogenetic relationships between extant organisms appeared inscrutable to rigorous analysis[1]. The underlying presumption, based on relative differences in what was broadly termed "complexity", was that evolution was a linear progression from the phenotypically simple prokaryotes - literally a reference to cells in a simple state "before the evolution of a nucleus" - to the phenotypically more advanced eukaryotes[2]. Empirical support for this position was limited to a small set of general observations. First, the presence of a membrane-bound nucleus enclosing the bulk of genetic material was seen, perhaps correctly, as an advancement from a more primitive cytoplasmic prokaryote-like arrangement; second, the complexity of genome organization into discrete chromosomes, and the eukaryote's tendency to monocistronic gene regulation and expression was seen, perhaps wrongly in retrospect, as indicative of an evolved sophistication.

The historical significance of archaebacteria is that their study undermined the foundations of this dichotomic dogma. Originally seen as a disjointed collection of idiosyncratic prokaryotes unified only in their inability to conform to taxonomic expectations, deeper study revealed underlying similarities in lipid composition, metabolism, and apparent adaptation to highly unusual ecological niches, and rendered the pre-existent conceptual background suspect.

The advent of two technologically driven disciplines - macromolecular sequencing and massive data analysis and storage - provided scientists with the ability to examine evolutionary processes which manifested themselves at a genotypic level unaccompanied by significant

phenotypic change. The ability to study deep evolutionary processes at this level entails a number of considerations. First, the genes studied have to be primordial, and therefore universally present, immediately excluding most elements of central metabolism and cell structure; second, they must be functionally conserved despite genetic variability.

Since the empirical notion of a unique and common ancestry was pivotal to the birth of phylogenetic analysis, the focus on genetic information processing systems - by definition present in the progenote - was a rational choice fulfilling these requirements. Initial studies were restricted to ribosomal RNA molecules for pragmatic reasons (they are easy to isolate and amenable to analysis using technologies then available) as well as conceptual ones (they are universally distributed, constant in function over all taxonomic groups, and highly conserved in both sequence and structure). Recently, these studies have been expanded to include genes coding for ribosomal proteins, RNA and DNA polymerases along with their associated factors, tRNA synthetases, and other universal elements of genetic information processing systems[3].

Divorced from the constraints of taxonomic construction based on a limited set of phenotypic characteristics, phylogenetic analysis began to reveal the minutiae of organismal relationships and evolutionary distances, permitting insight into the nature of the common ancestral state. It quickly became clear from the work of Woese and Fox that the existing taxonomic distinction between cells with and those without a nucleus was an artefact[4]. Perhaps more important from a philosophical perspective was the recognition of the underlying anthropocentricity which presumed evolution to be a linear progression from prokaryotic "simplicity" to eukaryotic "complexity". It was found that all extant life forms did not cluster around a linear continuum, but could be grouped into one of at least three apparently distinct genomic lineages. The eukaryotes remained a monophyletic kingdom, whereas what had been known as the prokaryotes did not appear to be a phylogenetically coherent group. There existed at least two types of prokaryotes, the true bacteria and the archaebacteria, each as distinctive as the eukaryote.

This revised postulate of a non-linear evolution, while important in dismantling the previous theory and replacing it with one more resonant with the data, may itself have been

FIGURE 1

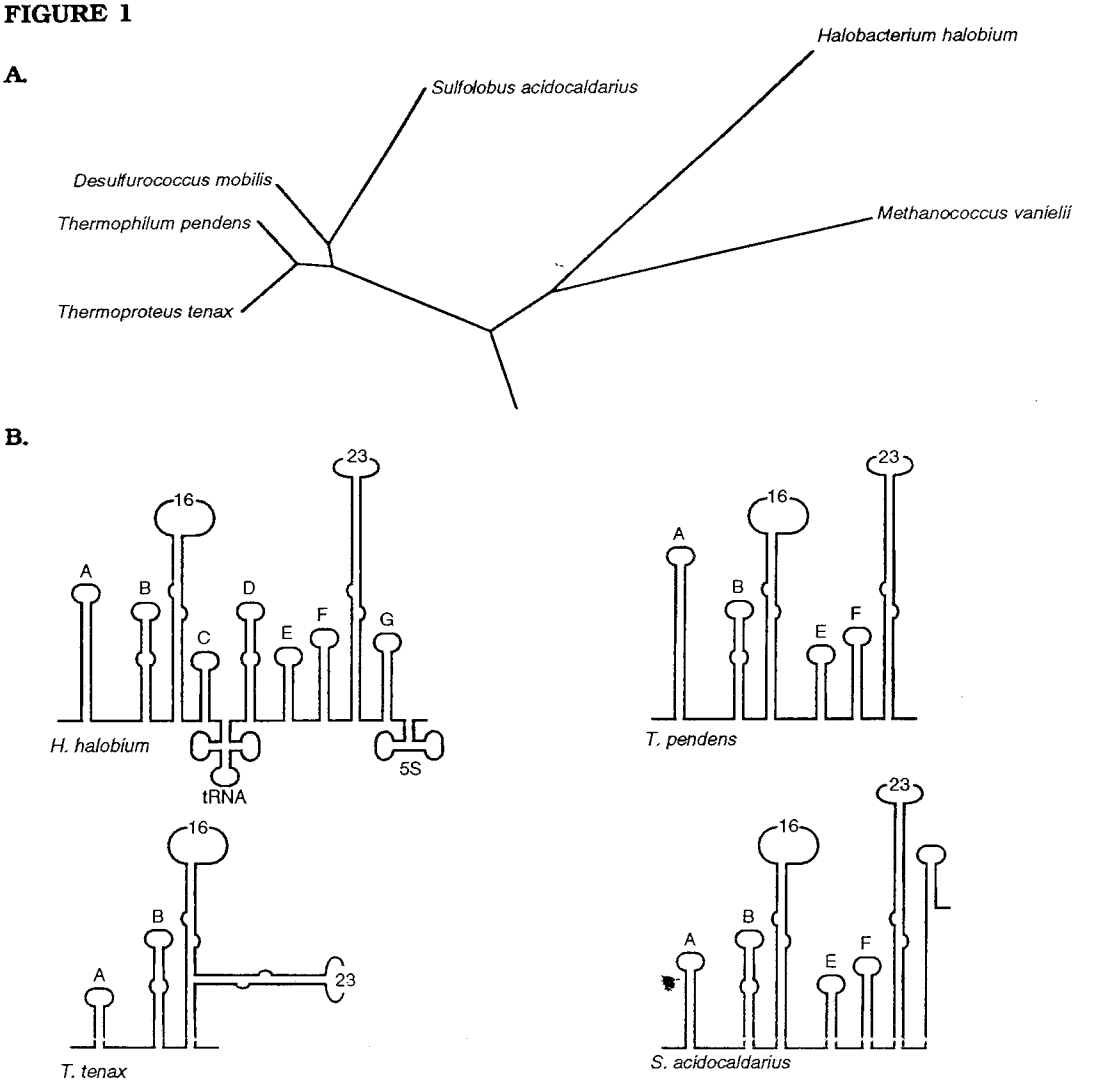


Figure 1: Comparative Phylogeny.

A. The Archaebacterial Phylogenetic Tree. The tree was derived from 23S rRNA sequences of the archaebacteria, and rooted relative to *Bacillus*. The extreme thermoacidophiles are on the left, and the methanogens and halophiles on the right. The branch lengths correspond to evolutionary distance. Other analytical methods yield trees with no common root for the archaebacterial branch[5].

B. Gene Organisation and Structure of Archaebacterial rRNA Transcripts. In *H. halobium*, the representative halophile, the operon content and gene organisation resemble those of the eubacteria. The other three strains, all sulfur-dependent thermoacidophiles, represent the extent of rRNA transcript structural variability. Helical nomenclature is taken from Garrett et. al.[5]. The complex structures of the 16S and 23S genes have been reduced to symbolic loops for the sake of visual clarity.

flawed by an unexamined presumption. Woese may have been guilty of presuming that archaebacterial distinctiveness implies archaebacterial taxonomic unity, a conclusion supported in part by his choice of statistical analysis techniques. According to his analysis, the seemingly monophyletic archaebacteria could themselves be divided into at least two major groups, one consisting of the sulfur-dependent extreme thermoacidophiles, the other of the methanogens and the derived extreme halophiles[6].

The phylogenetic rift between the two major groupings appears to be so deep however that the monophyletic nature of archaebacteria has been brought repeatedly into question. Many characteristic differences, including morphological structures on the large ribosomal subunit and the protein content of the ribosomes and RNA polymerases, exemplify this division and bring into question the notion of archaebacteria as a phylogenetically coherent group[7,8,9]. (For example, halobacterial ribosomes, like those of the eubacteria, contain few proteins greater than 30,000 daltons in mass, whereas the ribosomes of sulfur-dependent archaebacteria, like those of the eukaryotes, contain many such proteins.)

Proceeding from theoretical considerations, several statistical strategies have been developed subsequently to quantitate both the similarities and the differences in the data obtained from the sequences of homologous genes in different organisms. Although most methodologies usually agree on results for distantly related organisms, phylogenetically close groupings often give results contingent on the statistical method chosen. Three elements contribute to these discrepancies. First, it is impossible to verify the validity of the underlying conceptual bases by applying them to the data set and checking the constructed phylogenetic tree against the "true" phylogenetic tree, since the "true" tree is not known *a priori*. Second, evolutionary mechanisms active in one gene or organism need not be active in another. Phylogenetic trees, showing only the resultant distance or branching pattern without regard to the underlying mechanism, risk emphasizing methodological artefacts. Finally, substitution rates need not be constant between different branches or between different generations within a given branch.

One approach to shedding light on archaebacterial origins is to examine phenomena associated with components of the genetic information processing systems. As a companion data set, comparisons of enzymatic processes responsible for transcription initiation, termination, and processing may help clarify underlying mechanistic as well as structural similarities despite differences in molecular sequence data, and may ultimately aid in the discovery of new genes peripherally active in genetic information processing systems from which additional sequence data may be derived.

The objective of the present study is to analyze phenomena associated with the ribosomal RNA operon of the sulfur-dependent extreme thermoacidophile, *Sulfolobus acidocaldarius*, and compare them to analogous phenomena in eubacteria and both major branches of the archaebacteria.

1.1 PROMOTERS

It was noted by Pribnow after the sequence of the first few eubacterial promoters was obtained that each contained a hexanucleotide TATAAT sequence centred about 10 base pairs upstream of the mRNA transcription start point[10, 11]. Variants of a hexanucleotide TTGACA sequence about 35 base pairs upstream of the start site was identified subsequently[12, 13]. There appears to be a favoured spacing of 17 base pairs between the -35 and the -10 sequences.

The consensus sequence derived from a compilation of many diverse promoters gives rise to the strongest promoter, so that mutations which enhance the similarity of a particular promoter to the proposed consensus also enhance its activity[16]. The fact that sequences of *bona fide* promoters in *E. coli* often do not contain striking homologies to the consensus sequence, and that no wild-type *E. coli* promoter has been found that matches perfectly the consensus promoter suggests that promoter function is optimized to its particular context-specific need and not maximized to a state of constitutive transcription.

There are approximately 3000 molecules of RNA polymerase within an exponentially growing *E. coli* cell. Due to the combined effect of gene copy number, transcript size and

promoter strength, synthesis of stable RNAs in *E. coli* comprises about one half of the catalytic activity of RNA polymerase during exponential growth. The frequency of chain initiation by RNA polymerase, dependent on the nature of the promoter, occurs over a dynamic range of about 10^4 between promoters, with a maximum observed frequency of about one chain per second[14, 15].

The relation of structure to function is central to an understanding of biological processes[16]. There is a strong correlation between the nucleotide positions detected by protection assays of DNA at RNA polymerase contact sites and those that were implicated in promoter function by sequence comparisons alone. Physical studies show eubacterial RNA polymerase to be a large molecule that can span the entire -10 to -35 region, covering up to 65 base pairs upstream of the mRNA transcription start site[17, 18]. While the -35 region has been implicated in the recognition by the promoter-specific σ component of the RNA polymerase, the downstream -10 region has been implicated in transcription start site selection. The DNA duplex between the -10 region and the transcript start site is thought to be unwound or melted out by the RNA polymerase upon its entry into the promoter site[19]. The eubacterial promoter does not appear to contain a conserved sequence at or near the site of transcription initiation, although transcription was found to initiate predominantly from a single position 6 to 9 base pairs beyond the final T residue of the -10 region, usually from a template-specified purine. This bipartite model of the promoter is probably an oversimplification, however; both activation and repression of gene expression have been observed in *Escherichia coli*.

Virtually all of the known activator binding sites are located near or upstream of the -35 region, perhaps to minimize steric hindrance of the RNA polymerase. Positive activation by the CAP-cAMP complex for the *lac* and *gal* promoters seems to be dependent upon binding of the complex to a region of symmetrical DNA sequence located about 40 base pairs upstream of the mRNA start site[19, 21].

Repressors of transcription initiation act by binding to specific sites in regions as far upstream as -60 and as far downstream as +12. The tightest binding repressors can result in an absolute shut-off of chain initiation. Repression can be modulated by controlling the affinity of the protein to its specific DNA sequence. The *lac* repressor protein, for example, upon

binding the *lac* operator, inhibits transcription initiation. Because the operator overlaps the promoter, the polymerase does not bind to the promoter until the repressor is removed from the operator, making binding of the repressor and of RNA polymerase mutually exclusive competitive events. The repressor-operator complex is destabilized however by the binding of β-galactoside compounds by the repressor, allowing for polymerase binding and transcription initiation[21].

Regulation can also occur as a result of structural and conformational changes, such as methylation and negative supercoiling of the DNA template. Supercoiling results in considerably more diversity in the patterns of promoter strength than do mutations or ancillary proteins. The expression of the *gyrA* and *gyrB* genes, encoding the subunits of gyrase, is dependent on the supercoiling of DNA around the promoter, with transcription enhanced by the inhibition of gyrase activity[22].

In *E. coli*, each rRNA operon contains two tandem promoters, the first of which is responsive to both stringent control by ppGpp (itself responsive to either amino acid starvation or carbon and energy shiftdown) and to growth rate control; the second promoter is responsible for basal expression of the genes. Furthermore, an antitermination control mechanism has been postulated to prevent premature termination.

Archaebacterial promoter analysis was performed by a combination of nuclease S1 mapping and phylogenetic comparison of the upstream regions of genes encoding stable rRNA genes from a variety of species[24]. These upstream sequences indicate well-conserved elements thought to be involved in transcription initiation. Overlapping the mapped site of transcription initiation, the box B sequence consists of variants of the tetranucleotide motif (T/A)TG(C/A), with transcription usually initiating at the central G residue. A highly conserved TTTATA hexanucleotide box A sequence resides approximately 25 nucleotides upstream of the box B sequence. In these respects, the archaebacterial promoter is reminiscent of the eukaryotic RNA polymerase II promoter, which is characterized by an TATA box upstream of a short conserved sequence close to the site of transcription initiation[25]. The spacing between these two elements is roughly the same in both eukaryotes and archaebacteria[26]. Using deletion and linker substitution mutagenesis, an archaebacterial promoter region was dissected to examine

in more detail the functional significance of the promoter elements[27]. The transcription start site distal box A motif was found to be essential for transcriptional efficiency. It was also identified as the primary determinant of start site selection, in sharp contrast to the situation in the eubacterial promoter.

In contrast to the two tandem rRNA operon promoters in eubacteria, the rRNA operons of *Halobacterium cutirubrum* were shown to contain eight conserved 80 base pair repeats, each containing a functional promoter sequence[28, 29, 30]. An additional promoter sequence in the 16S and 23S intergenic space was discovered subsequently[31] and shown to be transcriptionally active at a relatively low rate[32]. It was suggested that the intergenic promoter does not violate the idea of a polycistronic rRNA operon; since the bulk of the transcripts are polycistronic, the intergenic promoter serves merely to fine-tune cellular levels of rRNA to stoichiometric levels in case of the disruption of molar equivalency. In effect, the promoter is the functional equivalent of the eubacterial antitermination mechanism.

A similar multiple promoter system, consisting of five promoters located 5' to the 16S gene, was shown to exist in the rRNA operon of the extreme thermoacidophilic archaebacterium *Desulfurococcus mobilis*[33]. On the basis of nuclease S1 mapping experiments, the system was postulated to differ from that of *H. cutirubrum* in that transcription initiating at an upstream promoter terminates in the vicinity of the next downstream promoter. The model proposed is one of polymerase loading at the single rRNA operon, presumably to maximize transcriptional activity.

In *Sulfolobus* sp. B12, the 16S-23S operon was shown to have a simple promoter structure: the sole upstream promoter was found to drive transcription of both the 16S and 23S genes[34]. It, and the promoter driving transcription from the unlinked 5S gene conformed well to the derived archaebacterial consensus sequences.

1.2 TERMINATORS

Comparisons of sequences at eubacterial termination sites have revealed three common features, generally found about 35 base pairs upstream of the transcription stop site[35]. First, an inverted repeat sequence can be found preceding known termination sites. There is substantial variation in both the length of the inverted repeat, determining the stem length, and the distance between the repeats, determining the loop size. Second, a G/C-rich sequence of variable length precedes the termination site. The G/C base pairs are thought to be important in stabilizing the stem and loop structures formed. Transcript termination has been shown to fail *in vitro* if GTP is replaced by ITP in the transcription reaction. Thus, in the case of *rho* factor-independent termination, the transcript is implicated in its own termination event. Finally, transcription tends to end in a run of four to eight consecutive T residues, with a short stretch of encoded U residues usually found at the terminus of the transcript[36]. The only terminators known to lack the oligo-T tract exhibit an absolute dependence on the *rho* factor for termination.

Controlling the amount of transcriptional read-through, or antitermination, in the region of the terminator represents an additional level of transcriptional control, independent of the promoter[40]. In the case of the tryptophan operon, for example, transcription termination is coupled to the levels of charged $tRNA^{trp}$ through translation of the small leader RNA transcribed from the region between the promoter and the attenuator[41, 42]. The 3' end of the *trp* leader can form one of two alternate and structurally exclusive stem and loop structures. Formation of the first, larger, and thermodynamically more stable stem precludes the formation of the termination structure, thereby inhibiting the termination event, and allowing read-through. In the presence of charged $tRNA^{trp}$ molecules, ribosomes, failing to stall in their translation of the leader sequence upon encountering two consecutive UGG codons, read through and sterically hinder the formation of the first structure. The termination stem forms instead, preventing transcription of the tryptophan biosynthetic genes.

An interesting example of the mechanistic variability available from a small subset of regulatory elements is the overlapping of promoter and terminator elements in *E. coli*. Prevalent mostly in phage, the system also exists within eubacterial genomes. For example, a terminator site for the *S10* ribosomal protein cistron overlaps with the -35 promoter site of the *spc* ribosomal protein cistron[43], allowing for additional levels of transcriptional control for interrelated genes within different operons. Elsewhere, the *tuf B* locus overlaps the promoter of the *sec E* locus, which in turn overlaps the promoter of the *L11* locus.

Archaebacterial transcriptional termination is not well characterized; agreement on what constitutes archaebacterial terminator consensus has not been reached[45-47]. The rRNA genes examined in thermoacidophilic[48], halobacterial[29], and methanogenic[49] species possess tetra- or penta-pyrimidine stretches in the vicinity of the mapped termination sites. A closer look reveals three or more fundamentally different sequence and structural motifs tentatively associated with termination.

Some of the halobacterial rRNA transcripts contain sequences that are virtually co-structural with eubacterial *rho* factor-independent terminators[28]. Two sequences of interest are found in the *H. cutirubrum* rRNA transcript 3' distal region. First, an inverted repeat capable of forming a 13 base pair helix followed by the sequence TTTTT is found immediately downstream of the 5S gene. The significance of this proposed structure is unclear, however, as no 3' terminus has been mapped to this sequence. A G/C-rich tract followed by an A/T-rich tract, similar to known terminators in other archaebacterial species, occurs downstream of this proposed stem-loop structure. No secondary structural elements are evident[29]. Again, no 3' terminus has been mapped to this sequence. In *Halococcus morrhuae*, a similar terminator-like helix can form, but lacks the subsequent oligo-U sequence[48]. A U_4 sequence is found, however, within the stem's apical loop. Both of these terminator-like structures located in halobacterial rRNA transcript 3' distal regions are followed by $tRNA^{Cys}$ genes. The significance of this is unclear.

A slight variant of this structural motif is found in the rRNA genes of methanobacteria. In *Methanobacterium thermoautotrophicum*[50], *Methanococcus voltae*[49], and *Methanococcus*

vannielii[51], termination occurs in a high T-residue content pyrimidine-rich sequence that precedes a non-coding inverted repeat. Termination must by necessity be different from the eubacterial *rho* factor-independent termination, which requires the inverted repeat to be transcribed before the RNA polymerase terminates the transcript. It is interesting to note that in the same organism, termination of the highly expressed structural genes for methyl CoM reductase α subunit is halobacterial-like in nature, occuring in an oligo-T region downstream of a hairpin structure[54].

The mechanism of transcriptional termination for the rRNA genes appears to be radically different in the sulfur-dependent thermoacidophilic archaebacteria. In both *Thermoproteus tenax*[55] and *Desulfurococcus mobilis*[33], termination occurs within a pyrimidine-rich region, which resembles the motif preceding the multiple terminators in eukaryotic 28S rRNA[56]. The experimental procedures used to map these archaebacterial terminators could not distinguish between nonspecific termination and exonucleolytic activity to explain the 3' heterogeneity observed.

It is not at all clear whether the apparent segregation of terminator motifs in archaebacterial rRNA operons has inherent deep evolutionary significance, or whether rRNA termination motif flexibility is common, requiring as a minimum only a T-residue rich sequence.

The spatial overlap of promoters and terminators, found predominantly in the eubacterial phage genomes, has also been observed in the first archaebacterial phage to be analyzed extensively. The UV-inducible phage SSV1, found in *Sulfolobus* sp. B12, contains typical thermoacidophilic terminators for the upstream transcripts nested between or juxtaposed upon the box A and box B sequence motifs for transcription initiation of the downstream transcripts[57]. Since promoters and terminators in eubacteria and archaebacteria are fundamentally different and the sequences are unrelated, the phylogenetically dispersed occurrence of this terminator-promoter overlap motif probably reflects more the strict requirement for phage genome compactness than it does any evolutionary implications of cross-kingdom similarity in transcriptional control.

1.3 Shine-Dalgarno Sequences

In all three archaebacterial branches (exemplified by *H. cutirubrum*, *M. voltae*, and *S. solfataricus*), purine-rich sequences complementary to the 3' end of the 16S rRNA have been found preceding the translation initiation codons of a variety of mRNA species[2]. It has been postulated that archaebacterial translation initiation may dependent on these purine-rich sequences in a manner similar to that of *E. coli*. However, the *H. halobium bop* gene (coding for bacterio-opsin), as well as some but not all *H. cutirubrum* ribosomal protein genes lack any such complementarity to the 16S 3' terminus in the region preceding the initiation codon, but do show complementarity in the immediate downstream translated region; these observations seem to indicate that the requirement for such Shine-Dalgarno sequences may not be absolute, or may only be restricted to a subset of genes. It has not yet been demonstrated that these complementary mRNA sequences actually function to position the ribosome correctly relative to the AUG initiation codon.

1.4 Conservation of rRNA Operon Gene Organization

The rRNA genes of eubacteria are clustered into polycistronic transcriptional units. Seven such rRNA operons exist in *E. coli*; their distribution on the chromosome is such that four of these are within 10 minutes of the origin of replication on the linkage map[58]. In eubacteria as well as in the chloroplasts[59-61], the gene order is 16S, tRNA, 23S, and 5S. Some rRNA operons contain additional tRNA genes distal to the 5S rRNA. In *E. coli*, 26% of the tRNA genes occur in the rRNA operons; in *Bacillus subtilis*, the clustering is more pronounced, with most of the tRNA genes being associated with rRNA operons[62]. The spacer regions flanking the structural genes are highly conserved between the different operons within the genome, attesting to their structural and functional importance. The advantages of evolving polycistronic gene organization for rRNA genes include the integration of cellular processes, like rRNA and tRNA synthesis, as well as assuring that the production of equimolar amounts of

rRNA molecules is maintained, allowing for the complex coordination of ribonucleoprotein structure assembly to yield the two ribosomal subunits.

Whereas the nature of promoter elements seems to attest to the unity of archaebacteria, the transcriptional linkage of rRNA genes differs sharply between the methanogens and halophiles on the one hand and the thermoacidophiles on the other. With the exception of the anomalous non-sulfur-dependent thermoacidophile *Thermoplasma acidophilum*[63, 64], where all three genes are transcriptionally unlinked, the rule amongst methanogens and halophiles is for a close linkage of the 16S, 23S and 5S genes, typically transcribed in this order. Furthermore, $tRNA^{Ala}$ genes have been found in the spacers between the 16S and 23S genes of these organisms. Finally, the copy number of rRNA operons varies between one and four for the methanogens and one or two for the halophiles. Clearly, the content, gene order, and copy number of rRNA operons in halobacteria and methanogens resemble those of the eubacteria.

These characteristics are sharply different in the sulfur-dependent archaebacteria, all of which possess only a single copy of the rRNA operon. In each case the 5S rRNA gene is unlinked to the 16S-23S rRNA operon[34, 55, 65, 66]. The sulfur-dependent archaebacteria are further distinguished by generally shorter intergenic regions containing no tRNA genes[67].

1.5 Transcript Structure

Internally complementary sequences capable of forming RNA duplexes between 5' and 3' sequences seems to be a common theme in both mature tRNA and pre-rRNA transcripts[68]. In *E. coli*, both the 16S and 23S rRNA precursor transcripts, upon binding ribosomal proteins, form giant ribonucleoprotein loops held together by their rRNA precursor processing stems[69]. Since these double-stranded RNA helices are the most stable secondary structural features in pre-16S and pre-23S rRNAs, they are readily observed by direct electron microscopy[58]. While the mature 23S termini form part of the helical region in the processing stem, mature 16S sequences do not. This structural conservation in organisms as diverse as eubacteria and archaebacteria suggests that these features must be important[70]. The benefits of maintaining

such complex spacer regions for rRNA genes are clear. They promote the formation of secondary structure during transcription of rRNA, as evidenced by the fact that pre-rRNA seems to form ribosomal particles *in vitro* more easily than does mature rRNA[71]. Furthermore, since the structural genes exist in the context of a longer transcript, the protection of rRNA termini from exonucleolytic activity is assured until the macromolecular assembly is structurally complete enough to protect its own termini[69]. (By way of example, it is thought that because of their compact structure, tRNA molecules are completely resistant to nucleases in E. coli *in vivo*, even under extreme conditions; the argument that structure confers protection is strengthened by experimental work showing tRNAs exhibiting mutations that weaken secondary structure are degraded[73]. Finally, the structural elements may act as signals for regulatory events during transcriptional initiation, transcript processing, and subunit assembly[62].

In addition to the conservation of rRNA operon gene organization across the eubacteria and archaebacteria in general, and the halophilic-methanogenic branch in particular, there is a remarkable conservation of stem-loop structural motifs, formed by the self-complementarity of RNA sequence transcribed from long, nearly perfect inverted repeats surrounding the large rRNAs. Amongst the archaebacteria, the transcripts generated by the extreme halophiles are structurally the most complex. As in the eubacteria, long inverted repeats flank both of the large rRNA genes, allowing for the formation of extended helical structures. While all archaebacteria exhibit several other partially conserved stem-loop helical structures, defined by their positions relative to the main processing stems[74], three are exclusive to the extreme halophiles, as exemplified by *H. halobium*. Helices C and G, preceding the tRNA and 5S rRNA, respectively, may participate in their processing from the primary transcript, and helix D, located within the 16S-23S intergenic spacer region, contains a promoter sequence shown to be transcriptionally active (Figure 1B).

Transcripts generated by the methanogenic and sulfur-dependent archaebacteria are structurally simpler, although no less interesting. Archaebacterial helices A, B, E, and F are present on the transcripts of most sulfur-dependent archaebacteria, with *T. tenax* and *P. occultum* being notable exceptions. In addition to an exceptionally short intergenic spacer, the

T. tenax transcript deviates most from a universally conserved structure in its bifurcated rRNA precursor processing stem .

1.6 Transcript Processing Sites and Pathways

In *E. coli* and other eubacteria, processing of the primary rRNA transcripts is a multistep, multienzymatic process. Full-length 30S polycistronic transcripts are not detectable in wild type eubacteria. In fact, it has been shown that the 16S rRNA precursor has been excised by the time the RNA polymerase moves on to transcribe the 23S gene[69]. The first step toward maturation is the cleavage by RNase III, a double-strand specific RNA endonuclease known to be critically dependent on the concentration of monovalent cations[75]. This enzyme has been shown to be responsible for the RNA processing of many other transcripts, including those of some T-phage. With notable exceptions in T-phage transcripts, RNase III cleaves both strands of its double-stranded RNA target at intervals of approximately 4 base pairs[76]. In both *E. coli* and *S. typhimurium*, this RNase III-mediated removal of the large rRNA precursors from the growing (and therefore incomplete) transcript occurs at staggered positions on opposite sides of the double-stranded stem, and is thought to involve two separate, but not necessarily independent reactions[58].

Upon RNase III cleavage, the 16S precursor retains long RNA sequences at both its 5' and 3' termini, whereas the 23S precursor retains only very short extraneous sequences, capable of forming less than 10 base pairs of additional helical structure. This difference may be related to the final secondary structures of the rRNA species, as the 5' and 3' termini of the mature 16S RNA are widely separated in the ribosome, whereas those of the 23S remain closely associated[62]. If, as has been suggested, the proximity of the 5' and 3' termini to each other in both the 16S and the 23S precursor are important for ribosomal protein binding and subunit assembly, then the 16S precursor would have to ensure its termini remain in contact via extraneous precursor sequences, whereas in the 23S precursor, the intrinsic complementarity of

the terminal sequences would be able to perform this task. Maturation to a functional state for the 16S rRNA, but not for the 23S rRNA, would require spatial separation of the two ends.

Not very much is known about the eubacterial 16S and 23S maturases. In the case of 23S rRNA, the proximity of generated 5' and 3' termini within the precursor structure may suggest that a single enzymatic activity is responsible for both events. This logic is not compelling however; tRNA maturation offers an empirical counter-example[77]. The reaction that generates the 3' end of the mature 16S rRNA, thought to be an endonuclease, is not as well characterized as the one that generates the 5' end. Whereas the ends generated by the 16S maturases are quite homogeneous, it appears that the *E. coli* 23S rRNA termini are somewhat heterogeneous, with several discrete species constituting each end[58].

Although the general mechanism for the release of rRNA precursors in archaebacteria appears to be similar to that of the eubacteria, very little is known of the details of processing. Two opposing bulges consisting of three unpaired nucleotides are staggered by three to four base pairs within the central portion of the 16S and 23S precursor processing stems; these constitute the processing sites responsible for the liberation of rRNA precursors via an endonucleolytic cleavage in a manner analogous to that of the RNase III-dependent precursor processing found in the eubacteria[49]. It has been inferred that, as in the case of eubacterial RNase III, the archaebacterial activities within the staggered cleavage sites are not ordered[33]. No conserved sequence motif is apparent around these endonucleolytic sites, but the purine content of the bulges is generally high[55]. As in the eubacteria, the processing order for the genes has been determined; 16S rRNA processing precedes 23S rRNA processing, probably for much the same set of reasons.

If understanding of eubacterial rRNA maturation can be said to be minimal, for archaebacteria it is virtually non-existent. Based on the varying quality of their mapping results, researchers have postulated that transcript maturation proceeds by either endonucleolytic cleavage, or alternatively by exonucleolytic degradation of processing intermediates.

1.7 Alternative Processing Pathways and Auxiliary RNases

While unprocessed tRNA precursors clearly cannot function in protein synthesis, the same apparently cannot be said about eubacterial rRNA; it is possible that very few of the rRNA processing reactions are required absolutely for the resultant rRNA species to function[58]. In *E. coli* RNase III mutants, uncleaved 30S transcripts are detectable, but are found not to be the precursors of functional 16S and 23S species; instead, functional rRNA molecules are derived from the cleavage of the growing transcript during transcription via some alternative processing pathways[23]. These processing events yield RNA intermediates not found in RNase III$^+$ cells. Most of the functional 23S RNA in the mutant is present as a longer precursor, with up to 97 and 53 extra nucleotides at the 5' and 3' ends, respectively. Since the RNase III$^-$ mutants, producing only abnormally long 23S RNA species, are still viable, it must be concluded that the resultant ribosomes are capable of carrying out protein synthesis[78].

In the RNase III$^-$ mutant, the 16S RNA matures normally despite the absence of RNase III processing of the 16S precursor processing stem[79]. Furthermore, experimental evidence has shown that deletion of up to 2/3 of the normal 16S rRNA leader, with a concomitant disruption of the 16S precursor processing stem past the RNase III processing site, does not prevent the formation of proper mature termini[70]. Because the removal of the self-complementary stem at the base of the 16S RNA precursor is essential to permit its 5' and 3' termini to be far apart, as they are in the mature and active 30S subunit of the ribosome, enzymes responsible for either alternative processing or intermediate (as opposed to precursor) processing must be active in the cell. This is neither an unlikely nor an inconceivable proposition, given the apparent necessity of spatial separation of the 16S rRNA ends.

Mutants grossly but not totally deficient in 16S rRNA processing are generally not viable[80]. Exceptions do exist, however; a mutant *E. coli* strain in which the 16S 5' terminus contains 50 extra nucleotides is slow growing but viable, suggesting that ribosomal assembly and function are not disrupted[69]. Such a mutant strain could still separate its termini despite the extra sequence at the 5' end, since the removal of extra sequence at the 3' end would

prevent the annealing of complementary flanking sequences. Mutant strains possessing extra sequence at the 3' end have not been observed. Whether such strains would be viable is uncertain; it is theoretically possible that, because the 3' end of the 16S rRNA contains the complement to the Shine-Dalgarno sequence, such terminal extensions would interfere with ribosome-mRNA association and translation initiation, resulting in protein synthesis inhibition and cell inviability.

By targetted disruption of U3 snRNA, a 220 nucleotide RNA transcript involved in cleavage at a specific site in the eukaryotic Internal Transcribes Spacer, and subsequent observation of the resultant altered levels of certain pre-rRNA intermediates, parallel rRNA processing pathways have also been shown to coexist simultaneously within a single cell in some, but not all species of *Xenopus*[93].

Given the importance of rRNA maturation, the necessity of fail-safe backup systems is obvious. Their presence in both eubacteria and eukaryotes, albeit using fundamentally different mechanisms, hints at the possibility of alternative processing pathways in archaebacteia.

2 MATERIALS AND METHODS

2.0 MATERIALS

Yeast extract, Bacto tryptone and Bacto agar were purchased from Difco Laboratories.

E. coli strains JM101, JM109, KW251 and LE392 were purchased from or obtained through Promega. Strain DH5α was obtained through BRL.

Vector plasmids pUC12, pUC13, pUC18, pUC19 and pBR322 were purchased from New England Biolabs. Vector phage M13mp12, M13mp13, M13mp18 and M13mp19 were purchased from Pharmacia. The vectors pGEM-3Zf(+/-), pGEM-5Zf(+/-) and pGEM-7Zf(+/-), helper phage R408, and λGEM11 packaging extracts were purchased from Promega.

Most restriction endonucleases and DNA and RNA modifying enzymes were purchased from Pharmacia. Proteinase K and MMLV-RT were purchased from BRL, Sequenase was purchased from United States Biochemical, RNAse T_1 was purchased from Boehringer Mannheim, and exonuclease III was purchased from Promega.

Ribonucleotides, deoxyribonucleotides and dideoxyribonucleotides were purchased from Pharmacia LKB Biotechnology. Radioactive α(^{32}P)NTPs, dNTPs, γ(^{32}P)ATP and random hexadeoxyribonucleotides were purchased from DuPont NEN Research Products. Sequence-specific oligonucleotides were synthesized at NAP UBC Oligonucleotide Synthesis Laboratory.

Acrylamide and N N' methylene bisacrylamide were purchased from BioRad Laboratories, genetic technology grade agarose was purchased from ICN Biomedicals, cesium chloride was purchased from Cabot Chemicals, β-mercaptoethanol (βME) was purchased from Matheson Coleman & Bell, isopropyl-β-D-thiogalactopyranoside (IPTG) was purchased from BRL and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (Xgal) was purchased from Biosynth AG. All other chemicals were purchased from either BDH, Fischer, or Sigma.

Hybond-N nylon membranes and filters, and Hybond M&G paper were purchased from Amersham. Zeta-Probe membranes were purchased from Bio Rad. Films for autoradiography (X-

Omat and XAR) were purchased from Kodak, and film for visualization of ethidium bromide stained DNA was purchased from Polaroid.

2.1 STRAINS, VECTORS, AND MEDIA

2.10 *Sulfolobus* Species Identification

A longstanding nuisance for *Sulfolobus* researchers has been the uncertainty regarding strain identification (Figure 2). Unbeknown to researchers, the two original isolates, Brock's *Sulfolobus* strain 98-3 and Pisciarelli's *Sulfolobus solfataricus* P1, were probably both mixed cultures with one predominant species. Differences in culturing techniques would favour the propagation of one species over another so that in time, the culture would become pure.

Brock's composite strain 98-3 was characterized at the Deutche Sammlung as *S. acidocaldarius*. Langworth, relying on the Deutche Sammlung nomenclature, also named his strain, obtained from Brock, as *Sulfolobus acidocaldarius* . He was, however, dealing either with a mixed culture, or a culture overgrown with *Sulfolobus solfataricus*. The culture was passed on and became Woese's original *Sulfolobus acidocaldarius* species.

Pisciarelli's *Sulfolobus solfataricus* strain P1, also a mixed culture with *Sulfolobus solfataricus* predominating, was cultured so as to favour the growth of *Sulfolobus acidocaldarius* either in Pisciarelli's lab or when it was acquired by Woese for phylogenetic analysis. Thus, both strains were mislabelled. That the Deutche Sammlung *Sulfolobus acidocaldarius* did not correspond to Langworth's *Sulfolobus acidocaldarius* was discovered subsequently.

The subsequent development of plating techniques allowed for colony purification - and Southern hybridization for unambiguous identification - of the two thermoacidophiles, but retracing the switch in nomenclature once the mixed and/or purified cultures had been disseminated to labs around the world has proved to be an ongoing irritant and source of confusion. For example, the DNA databases contain identical 5S RNA sequences attributed to the two different *Sulfolobus* species. The confusion extends to other rRNA sequences as well.

FIGURE 2

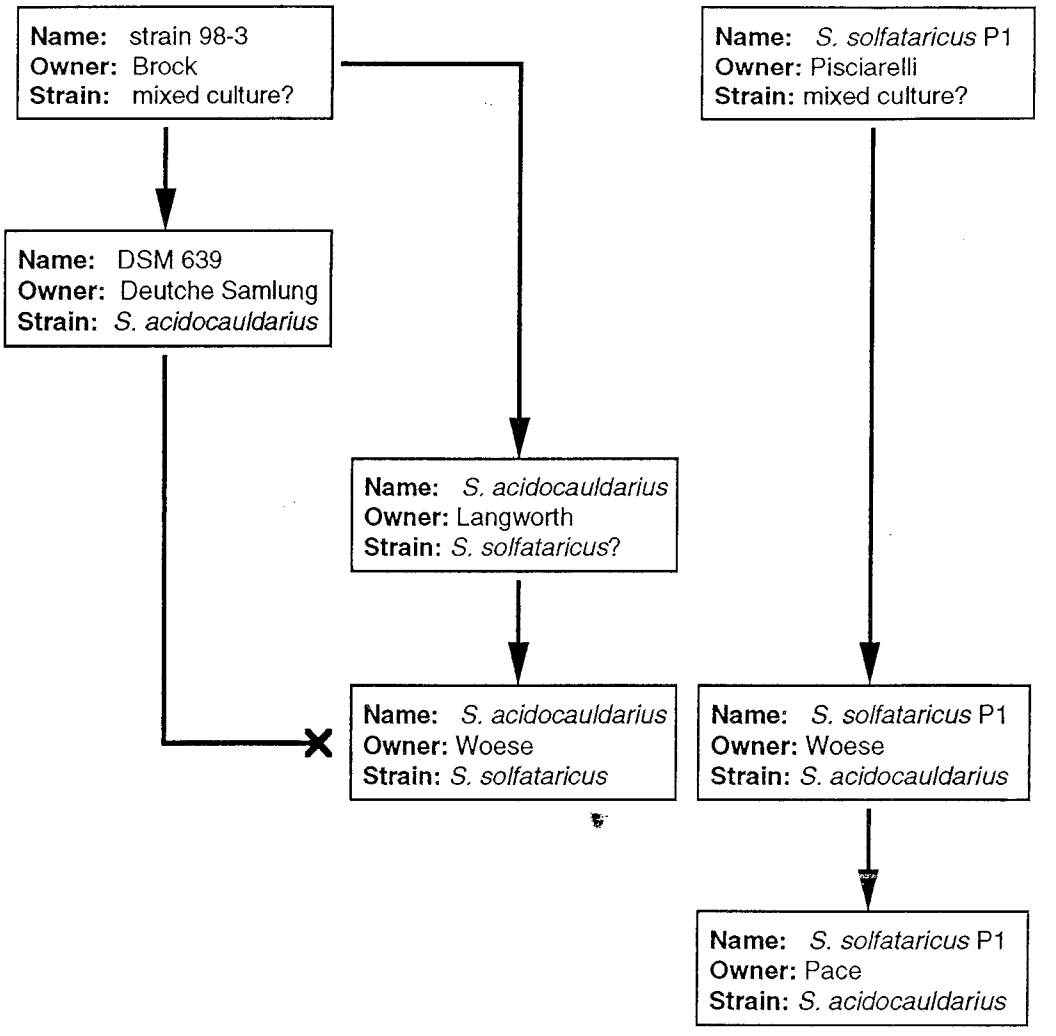


Figure 2: *Sulfolobus* Strain Histories. The histories of the original two *Sulfolobus* mixed cultures, as reconstructed by W. Zillig (personal communication). Arrows indicate the critical passage of cultures between laboratories which then dispersed the stocks to the scientific community. The "X" indicates recognised species differences between stocks originating from the same source. Question marks following strain designations indicate uncertainty regarding species identification. It is not clear at which point in the dissemination pathway, if any, the cultures became homogeneous. The strain of *S. acidocaldarius* used for this work was obtained from the laboratory of Norman Pace.

2.11 Plasmids, Phage Vectors, and Bacterial Strains

Sulfolobus acidocaldarius, previously misidentified variously as *Sulfolobus solfataricus* and *Caldariella acidophila* species, was obtained from Norman Pace. The *E. coli* strain LE392 (F-, *hsd*R514 (rk-, mk-), *sup*E44, *sup*F58, *lac*Y1, *gal*K2, *gal*T22, *met*B1, *trp*R55, λ–) was used for the construction of the *S. acidocaldarius* genomic library, and strain KW251 (F-, *sup*E44, *sup*F58, *gal*K2, *gal*T22, *met*B1, *hsd*R2, *mcr*B1, *mcr*A⁻, *arg*A81:*Tn*10, *rec*D1014) was used for the propagation of λ clones. Strains JM101 (*sup*E, *thi*, Δ(*lac-pro*AB), (F' *tra*D36, *pro*AB, *lac*IqZΔM15)), DH5aF' (F', Φ80d*lacZ*ΔM15, Δ(*lacZ*YA-*arg*F)U169, *rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17, *sup*E44, *rel*A1, λ-) and, where necessary, JM109 (*rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17, *sup*E44, *rel*A1, λ-, (F' *tra*D36, *pro*AB, *lac*IqZΔM15)) were used for genomic cloning, subcloning, and propagation of plasmids and M13 phage.

The phage vector λGEM-11 was used for the construction of the *S. acidocaldarius* genomic library. The plasmid vectors pUC18, pUC19, pGEM-3Zf(+), and pGEM-3Zf(-) and the phage vectors M13mp18 and M13mp19 were used for cloning of size-fractionated libraries, for subcloning, for the production of probes for nuclease S1 transcript mapping and Southern analysis, and for sequencing.

2.12 Media, Culture Conditions, and Storage

Liquid cultures of *S. acidocaldarius* were grown at 75°C in a well-aerated medium described by DeRosa (IR X). The medium consisted of 0.1% w/v Bacto yeast extract, 0.1% w/v Bacto tryptone, 22.5 mM KH_2PO_4, 20 mM $(NH_4)_2SO_4$, 2.25 mM $CaCl_2$, 800 μM $MgSO_4 \cdot 7H_2O$, and 0.1% v/v trace metal mix, and was maintained at a pH ranging between 3.8 and 4.2 by automated monitoring and addition of 7.5M H_2SO_4 during growth as required. The trace metal mix consisted of 875 μM $MnCl_2 \cdot 4H_2O$, 1.25 mM $Na_2B_4O_7 \cdot 10H_2O$, 75 μM $ZnSO_4 \cdot 7H_2O$, 30 μM $CuCl_2 \cdot 2H_2O$, 15 μM $NaMoB_4 \cdot 2H_2O$, 15 μM $VOSO_4 \cdot 2H_2O$, and 3.5 μM $CuSO_4 \cdot 7H_2O$. Cell stocks

of *S. acidocaldarius* were maintained for the short term by storage in growth medium at room temperature and for the long term by flash freezing cell pellets at -70°C.

The *E. coli* strains JM101, JM109 and DH5α were grown at 37°C in M9 minimal medium (50 mM Na_2HPO_4, 25 mM KH_2PO_4, 8.5 mM NaCl, 20 mM NH_4Cl, 1 mM $CaCl_2$ and 0.2%w/v glucose) as described by Miller when selection for the F factor was required, or in YT rich medium as described by Maniatis (86 mM NaCl, 0.8% w/v tryptone and 0.5% w/v Bacto yeast extract) when F factor selection was not required. Ampicillin was added to a concentration of 100 μM when required for plasmid selection, and 50 μM IPTG and 0.01% w/v Xgal were added to detect β-galactosidase activity. The *E. coli* strains LE392 and KW251 were grown in LB supplemented media (85 mM NaCl, 10 mM $MgSO_4$, 1% w/v Bacto tryptone, 0.5% w/v Bacto yeast extract, 0.2% w/v maltose). Agar plates contained 1.5% w/v agar, and top agar overlays contained 0.75% w/v agar. Cell stocks of *E. coli* were maintained in medium at 4°C for up to one month, on agar plates at 4°C for up to six months and in growth medium and 50% v/v glycerol at -70° C for long term storage.

2.13 Transformation of *E. coli*

Host *E. coli* cells were made competent for double stranded circular DNA uptake by the method described by Vierra[81] and Yanisch-Perron[82] and used within 12 hours. Overnight cultures of the appropriate host strain were grown up in minimal medium, then subcultured into YT rich medium and grown to an A_{600} of 0.5. The cells were pelleted by centrifugation in a Sorvall SS34 rotor at 5000 r.p.m. for 5 minutes and resuspended in 1/10 volume of 50 mM $CaCl_2$. For transformation, 200 μl aliquots of cells were mixed with 100 pg of DNA, left for 30 minutes at 0°C, and heat shocked for 2 minutes at 42°C. Frozen competent cell stocks were made up in transformation and storage buffer (LB broth (pH6.1), 10% PEG 3500, 5% DMSO, 10 mM $MgSO_4$, 10 mM $MgCl_2$), and stored at -70°C[83].

2.2 DNA SUBCLONING

2.20 Directional Cloning

Fragments of DNA were mapped with restriction enzymes to identify unique or rare sites, allowing for directional subcloning of smaller fragments into appropriate vectors. The potential for metabolic regulatory disruption in the *E. coli* host cells caused by the presence of ribosomal protein binding sites on RNA transcripts produced by the induction of the *lac* operator within the vector occasionally posed problems in subcloning DNA fragments, and resulted in persistent and reproducible deletions and rearrangements. Three solutions were employed to address this problem. First, the host strain JM109 was substituted for the more robust strain JM101. Second, where possible, fragments were subcloned in an orientation opposite to that of the *lac* operator so as to produce transcripts devoid of putative functional ribosomal protein binding sites. Third, where directional subcloning was not available or was found not to be effective, the β-galactosidase assay was abandoned in favour of colony or plaque screening by Southern hybridization.

2.21 Exonuclease III Generated Deletions

The subclones were further reduced in size by exonuclease III-generated deletions. In essence, the subclone was cleaved proximally to the insert at a restriction site whose recessed 3' end is an efficient substrate for exonuclease III digestion and distally at a restriction site whose 3' overhang is protected from exonuclease III digestion. A molar excess of exonuclease III digests DNA in the 3' to 5' direction in a time- and temperature-dependent manner. Nuclease S1 was used to digest the remaining single stranded DNA, Klenow fragment along with a mix of the four deoxyribonucleotides was used to fill in any recessed 3' ends, and T_4 DNA ligase was used to circularize the plasmid. Instead of yielding a evenly distributed set of deletions, the abundance of secondary structural elements within and around the rRNA genes, combined with the ability

of nuclease S1 to nick single stranded DNA, tended to result in deletions clustered around sequences coding for stem-loop structures.

2.22 Shotgun Cloning

Fragments not conducive to either directional cloning or exonuclease III-generated deletion were digested to completion with the frequently cutting endonucleases *Sau* 3A, *Taq* I or *Msp* I, and cloned into M13mp18 phage vectors. Clear plaques generated in the presence of Xgal and IPTG were prescreened by use of ddTTP sequencing (T-tracking) to eliminate sequencing redundancy.

2.3 Isolation of DNA and RNA

2.30 Isolation of *S. acidocaldarius* Genomic DNA

Liquid cultures of *S. acidocaldarius* were grown at 75°C to an A_{600} of 1.0. Since the cell wall and membrane appear to be unstable at 42°C and at 4°C, but quite stable at room temperature, the culture was quickly cooled to room temperature by swirling in an ethanol and dry ice bath. The cells were pelleted at 20°C, resuspended at 1 g cells/6 ml lysis buffer (150 mM NaCl, 100 mM EDTA, 2% w/v SDS), and lysed at 90°C for 5 minutes. The lysate was cooled by the addition of 4 ml 10X TE (100 mM Tris, 10 mM EDTA (pH8.0)), and extracted three times with redistilled phenol:chloroform (1:1) and two times with chloroform:isoamyl alcohol (24:1). To the lysate were added CsCl at 0.97g/ml and ethidium bromide to 8 mg/ml, and the DNA was purified by ultracentrifugation in a TLV 100 rotor for 6 hours at 100,000 r.p.m. The recovered DNA was dialysed against TE for 3 hours, and stored at -20°C.

2.31 Isolation of Plasmid DNA

Plasmid DNA was prepared in a manner essentially identical to that described by Maniatis[84], with the exception that lysozyme was omitted from the lysis step, and ammonium acetate was substituted for potassium acetate. Cells were pelleted from 1 ml of confluent medium by centrifugation for 1 minute in a Beckman microcentrifuge, resuspended in 100 µl resuspension buffer (25 mM Tris·HCl (pH8.0), 10 mM EDTA, 50 mM glucose) and lysed by the addition of 200 µl lysis buffer (200 mM NaOH, 1% w/v SDS). The bulk of cellular debris and genomic DNA was removed by the addition of 150 µl 7.5M NH_4OAc (pH 7.8), followed by centrifugation for 5 minutes. The supernatant was extracted once with phenol:chloroform (1:1 v/v). The DNA was precipitated by the addition of 2.5 volumes 95% ethanol, centrifuged for 20 minutes, dried under vacuum, and resuspended in 25 µl TE. Coprecipitated RNA was removed by the addition of 1 µl RNase A (1 mg/ml) followed by A 30 minute incubation at 37°C.

When the integrity of the supercoiled DNA was deemed essential, the DNA was either precipitated in 1.5 M NH_4Oac, 12% w/v PEG 8000 for 1 hour at 0°C followed by centrifugation for 20 minutes, or banded through an isopyctic CsCl gradient as described above.

2.32 Isolation of *S. acidocaldarius* RNA

For the isolation of large RNA species, cells were lysed and the proteins extracted as described in section 2.30, and solid LiCl was added to the supernatant to a final concentration of 2 M. The precipitated RNA was pelleted by centrifugation for 30 minutes at 9,500 r.p.m. in a Sorvall SS34 rotor. The RNA was resuspended in DEPC-treated water, ethanol precipitated, and centrifuged for 20 minutes. The RNA pellet was dried under vacuum, dissolved in RNA storage buffer (10 mM Tris, 1 mM EDTA (pH8.0), 0.1% w/v SDS, made up in DEPC-treated H_2O), and stored at -70°C.

For the isolation of an RNA population well-represented in small species, cells were again lysed and the proteins extracted as described above. The supernatant was layered over a

step gradient cushion (90 mM NaCl, 65 mM EDTA (pH 8.0), 1.2% w/v SDS, 5.7 M CsCl) and centrifuged for 9 hours at 100,000 r.p.m. in a TL 100 rotor. The RNA pellet was treated as described above.

For the isolation of rRNA species incorporated into mature ribosomes, cells were concentrated 10- to 50-fold and disrupted by passage through a French press. The pressate was centrifuged at 10,000 r.p.m for 10 minutes to remove debris and unbroken cells, and the supernatant was layered directly onto a 5-30% sucrose density gradient in TMA buffer (30 mM NH_4Cl, 50 mM Tris·HCl, 10 mM $MgCl_2$). The extract was centrifuged in a SW 27 rotor at 24,000 r.p.m. and 5°C for 12 hours. Collected fractions were scanned at λ=360 nm, and the 30S and 50S peaks were pooled.

2.33 Isolation of Phage DNA

Single stranded phage preparations of M13 derivatives were performed as described by Sanger [85]. Single stranded phage preparations of pEMBL and pGEM derivatives using R408 helper phage were performed as described by Dente[86].

2.34 Preparation and Screening of λ Libraries

A partially digested *Sau* 3A genomic library from *S. acidocaldarius* was constructed in λGEM-11 according to the manufacturer's instructions (Promega Technical Bulletin no. 057; Promega Protocols and Applications Guide). Serial dilutions of the restriction endonuclease *Sau* 3A were titrated against 20 µg genomic DNA to optimize fragment size distribution centred around 20 kilobase pairs. The digested DNA was extracted with phenol:chloroform (1:1 v/v), and resuspended in fill-in buffer (10 mM NaCl, 10 mM Tris·HCl (pH 7.5), 7 mM $MgCl_2$, 10 mM dATP, 10 mM dGTP). Klenow fragment (1 U) was added, and the mixture was incubated for 30 minutes at 37°C to partially fill in the *Sau* 3A-generated 3' recessed ends, leaving shortened single stranded protruding DNA (5' GA 3'). These fragments were then ligated into prepared

λGEM-11 arms (*Xho* I digested, dTTP/dCTP filled) whose protruding DNA (5'TC 3') has exclusive specificity for partially filled-in *Sau* 3A digested genomic DNA. The ligated DNA was packaged into λ particles, and stored at 4°C.

The λ phage were propagated essentially as described by Maniatis. For 200 ml preparations, $2X10^6$ p.f.u. of λ phage was added to 1 ml of an overnight culture of host cells grown in LB supplemented medium, incubated for 20 minutes at 37°C to allow adsorption of the phage to the cells, and used to inoculate 50-500 ml of LB medium. After cell lysis, chloroform (5 ml) was added to the lysate and incubation was continued for 5 minutes at 37°C. Bacterial debris was removed by centrifugation for 15 minutes at 9,500 r.p.m. in a GSA rotor, and the supernatant was removed. Cellular DNA and RNA were degraded by the addition of 200 μg each of RNase A, RNase T_1, and DNase I to the supernatant followed by a one hour incubation at 37°C. Polyethylene glycol 8000 and NaCl were added to final concentrations of 10% w/v and 1 M, respectively, and the λ phage particles were precipitated overnight at 0°C. The phage were collected by centrifugation for 30 minutes at 9,500 r.p.m. in a Sorvall SS34 rotor, and resuspended in SM buffer (20 mM Tris·HCl (pH 7.4), 100 mM NaCl, 10 mM $MgSO_4$). The DNA was extracted three times with phenol:chloroform, ethanol precipitated, dried under vacuum, and resuspended in 100 μl TE buffer.

2.4 GEL ELECTROPHORESIS

2.40 Native Gel Electrophoresis

Restriction endonuclease-digested DNA fragments were separated electrophoretically through horizontal agarose or horizontal polyacrylamide (acrylamide: N N' methylene bisacrylamide, ratio of 39:2) supports. Due to the preponderance of inverted repeats in the rRNA genes giving rise to potential secondary structural features in the DNA fragments, the rates of migration through agarose or polyacrylamide were dependent not only on size, but also on the ionic content of the buffer and the running temperature of the gels. For this reason,

electrophoretic condition were standardized: the buffer used for electrophoresis was 0.5X TBE (0.89 M Tris, 0.89 M boric acid, 0.025 M Na_2EDTA) with 250 ng/ml ethidium bromide, and electrophoresis was performed at 5 volts/cm. Under these conditions, the buffer temperature would not exceed 25°C.

Total cellular RNA was separated by electrophoresis through a vertical slab polyacrylamide support in a 0.5X TBE buffer made up with DEPC-treated water and 0.1% SDS.

2.41 Denaturing Polyacrylamide Gel Electrophoresis

DNA fragments from sequencing reactions, from nuclease S1 transcript mapping reactions and from primer extensions were separated on either 0.2 mm, 0.35 mm or 0.35 mm to 0.95 mm wedged denaturing vertical polyacrylamide gels. The gels were made in 0.5X TBE and 8.3 M urea, polymerised by the addition of $(NH_4)2S_2O_8$ and TEMED to 2 mM and 130 µM final concentrations, respectively. Electrophoresis was carried out in 0.5X TBE at 32 watts constant power. The gels were dried under vacuum.

2.5 Radioactive Labelling of RNA and DNA

2.50 Oligonucleotide Labelling

Oligonucleotides were labelled at their 5' ends. If the oligonucleotide contained a phosphate group at the 5' end, it was first dephosphorylated in SAP buffer (20 mM Tris·HCl (pH 8.0), 10 mM $MgCl_2$, 20 µg/ml DNA) by the addition of shrimp alkaline phosphatase (1 U/10 pmol DNA). The phosphatase was deactivated by heating to 65°C for 15 minutes. The labelling reaction took place in 50 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM DTT and 0.1 mM spermidine by the addition of T_4 polynucleotide kinase (1U) and 2 µl ($\gamma^{32}P$)ATP (110 mCi/ml) per picomole of oligonucleotide. The reaction was allowed to proceed at 37°C for 30 minutes. The

labelled oligonucleotides were purified either by ethanol precipitation or by exclusion chromatography on Sephadex G25 fine columns.

2.51 RNA Labelling

Because the 5' end of the 5S RNA molecule is shielded by its secondary structure, it is not an ideal substrate for the kinase labelling reaction. Instead, labelling was carried out as follows. The cytidine 3' phosphate (Cp) was labelled at its 5' end by T_4 polynucleotide kinase in the presence of $\gamma(^{32}P)$ATP under the conditions described above to yield cytidine 3',5' bis(phosphate) (^{32}pCp). This was then ligated to the 3' end of gel-purified 5S rRNA in the presence of RNA ligase buffer (50 mM HEPES, 20 mM $MgCl_2$, 3 mM DTT, 0.1 mM ATP, 10% DMSO, 0.001% w/v BSA) and T_4 RNA ligase (0.05 U/pmole RNA).

2.52 DNA Labelling

Labelling of DNA at the 5' end was carried out as described in section 2.50.

Labelling of DNA at the 3' end was carried out as described by Maniatis. DNA fragments digested with an endonuclease producing a 3' recessed end were labelled in Klenow buffer (10 mM NaCl, 10 mM Tris·HCl (pH 7.5), 7 mM $MgCl_2$) in the presence of the appropriate $\alpha(^{32}P)$ dNTP and Klenow fragment (3 U/pmol DNA) for 15 minutes at room temperature. The DNA was ethanol precipitated, dried under vacuum, and resuspended in TE buffer.

Random primer labelling of DNA was carried out by a modification of the method described by Fienberg[87]. Gel-purified DNA (50 ng) was mixed with random hexadeoxyribonucleotides (5 ng) to a final volume of 10 µl, denatured by boiling for 5 minutes, and cooled on ice. The labelling reaction was performed in a 50 µl volume of 50 mM Bis-Tris·HCl (pH 6.6), 20 mM NaCl, 5 mM $MgCl_2$, 5 mM βME, 20 µM dTTP, 20 µM dGTP, 20 µM dATP, and 0.4 µg/ml BSA by the addition of 5U Klenow and 5 µl $\alpha(^{32}P)$ dCTP (10 mCi/ml), followed by

incubation for 3 hours at room temperature. Unincorporated nucleotides were removed by ethanol precipitation of the labelled DNA.

2.6 SOUTHERN BLOTTING

2.60 Transfer

DNA was digested with a variety of restriction endonucleases and separated by size on agarose gels alongside ^{32}P-labelled size standards. The DNA was denatured *in situ* by soaking the gel for 30 minutes in denaturing solution (1.5 M NaCl, 0.5 M NaOH). The DNA was then transferred to a nylon membrane by blotting with transfer buffer (1.5 M NaCl, 0.25 M NaOH) for 12 hours. The membrane was washed in 2X SSPE (SSPE: 180 mM NaCl, 100 mM NaH_2PO_4, 10 mM EDTA (pH 7.7)) and dried at room temperature for an hour. The DNA was crosslinked to the membrane by a 2 minute exposure to UV light. The membranes were prehybridized for one hour in hybridizing solution (5X SSPE, 5X Denhardt's solution (Denhardt's solution: 0.02% w/v BSA, 0.02% w/v Ficoll, 0.02% w/v polyvinylpyrollidone), 0.5% w/v SDS) before the addition of the radioactive DNA probe. The hybridization temperature was adjusted empirically according to the nature of the probe used. After 12 hours of hybridization, the membrane was washed at room temperature for 10 minutes in 1X SSPE, 0.1% w/v SDS, twice at the hybridization temperature for 15 minutes in 0.1X SSPE, 0.1% w/v SDS, and exposed to film.

2.61 Colony and Plaque Lifts

Petri plates containing either colonies or plaques were cooled at 4°C for two hours before lifting. Nylon filters with orientation markers were carefully placed on the agar for 2 minutes. The orientation markers were transferred to the Petri plates, the filters were removed and placed colony side up onto denaturing solution for 10 minutes and then onto neutralizing solution (1.5 M NaCl, 0.5 M Tris·HCl (pH 7.2), 1 mM EDTA) for 5 minutes. The filters were allowed to air

dry for one hour before UV crosslinking. The filters were then rehydrated in 2X SSPE and the debris was gently rubbed off with a moistened tissue. Prehybridization, hybridization and subsequent washing were carried out as described in section 2.60.

2.7 DNA SEQUENCE ANALYSIS

2.70 Enzymatic Sequencing

Sequencing with Klenow by dideoxyribonucleotide chain termination was performed according to the method described by Sanger and modified by the manufacturer (Pharmacia).

Sequencing with T7 polymerase, Sequenase, and Sequenase version 2.0 by dideoxyribonucleotide chain termination was performed according to the method described by the manufacturers (Pharmacia, United States Biochemical).

Resolution of compressions, bandshifts and other artefacts introduced by secondary structure within the DNA template was addressed by a variety of means. If neither enzyme was capable of resolving a particular artefact, even with reactions run at temperatures elevated up to 45°C (with a concomitant increase in the enzyme concentration during the chase or termination step for Klenow or T7-derived enzyme, respectively), deoxyinosine 5' triphosphate was substituted for deoxyguanosine 5' triphosphate. DNA with sequencing artefacts persisting despite these efforts was subcloned in an effort to remove the offending structural element.

2.71 Maxam and Gilbert Sequencing

Chemical sequencing of DNA fragments was carried out essentially according to the method described by Maxam and Gilbert[20]. The DNA fragment was either labelled uniquely at the 3' end by Klenow enzyme or labelled at both 5' ends with T4 PNK and digested internally with a restriction endonuclease to yield a fragment with a uniquely labelled end. The labelled fragment was then purified electrophoretically through either polyacrylamide or agarose,

concentrated to 25000 disintegrations per minute per µl, as measured by Cerenkov counting, and heat denatured at 100°C for three minutes. Up to 50K cpm/reaction of end-labelled DNA was spotted onto strips of Hybond M&G filters kept at 0°C. The filters were rinsed twice with 95% ethanol, twice with distilled water, air dried for 5 minutes, and transferred to Eppendorf tubes. Since chemical sequencing was only performed to establish a size ladder with nucleotide precision for transcript mapping on DNA whose sequence had already been determined enzymatically, all four sequencing reactions were not required. The reactions performed were G (50 mM ammonium formate (pH 3.5), 0.7% DMS, 45 seconds, room temperature) and A+G (88% formic acid, 10 minutes, room temperature). The reactions were terminated by removing the strips from the reaction solutions, washing twice with distilled water, twice with ethanol, and air drying. Cleavage and removal of the DNA from the filter was achieved in one step by submersion of the filter in 100 µl of 10% piperidine and heating to 90°C for 30 minutes. The mixture was lyophilized twice and the DNA was resuspended in 2 µl of distilled water.

2.8 Transcript Mapping

2.80 Nuclease S1

Nuclease S1 protection of *S. acidocaldarius* rRNA transcripts was performed as follows. Fragments of DNA uniquely labelled at either the 5' or the 3' end were ethanol precipitated in the presence of RNA, resuspended in 20 µl S1 hybridization buffer (40 mM PIPES (pH 6.8), 400 mM NaCl, 1 mM EDTA) and 80 µl formamide, denatured for 20 minutes at 80°C, and hybridized to the RNA for 3 hours at temperatures optimized for G-C content (54°C-65°C). Unhybridized single stranded DNA and RNA was digested by the addition of 300 µl nuclease S1 digestion buffer (280 mM NaCl, 30 mM NaOac (pH 4.4), 4.5 mM $ZnCl_2$), 6 µg nonspecific single-stranded DNA, and nuclease S1 (100 U), and incubated for 30 minutes at 37°C. The protected products were ethanol precipitated, dried under vacuum, resuspended in 5 µl of formamide dye mix,

denatured, and loaded on denaturing polyacrylamide gels alongside the DNA probe's Maxam and Gilbert sequencing ladder.

2.81 Primer Extensions

Primer extension reactions by a variety of reverse transcriptase enzymes (M-MLV reverse transcriptase, AMV reverse transcriptase) were carried out according to the methods described by the manufacturers. A sequence-specific oligonucleotide primer labelled at the 5' end was ethanol precipitated with RNA, resuspended in 10 µl reverse transcriptase annealing buffer, annealed for 5 minutes at 65°C, cooled slowly to the reaction temperature (37°C-45°C) and incubated for an additional hour. Primer extensions were carried out by the addition of 10 µl reaction buffer and 1 µl of RNase inhibitor, followed by incubation for 1 hour. The products were ethanol precipitated, dried under vacuum, resuspended in 5 µl of formamide dye mix, denatured, and run on denaturing polyacrylamide gels alongside a template sequencing ladder.

2.9 Endonuclease Assays

2.90 Construction of the pGEM-3Zf(+)PCGene Assay Set

The DNA in the vicinity of the 16S-23S operon had been cloned as two fragments. A 7.7 kilobase pair *Hind* III fragment containing both the 16S and the 23S coding regions along with 97 base pairs of 5' flanking and 2800 base pairs of 3' flanking sequence had been cloned into the unique *Hind* III site of pBR322 to generate the clone pBR322(RF3), and an 860 base pair *Sau* 3A fragment containing mostly 5' flanking sequence but only 10 base pairs of the 5' end of the 16S coding sequence was cloned into the *Bam* HI site of pEMBL-18 to generate the clone pEMBL 18(RF2). The clone pGEM-3Zf(+)PCGene, with contiguous 5' flanking sequence and a 16S 5' coding sequence long enough to function in transcript mapping experiments was constructed (Figure 3A). To make the construct, a truncated 754 base pair *Sau* 3A-*Hind* III

FIGURE 3

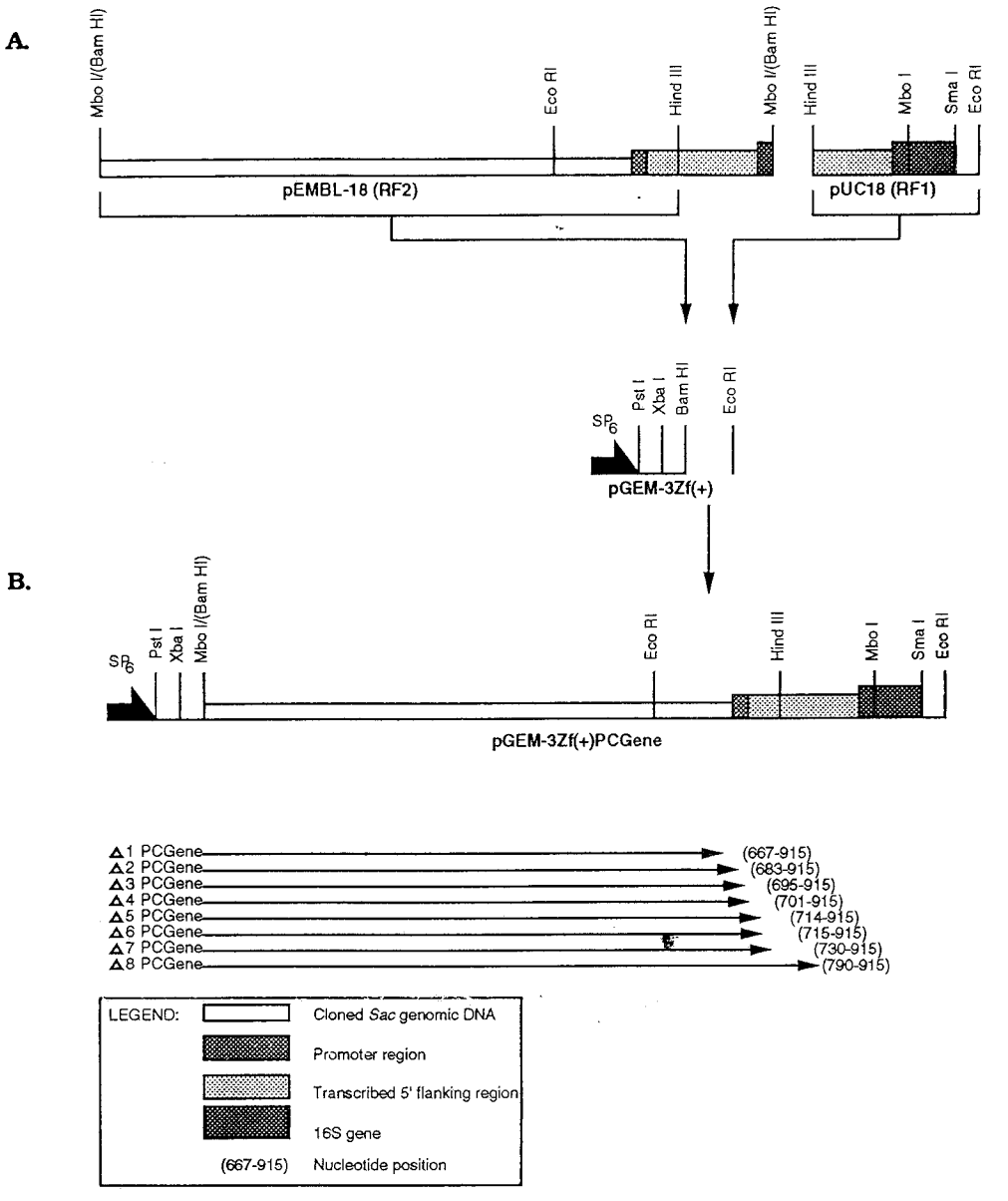


Figure 3: Construction of pGEM-3Zf(+)PCGene.

A. Construct Components. To create a clone contiguous through the transcribed 5' leader and the 5' end of the16S gene, a 750 base pair *Mbo* I-*Hind* III fragment from pEMBL-18(RF2) and a 184 base pair long *Hind* III-*Eco* RI fragment from pUC18(RF1) were co-ligated into a *Bam* HI-*Eco* RI cut pGEM-3Zf(+) vector.

B. Deletions. Exonuclease III-generated deletions were screened and 8 deletions (Δ1PCGene-Δ8PCGene) clustering around the mapped transcriptional start site at position 708 were chosen. *In vitro* transcription from the vector's SP_6 promoter was used to create transcripts resembling the *S. acidocaldarius* rRNA operon transcribed 5' leader.

fragment from the 5' end of pEMBL-18(RF2) was purified. A *Hind* III- *Sma* I fragment from the 5' end of pBR322(RF3) clone had been subcloned into pUC18 to generate the clone pUC18(RF1), and was purified with part of the pUC18 multiple cloning site as a 171 base pair *Hind* III- *Eco* RI fragment. The 754 base pair *Sau* 3A-*Hind* III fragment and the 171 base pair *Hind* III- *Eco* RI fragment were co-ligated into the *Bam* HI and *Eco* RI sites of pGEM-3Zf(+) to generate a 920 base pair fragment containing 849 base pairs of 5' flanking and 71 base pairs of 16S 5' coding sequence. This construct was used in nuclease S1 mapping of the 16S-23S operon promoter region.

Deletions were generated in the 5' flanking region with exonuclease III using *Pst* I as the protecting and *Xba* I as the non-protecting enzymes, and sequenced to verify the extent of the deletions. The resultant clones were labelled pGEM-3Zf(+)Δ1PCGene to pGEM-3Zf(+)Δ8PCGene (Figure 3B).

2.91 Preparation of *S. acidocaldarius* Cell-free Extract

Crude cell-free extract from *S. acidocaldarius* was made according to a modification of the method of Hudelpohl[39]. *S. acidocaldarius* culture grown at 75° C to A_{600}=0.5 was quick-cooled in an ethanol-dry ice bath to room temperature. The cells were pelleted at room temperature, washed twice in distilled water, and resuspended in four packed-cell volumes of extraction buffer (50 mM Tris-HCl (pH 8), 15 mM $MgCl_2$, 1 mM EDTA (pH 8), 1 mM DTT). The cells were lysed by sonication on ice in six 30 second bursts, and the lysate was centrifuged at 4°C, 9500 r.p.m. for 20 minutes. Glycerol was added to a final concentration of 20% (v/v). The extract was flash-frozen in an ethanol-dry ice bath and stored at -70° C.

2.92 *In Vitro* Transcription By SP_6 Polymerase

Run-off transcription from the PCGene construct produced RNA transcripts resembling the 16S-23S transcribed 5' flanking leader, along with the start of the 16S gene. The RNA,

transcribed by Promega SP_6 RNA polymerase and incorporating $\alpha^{32}P$ CTP into the transcript, was driven from vector-contained SP_6 promoters according to the manufacturer's directions, with the following modifications. The RNA produced was not treated with RNase-pure DNase to remove the template DNA, nor was the RNA gel purified to remove premature termination products. Instead, the reaction was extracted once with phenol:chloroform (1:1 v/v), precipitated by the sequential addition of 1/3 volume 7.5M NH_4OAc and 2.5 total volumes of 95% ethanol, centrifuged for 20 minutes, dried under vacuum, and resuspended in 50 µl of DEPC-treated dH_2O. The final concentration was expected to vary between 200-500 ng/µl.

2.93 Processing by S. *acidocaldarius* Cell-free Extract

The standard 20 µl *in vitro* RNA processing reaction was carried out by preincubating 1 µl PCGene RNA substrate, 4 µl 10X reaction buffer, and 7.5 µl dH_2O at 75° C for 10 minutes; 7.5 µl of cell-free extract, preincubated separately for 3 minutes, was then added to the standard reaction mix. The reaction was left to proceed for 15 minutes, and was stopped on ice by the addition of 30 µl ice-cold stop mix (50 mM EDTA (pH 8), 0.5% SDS (w/v)), extracted once with phenol:chloroform (1:1), precipitated by the sequential addition of 1/3 volume 7.5 M NH_4OAc and 2.5 total volumes of 95% ethanol, centrifuged for 20 minutes, and dried under vacuum. The labelled processing precursors, intermediated, and products were separated by size on a denaturing polyacrylamide gel and visualized by autoradiography.

Except for the amount of PCGene RNA substrate, all the parameters of the standard reaction were altered individually to assay for their effect on processing efficiency.

To optimize cleavage, a variety of parameters were altered. First, the reaction was assayed at 55° C, 65° C, 75° C, and 85° C. Then the reaction was assayed at varying salt compositions and concentrations. The 10X test buffers all contained 500 mM Tris·HCl (pH 8) and 100 mM $MgCl_2$. The 10X Buffer "OO" contained no additional salts, 10X buffer "NO" contained 400 mM NaCl, 10X buffer "OK" contained 400 mM KCl, and 10X buffer "NK" contained 200 mM NaCl and 200 mM KCl. Pharmacia's One-4-All buffer, whose composition was not

made available by the manufacturer, was used as the assay standard. All buffers were tested at 0.5X, 1X, 2X, 3X, and 4X final concentrations.

To identify which cleavage product corresponds to the 3' end of the PCGene RNA substrate, the transcripts were produced without incorporation of radioactive nucleotides. After transcription, the RNA was labelled with radioactive ^{32}pCp at its 3' end and only then subjected to the standard reaction. Upon cleavage, only the product corresponding to the 3' end of the substrate should be identifiable.

To identify the 5' end of the transcript, standard reactions were carried out using Δ1- Δ8 PCGene RNA substrates, transcribed from the 5' deletion series of the PCGene clones.

To identify the sequences required for effective cleavage, Δ3 and Δ8 PCGene RNA as control and assay, respectively, were subject to cell-free extract for 1, 2, 4, 8, 16, 32, and 64 minutes.

To assay for the possibility of specific cleavage inhibition by incorporation of NTPαS into the substrate, the SP_6 polymerase reaction of the Δ3PCGene DNA template was carried out with all 16 permutations of NTP/NTPαS nucleotides in an attempt to correlate the nucleotide incorporated with the nature and extent of cleavage inhibition.

To investigate the potential for self-cleavage or for RNA-mediated cleavage, the PCGene transcript was assayed in the absence of cell-free extract. In its place, total cellular RNA (10 μg) and nucleotides (to a final concentration of 0.5mM each NTP and dNTP), or nucleotides alone, were added to the reaction mix.

Finally, to assay for potential RNA mediated cleavage, the cell-free extract was preincubated at 37° C for 15 minutes in the presence of either BSA or RNase A, as control and assay, respectively, preincubated at 75° C for 15 minutes to denature the non-*Sulfolobus* proteins (i.e. BSA or RNase A), and added to the standard reaction. The reaction was stopped as usual, but extracted with phenol:chloroform immediately to remove the RNase A before it renatured and regained activity, destroying the PCGene RNA substrate.

3 RESULTS

3.0 THE 5S rRNA GENE

An attempt to clone the 5S rRNA gene from *S. acidocaldarius* (then still identified as *S. solfataricus*) using the 20mer oligonucleotide oPD1 complementary to the 3' end of the sole published 5S rRNA sequence (attributed to *S. acidocaldarius*) was unsuccessful; the only positively hybridizing sequence in the genome was shown to lack the *Bam* HI restriction site internal to the published 5S sequences. On the assumption that sequence variability within the 5S gene might account for the absence of the conserved *Bam* HI site, the positively hybridizing DNA was cloned into a pUC13 cloning vector as a 1.7 kilobase pair fragment and partially mapped with uniquely cutting restriction endonucleases. When the oligonucleotide oPD1 was used as a primer for enzymatic sequencing, the sequence generated was shown to exhibit no similarity to 5S rRNA downstream from the oligonucleotide binding site.

Since there is a precedent for introns within archaebacterial tRNA genes[37, 38] and 23S rRNA genes[44], as well as for intervening sequences within 5S genes[53], the lack of sequence similarity immediately proximal to the oligonucleotide primer binding site could be indicative of a similar phenomenon in *S. acidocaldarius*. If an intervening sequence was present, sequence similarity to the 5S rRNA should resume at a reasonable distance upstream of the oligonucleotide binding site. The sequencing of exonuclease III-generated deletion clones 900 base pairs upstream of the binding site failed to reveal any additional similarity to the database 5S rRNA sequence. In addition, primer extension analysis using total cellular RNA failed to yield any extension product. Contrary to these results, it had been reported by then that the 5S rRNA sequences from both *S. acidocaldarius* and *S. solfataricus*, obtained by RNA sequencing for the former and DNA sequencing for the latter, were identical[52, 72].

A second attempt to clone the 5S gene using gel-purified 5S rRNA end-labelled with pCp and RNA ligase was successful. Total genomic DNA was digested with a variety of restriction endonucleases, including *Bam* HI. The probe hybridized to a single DNA sequence for all but

the *Bam* HI digests, where two bands were observed (as was expected for a *Bam* HI-containing 5S gene sequence). The probe was then used to identify a subset of phage clones from a constructed λGEM11 genomic library. The 5S gene was subcloned on a 1027 nucleotide *Pst* I-*Xba* I fragment and its sequence was determined. By phylogenetic comparison, the 124 nucleotide 5S rRNA gene was localized between positions 540 and 663. A partial restriction map and the location of the 5S gene are illustrated in Figure 4.

Transcription initiation and termination sites and mature rRNA end sites were located using a combination of nuclease S1 protection assays. To characterize the 5' end of the 5S rRNA gene transcript, the (-) strand of a 193 nucleotide *Sac* I-*Bam* HI fragment was 5' end-labelled at nucleotide position 626, located within the 5S gene's *Bam* HI site (Figure 5). The sole product protected from nuclease S1 digestion by total cellular RNA was 87-89 nucleotides in length, corresponding to a 5' end at nucleotide position 538-540. Longer protection products were not detected. The 3' end of the 5S rRNA gene transcript was located using the (-) strand of a 223 nucleotide *Bam* HI-*Nsi* I fragment, this time labelled at the 3' end at nucleotide position 626. The nuclease S1 resistant product was 38-39 nucleotides in length, corresponding to a 3' transcript end at or near nucleotide position 663-664. Again, longer protection products were not detected.

The mature 5S coding sequence was shown to extend from nucleotide position 540 to position 663. Since no RNA protection products longer than those generated by the mature rRNA were observed using either the 5' or 3' end labelled probes, it may be assumed that the primary transcript of the 5S rRNA gene is probably not processed at either the 5' or the 3' end but is assembled directly into 50S subunits. The G residue at position 540, corresponding to both the site of transcription initiation and the 5S rRNA mature 5' end, is located within a promoter-like box B element that is preceded at an appropriate distance by a box A element (Figure 4).

The sequence of the 5S coding region exhibits 16 nucleotide differences when compared to the published sequences of either *S. acidocaldarius* or *S. solfataricus*, and 18 nucleotide differences when compared to the related *Sulfolobus* strain B12 sequence (Figure 6B). All of

FIGURE 4

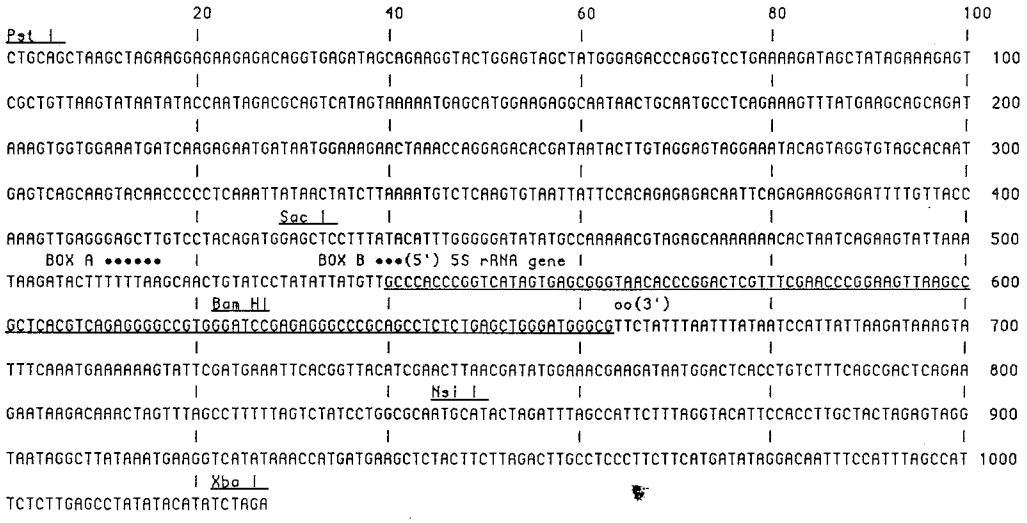


Figure 4: 5S Sequence. The sequence of the 1027 base pair Pst I-Xba I fragment containing the 5S gene, along with 5' and 3' flanking sequence, is shown. All references in the text to nucleotide position are relative to this sequence. The sequence corresponding to the 5S structural gene is underlined. The promoter box A and box B motifs are indicated by solid circles and the transcript termination site mapped by nuclease S1 protection is indicated by open circles. Selected hexanucleotide-specific restriction endonuclease sites as well as the transcript 5' and 3' end sites are indicated.

FIGURE 5

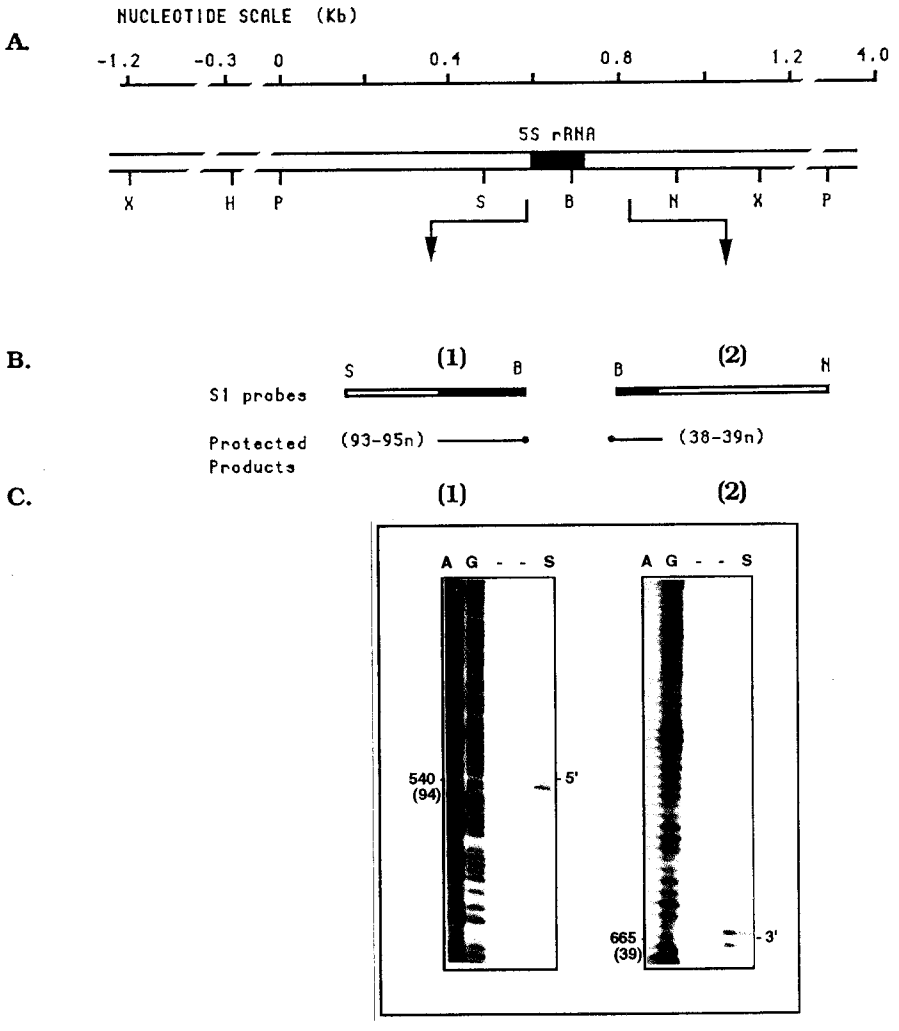


Figure 5: Nuclease S1 Mapping of the 5S Gene.

A. Genomic Restriction Map. The partial genomic restriction map in the vicinity of the 5S gene, indicated in black, is shown. The numbering is relative to the nucleotide sequence of Figure 4. X=*Xba* I: H=*Hind* III: P=*Pst* I: S=*Sac* I: N=*Nsi* I: B=*Bam* HI.

B. Probes Used for Nuclease S1 Mapping. Restriction endonuclease-digested fragments, uniquely labelled at the appropriate end and used for nuclease S1 mapping of (1) the 5' end and (2) the 3' end of the 5S gene, are shown.

C. Nuclease S1 Mapping Results. The nuclease S1 protection products are presented alongside DNA sequence. A = G reaction; G = A+G reaction; S = nuclease S1 reaction. Transcript start and stop sites are indicated to the right of the lanes. 5' = transcript start site; 3' = transcript end site. The position of the sites relative to the sequence in Figure 4 (upper) and the length of the protected product (lower, in brackets) is indicated to the left of the lanes.

Figure 6: Phylogenetic Considerations Surrounding the 5S Gene.

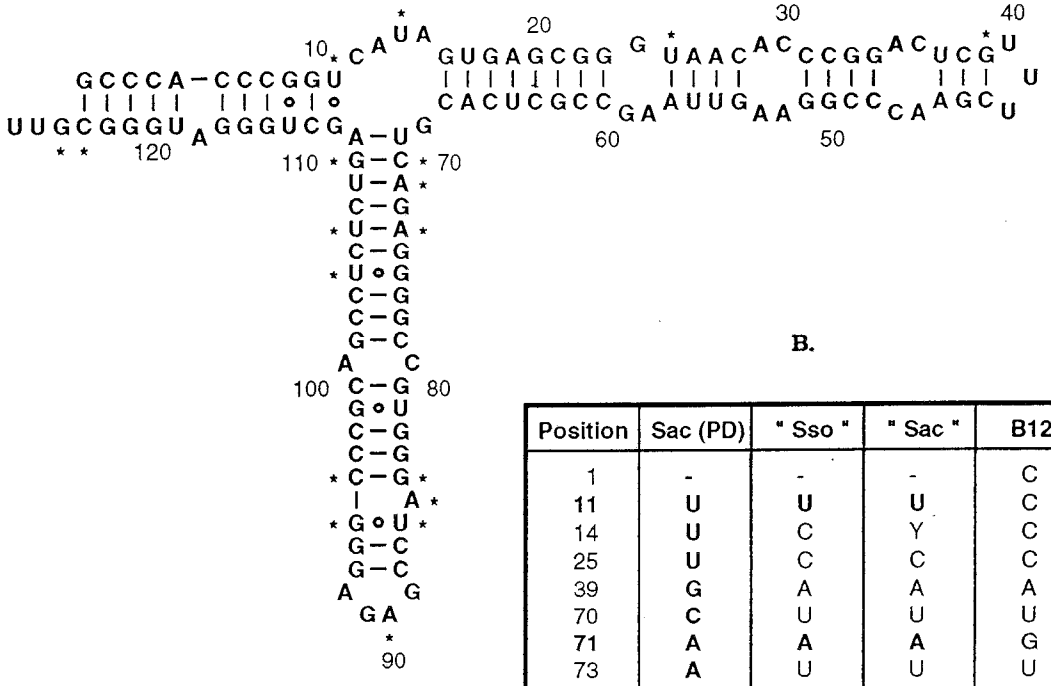
A. Secondary Structure of the Transcript. The predicted secondary structure of the 5S rRNA is based on conserved archaebacterial 5S structure. The numbers are relative to the transcription start site. Solid lines (-)represent standard hydrogen bonded base pairing; open circles (◦) represent non-standard base-pairing; asterisks (*) indicate points of sequence dissimilarity relative to 5S sequences obtained from other *Sulfolobus* species.

B. Nucleotide Concordances. Comparison of nucleotide differences among the three published *Sulfolobus* 5S sequences and the derived sequence. Bold letters indicate nucleotide identity with the derived sequence. Note that at positions 11 and 71, only B12 varies relative to other sequences. "Sso"= *S. solfataricus*; "Sac"= sequence published as *S. acidocaldarius*; B12= *Sulfolobus* species B12; Sac(PD)= *S. acidocaldarius* (Patrick Dennis lab stock).

C. "Sac" 5S Specific Oligonucleotide. The sequence for the oligonucleotide oPD1 was based on complementarity to the 3' end of the published *S. acidocaldarius* ("Sac") sequence. This sequence turned out to be incorrect. The actual sequence (Sac(PD)) contains four nucleotide differences in the 20 nucleotide sequence relative to the published sequence. The 3' terminal T residue of oPD1, while complementary to the A residue in the published "Sac" sequence, fails to complement the corresponding residue in the actual Sac(PD) and accounts for its complete failure to prime in reverse transcription reactions. Open circles indicate mismatches.

FIGURE 6

A.



B.

Position	Sac (PD)	" Sso "	" Sac "	B12
1	-	-	-	C
11	**U**	**U**	**U**	C
14	**U**	C	Y	C
25	**U**	C	C	C
39	**G**	A	A	A
70	**C**	U	U	U
71	**A**	**A**	**A**	G
73	**A**	U	U	U
84	**G**	A	A	A
85	**A**	U	U	U
86	**U**	A	A	A
90	**A**	U	U	U
95	**G**	A	A	A
96	**C**	U	U	U
105	**U**	C	C	C
107	**U**	A	A	A
110	**G**	A	A	A
123	**C**	U	U	**C**
124	**G**	U	U	U

C.

oPD1 3' T G A T T C G A C C C T A C C C A A A A 5'

Sac (PD) 5' T C T G A G C T G G G A T G G G C G T T 3'

these sequences can be folded into the universal secondary structure typical of 5S rRNA molecules. Specifically, in all four thermoacidophilic sequences, the propensity for base pairing is enhanced and helix formation is extended in the context of the universally conserved 5S structure.

The nucleotide substitutions in the sequence obtained, relative to the three other published sequences, do not disrupt the predicted structure. When substitutions occur within helical regions, they are compensatory, and generally act to increase hydrogen bonding across the helix; when substitutions occur at internal helical bulges, they sometimes enhance helix formation. Overall, the predicted structure of the *S. acidocaldarius* folded transcript appears to be more stable than those predicted for the three other known *Sulfolobus* sequences (Figure 6A).

Comparison of the oPD1 20mer oligonucleotide sequence with the experimentally obtained 5S gene sequence indicates the presence of 4 mismatched residues. One of these, a T residue at the 3' end of the oligonucleotide, appears opposite a non-complementary T residue in the 5S gene sequence and explains the complete failure of the oligonucleotide to act as an effective primer in the primer extension assay.

3.1 THE 16S-23S RRNA GENES

The single copy, closely linked 16S and 23S rRNA genes were originally isolated by Norman Pace on a 7.7 Kbp *Hind* III fragment. This fragment contained only one hundred nucleotides of 5' sequence flanking the 16S gene. Alignment with the known *Sulfolobus* strain B12 16S 5' flanking sequence indicated, and nuclease S1 transcript mapping of the *S. acidocaldarius* 16S 5' flanking sequence confirmed, that the transcription initiation site was located upstream of the cloned region. To clone a bigger piece of DNA with more 5' flanking sequence, a 109 nucleotide *Hind* III-*Mbo* I fragment - encompassing the short flanking sequence and nine nucleotides of the 5' end of the 16S gene - was used as a probe for Southern hybridization of restriction enzyme digested genomic DNA to identify and clone an overlapping 857 nucleotide *Sau* 3A1 fragment. The nucleotide sequences of the two fragments in the region

of overlap were identical, indicating that they were derived from the same species, as was to be expected since the lab stock of *S. acidocaldarius* was also obtained from Pace. Using the two clones, 7151 nucleotides of contiguous sequence was determined (Figure 7). Subsequent references to nucleotide position are relative to this sequence.

Several sequencing strategies were used. Initially, the large *Hind* III clone was digested with tetranucleotide-specific restriction endonucleases, and the small fragments (50-350 base pairs in length) were shotgunned into M13. Approximately 13 kilobases of sequence was generated from a total of 67 subclones obtained through screening.

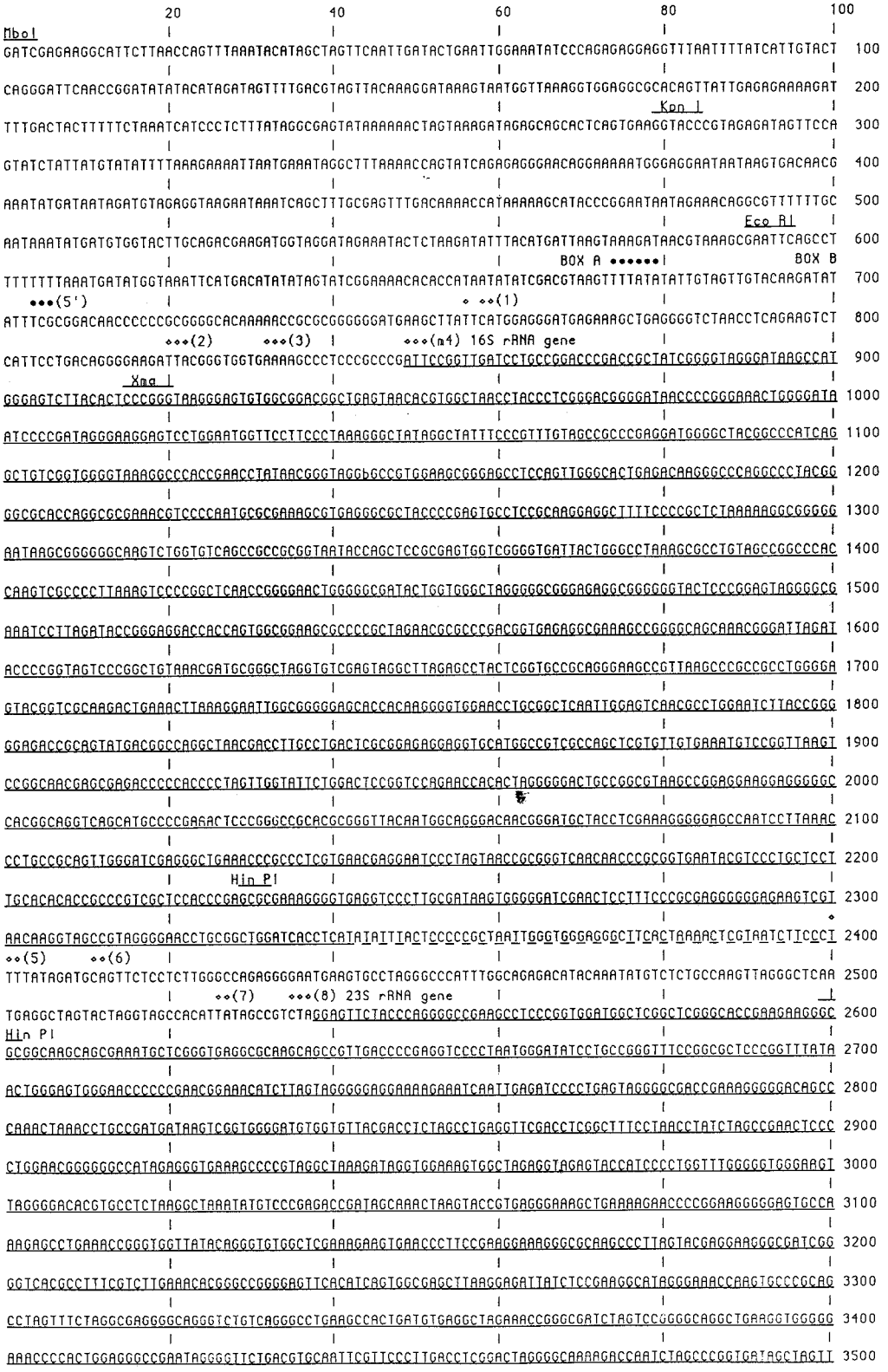
Alignment of the subclones into coherent contig maps was rendered impossible by the limitations of the computer software then available (DNA Inspector^{TM} and MacGene^{TM}). Extended inverted repeats in the structural genes resulted in persistent alignment of noncontiguous DNA. Manual alignment and ordered presentation of subclone sequences to the software resulted in valid contig seeds for some, but not all, of the sequences generated.

Additional contig seeds were generated by sequences of nested and ordered exonuclease III-generated deletions. While these sequences greatly increased confidence in the contigs generated, DNA regions not conducive to the production of stable exonuclease III-generated subclones resulted in two areas of contig uncertainty, both located within the 23S gene. Sequence for these two gaps was generated from small, defined directional subclones, and a single coherent contig was finally obtained. A partial restriction map of this region of the chromosome, and the locations of the 16S and 23S genes, are depicted in Figure 8A.

This region contains a 1493 nucleotide segment that corresponds exactly to the 16S rRNA gene sequence determined previously[88] (position 849-2341), and a 3052 nucleotide segment that corresponds exactly to the 23S rRNA gene sequence of *S. acidocaldarius* (position 2538-5579: Figure 7). A 195 nucleotide intergenic spacer separates the 16S and 23S genes. Nuclease S1 mapping indicates that the first 59-69 nucleotides of the intergenic spacer sequence remain attached to the 3' end of the 16S rRNA even after assembly into 30S ribosome subunits.

Figure 7: 16S-23S Operon Sequence. The sequence generated from two overlapping clones, the 860 base pair *Sau* 3A fragment (position 1-957) and the 6401 nucleotide *Hind* III- *Bgl* II fragment (position 750-7151), is shown. All references to nucleotide position in the text are relative to this sequence. The sequences corresponding to the 16S and the 23S structural genes are underlined (solid line), as is the additional sequence found attached to the mature 16S 3' terminus (dashed line). The promoter box A and box B motifs are indicated by solid circles, and the mapped 3' end site is indicated by open circles. Selected hexanucleotide-specific restriction endonuclease sites, as well as those tetranucleotide-specific restriction endonuclease sites used in nuclease S1 mapping experiments, are indicated. Processing sites within the rRNA transcript, as well as the transcript 5' end sites, are indicated by open squares.

FIGURE 7


20
40
60
80
100
Mbo I
Kpn I
Eco RI
BOX A
BOX B
•••(5')
••(1)
••(2)
•••(3)
•••(4) 16S rRNA gene
Xma I
Hin PI
••(5)
••(6)
••(7)
•••(8) 23S rRNA gene
Hin PI

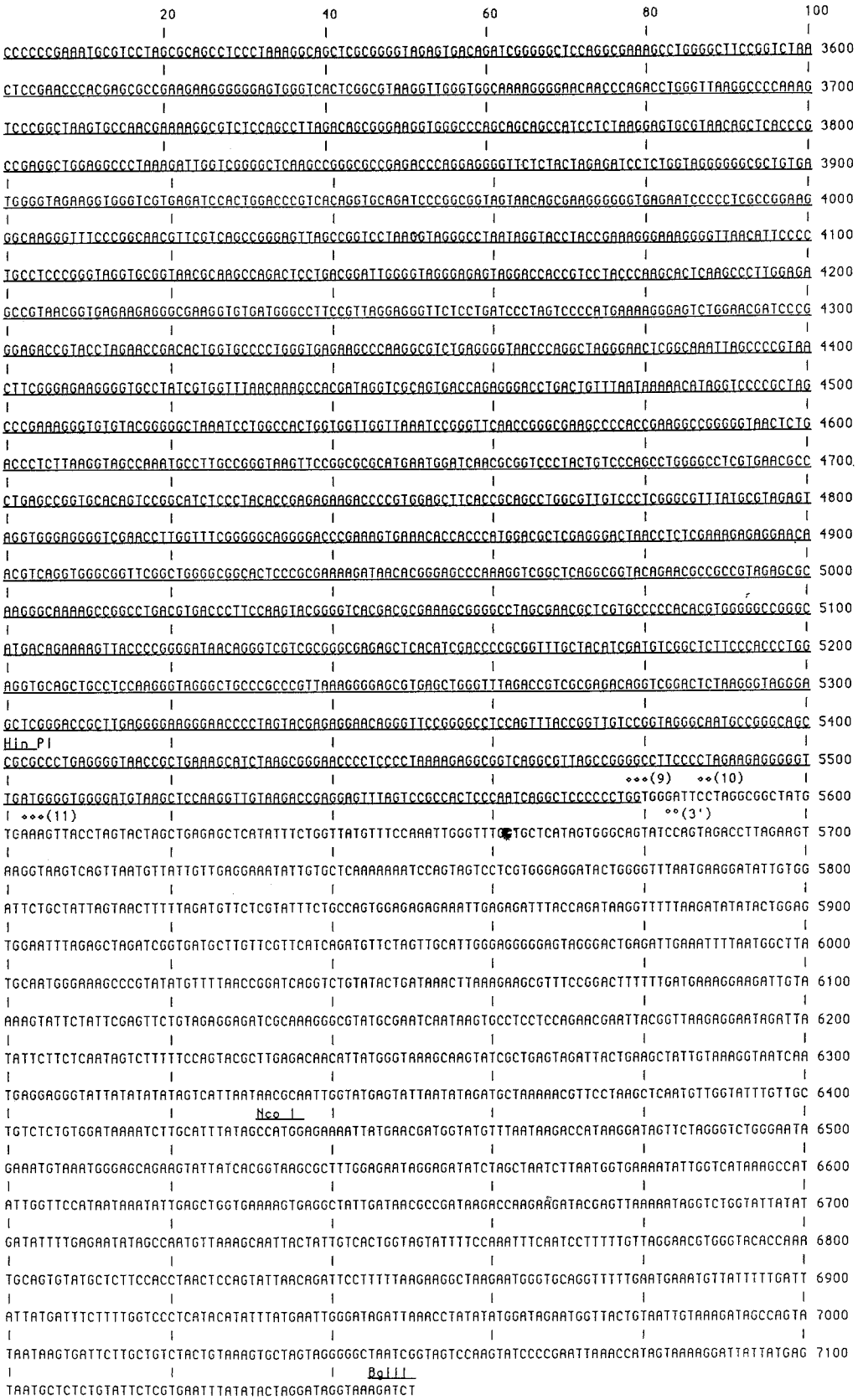

20
40
60
80
100
Hin PI
•••(9)
••(10)
•••(11)
°°(3')
Nco I
BglII

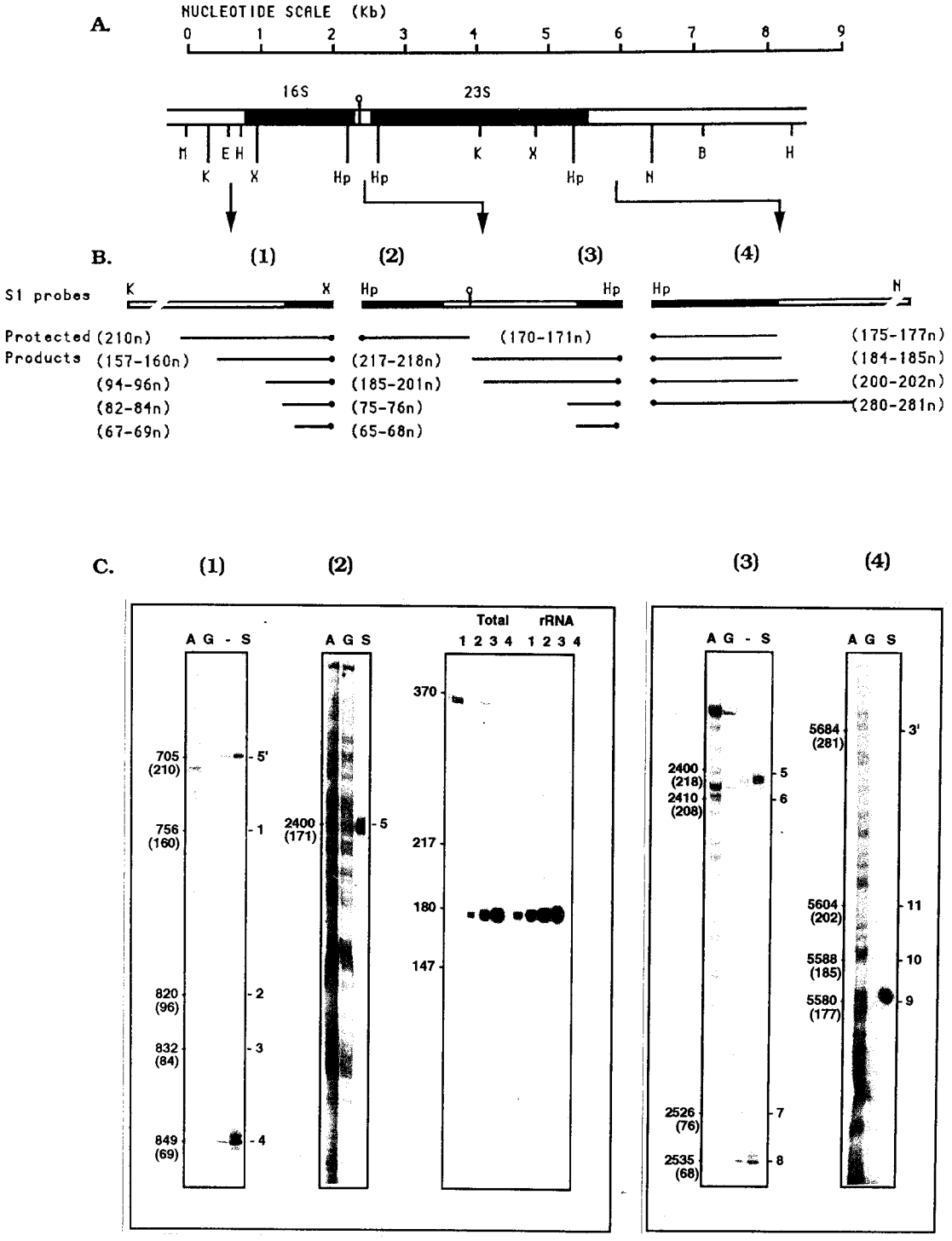
Figure 8: Nuclease S1 Mapping of the 16S-23S rRNA Operon.

A. Genomic Restriction Map. The partial genomic restriction map in the vicinity of the 16S-23S genes, indicated in black, is shown. The numbering is relative to the nucleotide sequence of Figure 7. M=*Mbo* I: K=*Kpn* I: E=*Eco* RI: H=*Hind* III: X=*Xma* I: Hp=*Hin* PI: V=*Eco* RV: N=*Nco* I: B=*Bgl* II.

B. Probes Used for Nuclease S1 Mapping. Restriction endonuclease-digested fragments, uniquely labelled at the appropriate end and used for nuclease S1 mapping of (1) the 16S 5' end, (2) the 16S 3' end, (3) the 23S 5' end and (4) the 23S 3' end, are shown.

C. Nuclease S1 Mapping Results. The nuclease S1 protection products are presented alongside DNA sequence. A = A reaction; G = G+A reaction; S = nuclease S1 reaction. Transcript start and stop sites, as well as mapped processing sites are indicated to the right of the lanes. 5' = transcript start site; 3' = transcript end site; numbered sites are identical to those depicted in Figure 11. The position of the sites relative to the sequence in Figure 7 (upper) and the length of the protected product (lower, in brackets) is indicated to the left of the lanes. The autoradiograms correspond to (1) the 16S 5' end, (2) the 16S 3' end, (3) the 23S 5' end and (4) the 23S 3' end. For (2) the nuclease S1 mapping of the 16S 3' end, both total RNA and ribosome-extracted rRNA were used for protection of the probe DNA at a variety of molar ratios. Lanes 1: 1.0 DNA:0.3 RNA; Lane 2: 1.0 DNA:1.0 RNA; Lane 3: 1.0 DNA:3.0 RNA; Lane 4: 1.0 DNA:9.0 RNA.

FIGURE 8


A.
NUCLEOTIDE SCALE (Kb)
0 1 2 3 4 5 6 7 8 9
16S
23S
M K E H X Hp Hp K X Hp N B H
B.
(1) (2) (3) (4)
S1 probes
K X Hp Hp Hp N
Protected Products
(210n)
(157-160n)
(94-96n)
(82-84n)
(67-69n)
(170-171n)
(217-218n)
(185-201n)
(75-76n)
(65-68n)
(175-177n)
(184-185n)
(200-202n)
(280-281n)
C.
(1) (2) (3) (4)
A G - S
A G S
Total rRNA
1 2 3 4 1 2 3 4
705 (210) 5'
756 (160) 1
820 (96) 2
832 (84) 3
849 (69) 4
2400 (171) 5
370
217
180
147
2400 (218) 5
2410 (208) 6
2526 (76) 7
2535 (68) 8
5684 (281) 3'
5604 (202) 11
5588 (185) 10
5580 (177) 9

As in other thermophilic species of archaebacteria, the 16S-23S intergenic space does not contain a recognizable tRNA gene nor a potential tRNA-like structure, and the region distal to the 23S gene contains neither a 5S nor a 5S-like rRNA sequence. Although there are short open reading frames in the 5' and 3' flanking sequences, none encode polypeptides that exhibit homology to any known protein sequences.

3.10 Structure and Mapping of the 16S-23S Transcript

The 16S and 23S rRNA genes are transcribed as part of a single long primary transcript. Nuclease S1 protection studies have identified putative 5' and 3' transcript end sites at or near positions 706 and 5683, respectively (Figure 11). In addition to the mature 16S and 23S coding regions, the primary rRNA operon transcript contains (i) a 143 nucleotide leader preceding the 16S gene (position 706-849), (ii) a 195 nucleotide spacer between what is recognized phylogenetically as the 3' end of 16S rRNA and the 5' end of 23S rRNA (position 2342-2538), and (iii) a trailer located 3' to the 23S gene (position 5579-5683). Numerous processing intermediates as well as the mature 16S and 23S rRNAs are derived from endonucleolytic and exonucleolytic processing of the 4986 nucleotide primary transcript. To understand these intermediates and appreciate their relationship to the processing pathway and ribosome assembly, it is necessary to visualize processing sites within the contexts of both primary sequence and putative secondary structure.

Located within the non-coding domains of the primary transcript are seven regions of inverted repeat symmetry that have the potential to form hairpins. The two largest of these contain the mature 16S and 23S sequences as part of their apical loops and superficially resemble the RNase III precursor processing stems within the primary rRNA operon transcripts of the eubacterium *E. coli*. In known archaebacterial rRNA operons, these prominent helical structures form a well defined substrate that is recognized by an RNA endonuclease. The substrate consists of two three-nucleotide bulges on opposite sides of the helix separated by precisely four base pairs. Cleavage within these bulges liberates the precursor 16S and 23S

molecules from their respective precursor processing stems within the primary transcript; the precursor sequences are normally removed in subsequent stages of maturation and ribosome subunit assembly. In addition, there are two smaller helices (A and B) in the 5' leader, two helices (E and F) in the spacer, and a single helix (the 3' end stem) in the 3' trailer.

3.11 The 5' Flanking Region of the 16S Gene

Transcription and processing of the primary transcript from the 16S-23S rRNA bicistronic operon is more complex than that of the 5S gene. The putative transcription initiation site and sites generated by endonucleolytic processing of the primary rRNA transcript within the 16S 5' flanking region were identified using a 636 nucleotide *Kpn* I-*Xma* I fragment 5' end labelled on the (-) strand at nucleotide position 919, 70 nucleotides downstream of the 16S gene 5' end. The probe was hybridized to total cellular RNA and the resulting hybrids were digested with nuclease S1. Five different protection products of lengths 67-69, 82-84, 94-96, 157-160 and 210 nucleotides were detected (Figure 8C(1)).

The longest observed product has a 5' end site at nucleotide position 705. The G residue at this position is located within a promoter-like box B sequence and is preceded by an appropriately positioned box A sequence. This fact was verified by reverse nuclease S1 mapping. When the 325 nucleotide *Eco* RI-*Xma* I fragment 3' end-labelled on the (-) strand at nucleotide position 593 and 594 was used as a probe in nuclease S1 protection assays, no protection products were observed. This indicates an absence of long transcription products with initiation points upstream of the *Eco* RI site, and implies that transcription initiation for the full length transcript occurs only downstream of the *Eco* RI site. The presence of putative promoter-like sequence elements upstream of the above-mentioned promoter will be discussed in section 4.00.

The shortest nuclease S1 protection product detected with the 5' end labelled 636 nucleotide *Kpn* I-*Xma* I fragment is 67-69 nucleotides in length and corresponds to protection by mature 16S rRNA with an end site at or near nucleotide position 849. This site is located at

the top of the 16S precursor processing stem. Two possible alternative structures for this region are depicted in Figure 11A.

The three other protection products of length 82-84, 94-96, and 157-160 have 5' end sites at nucleotide positions 832-834, 820-822 and 756-759, and represent novel RNA processing intermediates. The central region of the 16S precursor processing stem is found in all other archaebacterial species to contain a substrate recognized by an RNA endonuclease. In *S. acidocaldarius*, this central region does not conform well to the structure required for endonucleolytic removal of precursor 16S rRNA from the primary transcript. However, a minor but energetically less favorable rearrangement could generate the requisite structure containing the two three-nucleotide loops on opposite sides of the helix, separated by precisely four base pairs (Figure 11B(2)); endonuclease cleavage within the 5' loop between nucleotides 832 and 834 results in the processing intermediate that protects a fragment 82-84 nucleotides in length. Judging from the relative intensity of the bands produced, this does not appear to be a major or stable intermediate in the processing pathway. Of the two remaining 5' end sites, one occurs in an unpaired sequence immediately preceding helix B (position 756-759) and the other occurs within the structural region at the base of the 16S precursor processing stem (position 820-822). The significance of these sites will be discussed in detail in section 4.1.

3.12 The 16S-23S Intergenic Space

Transcript end sites generated within the 16S-23S intergenic spacer region were detected initially using a 370 nucleotide *Hin* PI fragment 3' end-labelled on the (-) strand at nucleotide position 2231, located within the 3' end of the 16S gene. When total RNA was used to protect the probe, only a single heterologous product approximately 171 nucleotides in length was protected from nuclease S1 digestion. This corresponds to a 3' end site near nucleotide 2400. Surprisingly, this is a predominant product and no shorter fragments of about 152 and 112 nucleotides, corresponding to protection by the precursor processing intermediate, expected in archaebacteria, at position 2382 and the 3' end of mature 16S rRNA at nucleotide position

2342, respectively, were apparent (Figure 8C(2)). This unexpected result suggests that the 16S rRNA found in the *S. acidocaldarius* ribosome is not trimmed to the expected 3' maturation site, but instead retains 58-60 nucleotides of what is normally intergenic precursor sequence .

To verify this result and to map the 3' end of the 16S rRNA more precisely, several additional experiments were performed. First, rather than using total cellular RNA in which the longer and therefore more readily hybridizing precursor RNA may be relatively over-abundant, 16S rRNA was isolated from 30S ribosomal subunits that had been purified by sucrose density gradient centrifugation. This rRNA should be relatively free of primary transcript. Furthermore, the amount of RNA employed in the nuclease S1 protection assay was titrated against the DNA probe such that the molar ratio of 16S rRNA:probe DNA ranged form 0.3:1 to 9.0:1. This reduction in RNA concentration should virtually eliminate any masking of a "true" 16S 3' mature end protection signal caused by preferential hybridization of the longer processing intermediate to the DNA probe (and resulting in an inherently more stable RNA:DNA duplex). Figure 8C demonstrates that again, only a single 160-170 nucleotide product is protected. Essentially identical patterns are observed using 30S ribosomal subunit RNA or total cellular RNA. At submolar RNA input concentrations, the product resolves into two distinct bands.

When the *Hin* PI fragment 5' end labelled on the (-) strand at nucleotide position 2601 within the 5' end of the 23S gene was used as probe, four products of length 66-68, 75-76, 190-191 and 198-201 were observed. The shortest of these, as expected, corresponds to protection by the 5' end of mature 23S rRNA at nucleotide position 2538. The 75-76 nucleotide product corresponds to protection by the precursor 23S rRNA liberated by endonucleolytic cleavage within the conserved archaebacterial processing bulge in the 23S precursor processing stem of the primary transcript (position 2526-2527). The final pair of longer protection products exhibits heterogeneous 5' ends in the region between nucleotides 2400-2403 and 2409-2411. The most intense products have 5' ends in the unpaired region at the base of the precursor processing stem (position 2400).

It is important to note that there is virtually no full-length protection of either the 5' or the 3' end-labelled *Hin* PI fragment. This implies that processing within the 16S-23S intergenic

space occurs very quickly after the RNA polymerase has traversed the region, so that unprocessed transcript does not accumulate. These results seem to indicate that the initial event is an endonucleolytic cleavage within the exposed sequence between the 16S processing stem and helix E, which appears to be rapidly trimmed to position 2400, and assembled directly into 30S ribosomal subunits without the expected maturation found in other archaebacterial, eubacterial and eukaryotic species.

After the RNA polymerase has traversed the 23S gene, the 23S precursor processing stem can form by hydrogen bonding of the complementary inverted RNA repeats. The RNA can then be cleaved at the precursor processing site (position 2527). The precursor 23S RNA is subsequently trimmed to generate the mature 5' end of the 23S rRNA at nucleotide position 2538.

3.13 The 3' Flanking Region of the 23S Gene

The 3' endonucleolytic sites were detected using a 930 nucleotide *Hin* PI-*Nco* I fragment 3' end labelled on the (-) DNA strand at nucleotide position 5403. Four products of lengths 175-177, 185-186, 201-203 and 281-282 were protected, corresponding to 3' transcript end sites near nucleotides 5579, 5588, 5604 and 5684, respectively. The shortest product, with an end site near 5579, was also the most abundant, resulting from protection by the 3' end of mature 23S rRNA. The end site near nucleotide 5604 occurs within the bulge region of the precursor processing stem and results from precursor excision by the substrate-specific precursor-processing endonuclease typically found in archaebacteria. The relative amount of this precursor 23S rRNA is rather low, implying that it is rapidly processed to the more abundant product with its 3' end site at nucleotide 5579. The most distal 3' transcript end site at nucleotide 5684 occurs immediately following the 3' end stem and may represent either the site of transcription termination or the most distal processing site detectable. The other product, with a 3' end near nucleotide 5588, occurs within the 23S precursor processing stem at a

position opposite the 5' end of mature 23S rRNA. Its metabolic significance will be discussed in section 4.1.

3.2 Assays for Endonucleolytic Activity

The nuclease S1 mapping results clearly indicate that *S. acidocaldarius* has, superimposed upon the sites and structures of the "classical" archaebacterial rRNA maturation pathway, an absolutely novel set of processing sites which do not fit into any known maturation scheme. Specifically, sites 2, 5, 6, and 9 have no observed equivalents in other species, and the descending side of the 16S precursor processing loop lacks both the staggered partner for site 3 and a proper 3' maturase site (Figure 11). The presence of significant sequence similarity around these errant processing sites (discussed in section 4.10) as well as their strategic location within possible conformational variants of the transcribed noncoding sequences (discussed in section 4.12) warranted further experimentation. If indeed the novel processing events occur *in vivo* in the context of this proposed alternative structure, then it should be possible to design constructs whose transcripts can assume this postulated alternative structure and can be processed *in vitro*.

The constructs designed for this purpose are the pGEM-3Zf(+)PCGene deletion series, discussed in section 2.90. The radioactive *in vitro* transcripts driven from the SP6 promoters of these constructs were subjected to *S. acidocaldarius* cell-free extracts and their processing was monitored by electrophoresis and autoradiography.

3.20 Optimization of *In Vitro* Processing of PCGene RNA

Cleavage of the 264 nucleotide Δ3PCGene RNA transcript in the cell-free extract standard reaction generates four major sets of products with apparent lengths of 33-36 nucleotides, 47-50 nucleotides, 63-67 nucleotides and 106-110 nucleotides (Figure 9). In the optimization of cleavage, a variety of parameters were altered.

Figure 9: *In Vitro* Analysis of rRNA Processing.

A. Standard Reaction.

(1) PCGene Substrate. A schematic representation of the PCGene RNA transcript in the postulated cloverleaf configuration is shown. The start sites for transcripts derived from the deletion clones are indicated as Δ1 to Δ8. Note that Δ1 to Δ4 contain 5' sequence preceding the *in vivo* transcription start site. The clone Δ4PCGene best simulates the actual *in vivo* transcript, with only 5 additional nucleotides at the 5' end. Note also that the clone Δ8PCGene lacks the novel processing site 1. Astrisks indicate processing sites.

(2) Processing Products. The Δ4PCGene RNA transcript was used as a substrate to assay for novel endonucleolytic activity under defined temperature and salt conditions. Four discreet bands, corresponding to the products of three cleavages, accumulate over time. The fragments are: *5'-1*=transcript 5' end to novel processing site 1; *1-2*=novel processing site 1 to novel processing site 2; *2-m4*=novel processing site 2 to 16S 5' maturase site 4; *m4-3'*=16S 5' maturase site 4 to transcript 3' end. The lanes represent various timepoints: 1=1 min.; 2=2 min.;3=4 min.; 4=8 min.; 5=16 min., 6=32 min.

B. Structural Requirements.

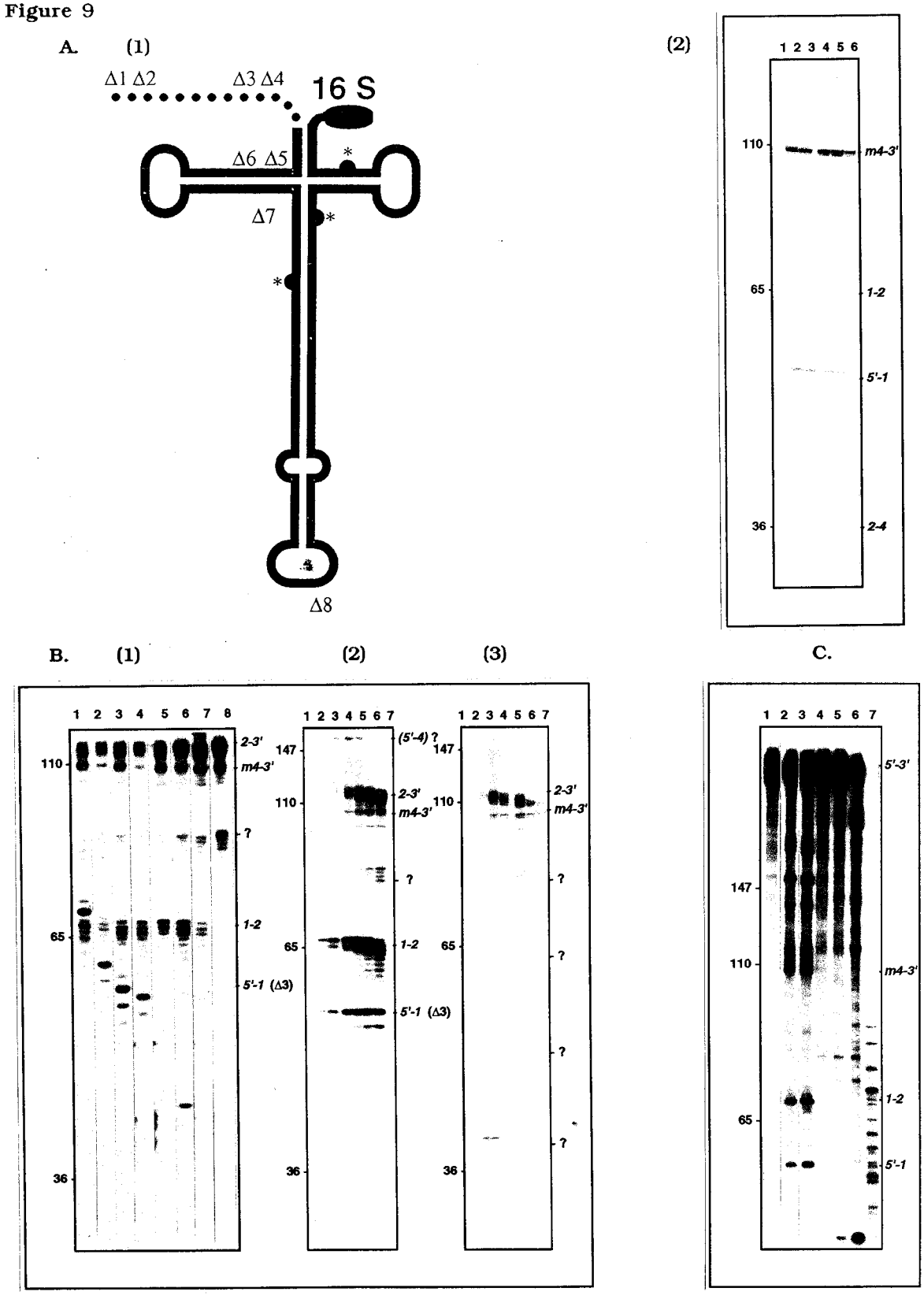
(1) PCGene Deletion Series. Exonuclease III-generated deletions on the PCGene clone were used to probe the structural requirements of the PCGene transcripts and to identify the fragment corresponding to the 5' end of the transcript. The extent of the 5' deletion affects the size of fragment *5'-1* (transcript 5' end to novel processing site 1) in all deletion transcripts. Lane numbers correspond to deletion clone numbers.

(2) Δ3PCGene Deletion. A processing time course at increased salt concentration reveals transient fragments and artefacts in addition to the expected cleavage products. Lane numbers are essentially the same as in A(2).

(3) Δ8PCGene Deletion. A processing time course at increased salt concentration reveals transient fragments and artefacts in addition to the expected maturation product. The detrimental effect on processing of the removal of site 1 is apparent. Lane numbers are essentially the same as in A(2).

C. RNase A Inactivation. The presence of low levels of RNase A was shown to inhibit the novel processing reaction. At high RNase A concentrations, the PCGene RNA substrate was destroyed. Lanes: 1= full length transcript; 2-7 = 10-fold increases in RNAse A, from 4 pg (lane 2) to 400 ng (lane 7).

Figure 9


A.
(1)
Δ1 Δ2
Δ3 Δ4
16 S
Δ6 Δ5
Δ7
Δ8
(2)
1 2 3 4 5 6
110
65
36
m4-3'
1-2
5'-1
2-4
B.
(1)
(2)
(3)
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7
2-3'
m4-3'
?
1-2
5'-1 (Δ3)
(5'-4) ?
147
C.
5'-3'

1. When the reaction was assayed at 55° C, 65° C, 75° C, and 85° C, cleavage was observed over the full temperature range, but length heterogeneity, probably due to exonucleolytic trimming, was found to increase with temperature.

2. When the reaction was assayed at varying salt concentrations to define optimal reaction conditions, the inhibitory effects of monovalent ions became apparent.

a. With the Pharmacia One-4-All buffer, increasing buffer concentration resulted in increased resolution through the retardation of exonucleolytic trimming. At concentrations higher than 2X, the short 33-36 nucleotide product was not discernible on the autoradiogram, but bands of approximately 80, 130 and 150 nucleotides became more prominent. Compare, for example, Figure 9A(2) with Figure 9B(2).

b. Buffer "OO", containing 10 mM $MgCl_2$, but neither NaCl nor KCl, produced results similar to those observed with the Pharmacia buffer. At 3X concentration, the short product was not observable, and the minor bands again became more prominent. At the 4X concentration, the reaction was partially inhibited.

c. Buffers "NO", "OK", and "NK", containing 50 mM NaCl, KCl, or 25 mM both, respectively, in addition to 10 mM $MgCl_2$ resulted in moderate reaction inhibition at the 2X concentration, and complete inhibition at the 3X concentration. All subsequent reactions were carried out in 2X concentration Pharmacia One-4-All buffer as the reaction standard. Even though three artefactual bands (80, 130, and 150 nucleotides long) appear, decreased length heterogeneity of the remaining bands favoured these conditions.

3.21 Characterization of *In Vitro* Processing of PCGene RNA

When the Δ3PCGene RNA transcript was uniquely labelled at its 3' end and subject to the standard reaction, the only major product observed was the 106-110 nucleotide fragment *m4-3'*, generated by processing at site 4, the site of 16S 5' maturation (position 850-915).

When the Δ1PCGene to Δ6PCGene RNAs, transcribed from a 5' deletion series of the PCGene DNA, were used as substrates in standard reactions, three of the four products

generated from each substrate were identical in length. Product *5'-1*, generated by cleavage at site 1 has the apparent length of 70, 59, 57, 56, 43 and 42 nucleotides in the Δ1, Δ2, Δ3, Δ4, Δ5, and Δ6PCGene reactions, respectively, and identifies the 5' end of the transcripts (Figure 9B(1)). Note that in Δ7PCGene, the fragment was too short to visualize, and that Δ8 lacks site 1 completely.

With the 5' terminal fragment *5'-1* identified by size differences among the deletion constructs and the 3' terminal fragment *m4-3'* identified by unique 3' labelling of the RNA before processing, the remaining two fragments *1-2* (position 756-820) generated by processing at sites 1 and 2, and fragments *2-4* (position 820-850) generated by processing at sites 2 and 4 were identified by length alone. That in the case of these latter fragments processing took place at site 2 rather than at site 3 has not been proven experimentally, but is suggested by the model proposed in section 4.131. Further support for this proposition was gathered from the following experiment. When Δ3 and Δ8 PCGene RNA as control and assay, respectively, were incubated with cell-free extract for 1, 2, 4, 8, 16, 32, and 64 minutes, differences in the products generated were used to identify regions required for effective cleavage. In the reaction using Δ8PCGene as the substrate, the only major fragment in common with the Δ3PCGene set was the 106-110 nucleotide *m4-3'* fragment. In addition, the Δ8PCGene reaction exhibited a transient accumulation of two fragments, 40 and 49 nucleotides in apparent length, in the first 8 minutes of reaction. Instead of increasing in length heterogeneity with increasing incubation (as would be expected due to nonspecific exonucleolytic degradation), the fragments disappeared altogether; a concomitant appearance of a single 63 nucleotide band was observed (Figure 9B(3)). Based on the proposed alternative structure (Figure 14), while processing took place as expected at site 4 of Δ8PCGene, processing at site 2 was inhibited in the Δ8 PCGene RNA, presumably because of the disruption of the double-stranded helix around site 2.

When the Δ3 PCGene DNA template was assayed for the possibility of specific cleavage inhibition by incorporation of NTPαS into the substrate, the SP_6 polymerase transcription reaction was carried out with all 16 permutations of NTP/NTPαS nucleotides. Incubation of the transcript with the cell-free extract for times of 0, 4, 16, and 64 minutes was used to track band

accumulation. In nearly every instance where an NTPαS had been incorporated, the cleavage reaction was substantially inhibited, and none of the major species were produced. Of the four major species expected, the 106-110 nucleotide *m4-3'* fragment, corresponding to the 5' portion of the 16S rRNA, appeared only in reactions with RNA transcripts containing ATPαS or UTPαS or both. Additional bands appeared in some of the reactions. A 63-67 nucleotide long fragment, for example, appeared only in the reaction containing CTPαS. In the reaction containing ATPαS, two bands with apparent lengths of 130 and 135 nucleotides, and in the reaction containing UTPαS and CTPαS, two bands with apparent lengths of 135 and 165 nucleotides were observed. Processing appears to be completely inhibited in transcripts incorporating GTPαS since no cleavage products were observed.

When the cell-free extract was preincubated at 37°C for 15 minutes in the presence of either BSA or RNase A as control and assay, respectively, preincubated at 75° C for an additional 15 minutes to denature the non-*Sulfolobus* proteins, and added to the standard reaction, RNase A-mediated inhibition of the cleavage reaction became evident. In the reactions containing up to 40 pg of RNase A, cleavage was unaffected, and the expected bands were observed. In the reaction containing 400 pg of RNase A, cleavage was completely inhibited, and only the full length transcript was visible. At increasing RNase A concentration, substrate degradation became prominent due to the incomplete inactivation of the large amount of RNase A added (Figure 9C). Thus, as evidenced by the RNase A-mediated enzymatic inhibition, an RNA moiety in the cell-free extract is required for the processing of the PCGene RNA transcript.

Finally, when the reaction was assayed in the absence of cell-free extract and the presence of either total cellular RNA (10 μg) and nucleotides (to a final concentration of 0.5mM each NTP and dNTP), or nucleotides alone, no substrate cleavage was observed, even upon overexposure of the autoradiogram. Under the conditions used, neither cellular RNA nor the PCGene transcript RNA is sufficient for processing.

4 DISCUSSION

4.0 SEQUENCE AND STRUCTURE MOTIFS

4.00 Promoter Alignment

Both general and specific sequence motifs are associated with the promoter elements responsible for transcription of stable rRNA genes in thermophilic archaebacteria[24]. These sequences have been termed box A, a general motif centred between 30 and 35 nucleotides upstream of the transcription initiation site, and box B, a stable rRNA specific motif overlapping the transcription start site. It has been suggested that the box A motif may be the functional equivalent of the TATA sequence in eukaryotic RNA polymerase II promoters[25]. Both of these motifs, with their appropriate spacing are found in front of both the *S. acidocaldarius* 5S gene at positions 511-516 and 539-541, respectively (Figure 4), and the 16S-23S operon at positions 674-679 and 704-706, respectively (Figure 7) . By primer extension and nuclease S1 protection, primary transcript 5' end sites have been mapped at or near the G residue at position 600 and 706 in the respective box B motif.

The TTTTTA box A motif of the 5S rRNA gene promoter is located at position 511, 29 nucleotides upstream of the transcription initiation site. The T residue of the TGC box B motif is located one nucleotide upstream of the transcription start site and overlaps the 5S rRNA gene sequence beginning with the 5' terminal G residue at position 540.

The TTTATA box A and TCG box B motifs of the single 16S-23S rRNA operon promoter are located at positions 674, 31 nucleotides upstream of the transcription initiation site and at position 705, respectively; transcription initiation occurs near the G residue within the TCG box B motif at position 706, producing a primary transcript with a 143 nucleotide long 5' leader.

Sequences around the transcript initiation sites of both the 5S and the 16S-23S transcripts are aligned in Figure 10A, along with the well-characterised homologous sequences from the rRNA genes of *Sulfolobus* strain B12, the transcribed genes of the phage particle SSV1,

FIGURE 10

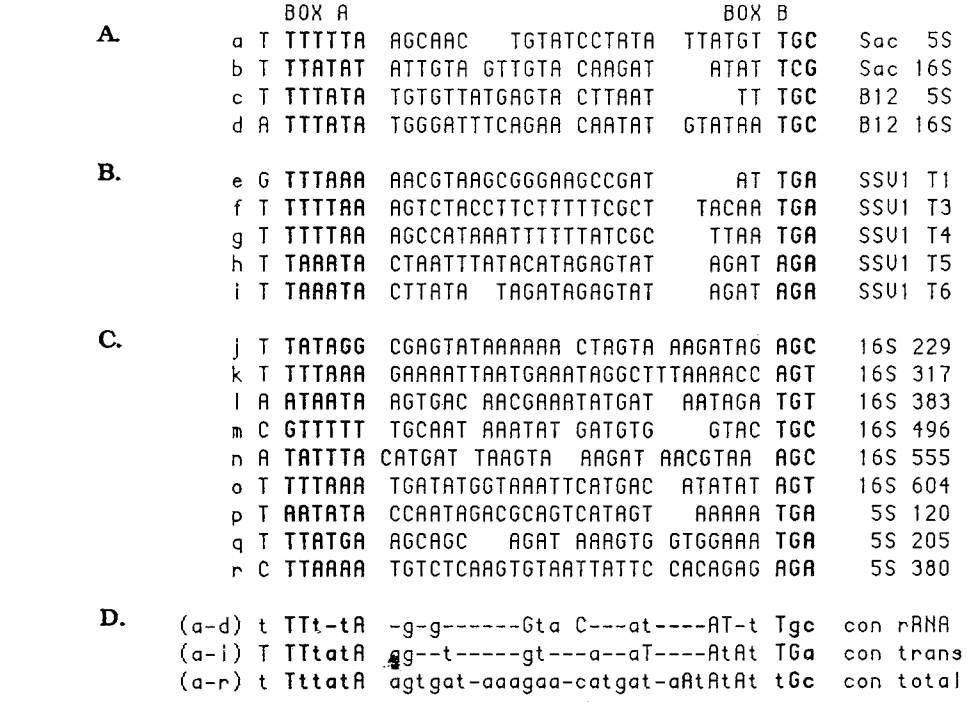


Figure 10: Alignment of Promoters and Putative Promoter-like Elements.

A. *Sulfolobus* rRNA Promoters (1-4). The promoters of the 5S gene and the 16S-23S operon of *S. acidocaldarius* (Sac) are aligned with those of *Sulfolobus* strain B12 (B12) relative to the box A and box B consensus sequences (both indicated by bold letters) according to the constraints described in the text[25].

B. *Sulfolobus* Phage SSV1 Promoters (5-9). The promoters of the mapped Sulfolobus strain B12 transcripts (T1-T6) are aligned relative to the box A consensus sequence and the transcription start site (both indicated in bold letters). The rRNA-specific box B sequence is, by definition, absent.

C. *S. acidocaldarius* Promoter-like Elements (10-18) Sequences upstream of the 5S gene and 16S-23S operon of *S. acidocaldarius* are aligned. The numbers refer to nucleotide positions relative to Figure 4 and 7 for the 5S and 16S-23S proximal sequences, respectively.

D. Consensus Sequences. The consensus sequences derived from *Sulfolobus* rRNA promoters (con rRNA) and *Sulfolobus* phage SSV1 promoters (con trans) are compared to the cumulative consensus sequence derived from all promoter and promoter-like sequences (con total). Residues present in less than 50% of the sequences are indicated by a dash, residues present in more than 50% of the sequences are indicated by lower case letters, and residues present in more than 75% of the sequences are indicated by upper case letters.

and the promoter-like sequences upstream of the *S. acidocaldarius* 16S-23S and 5S genes. Variability in promoter strength and the general evolutionary plasticity of promoter sequences makes such alignments tentative at best. The alignment was performed manually with the following constraints. First, since the promoters exhibit variability in the distance between the box A and box B motifs, gaps had to be introduced into the alignment; the imposed gap number was kept to a minimum to avoid artefactual alignments. Second, the gap position was maintained constant in all sequences on the basis of the underlying supposition that since the polymerase or transcription factor which recognizes the sequence makes contact with the DNA only at defined positions extending over several nucleotides, sequence conservation would cluster around these contact points, and length heterogeneity could only occur outside. Specifically, the alignment was performed relative to the box A and box B elements, which define the outer extremities of the promoters. The distance between these two elements varies between 27 and 33 base pairs. The derived consensus sequence (Figure 10D) shows aligned residues with identity greater than 50% in lower case letters and those with identity greater than 75% in upper case letters. While extending somewhat the conservation around the box A and box B motifs, the consensus shows minimal additional similarity outside these previously defined elements.

Promoters from the *Sulfolobus* phage SSV1 transcripts are shown in Figure 10B[25]. Although the rRNA-specific box B motif is not conserved in mRNA transcripts, the fact that transcription initiates on a G residue, identified by transcript mapping in each case, allows alignment of the promoter region. A consensus derived from both the rRNA genes and the phage genes ("con trans") is fundamentally the same as that derived from the rRNA genes alone ("con RNA"). Again, conservation around the known motifs is extended somewhat, but no new elements are indicated.

An example of a multiple promoter system, with termination and transcriptional reinitiation at the next downstream promoter, has been found in the rRNA operon of at least one thermoacidophile. In *D. mobilis*, multiple promoters were first noted by sequence comparison to the thermoacidophile promoter consensus sequence, but no transcriptional

readthrough could be detected from any but the most proximal promoter[33]. To show that this was due to transcriptional termination rather than an absence of activity, restriction sites downstream of the promoters were used to generate the DNA probes used for detecting transcriptional activity. In virtually all known cases of halophilic archaebacteria, the rRNA operon contains multiple tandem promoters, but not the associated terminators found in *D. mobilis*; these halophilic promoters result in transcriptional readthrough that can be mapped using standard techniques. The promoters are spaced at approximately 125 nucleotide intervals in the 5' flanking region. The single rRNA operon in *H. cutirubrum*, for example, has eight separate promoters, functioning with varying efficiency, spread over one kilobase of sequence[28, 29]. Such systems, whether they involve termination and reinitiation or merely read-through, are postulated to increase polymerase loading and maximize transcription from the single copy rRNA operon during exponential growth.

In *S. acidocaldarius*, nuclease S1 protection assays indicate that transcription appears to initiate from single sites for both the 16S-23S and the 5S transcripts. Closer examination of the upstream region of the *S. acidocaldarius* 16S-23S operon, however, reveals six sequences, at positions 229, 317, 383, 496, 555, and 604, exhibiting similarity to the promoter consensus sequence. In the absence of detectable transcriptional read-through in mapping experiments, their significance is unclear. If transcription initiates at these sequences, but terminates before reaching the next putative promoter, nuclease S1 mapping analysis would not reveal the presence of these transcripts. Alternatively, these regions could be involved in polymerase association without transcription initiation, serving merely to increasing polymerase concentration in the vicinity of the operon.

Intergenic promoters with low transcription initiation rates have been reported in halophilic archaebacteria, and are assumed to rectify stoichiometric imbalance caused by premature transcriptional termination[31, 32]. The presence of a cryptic promoter-like sequence in the intergenic spacer region between the 16S and 23S genes of *S. acidocaldarius* is enigmatic. The putative sequence's box A, at position 2401-2406, is located at the foot of the 16S pecursor processing stem, with the putative transcription start site at a G residue at position 2435

within the stem of helix E (Figure 16). Transcripts initiated from this promoter would be detectable even if present at only 10% of the normal level; nuclease S1 mapping, however, does not reveal the presence of any transcripts originating from this position.

4.01 Terminators

Termination of the *S. acidocaldarius* 5S transcript, like initiation, is coincidental with its mature end, with no RNA processing required for the formation of a mature species. Termination takes place at positions 666-665 within the first pair of T residues in a long T-rich tract, resulting in 3' end heterogeneity over these two nucleotides. Whether this is the result of non-specific termination or of exonucleolytic activity, or whether it is an artefact of nuclease S1 mapping cannot be rigorously determined by the procedure used.

Without a biochemical basis, there is no rational framework in which to discuss, other than by analogy, the nature of archaebacterial termination. Conceptually, the termination event for the 5S transcript is most similar to the *rho* factor independent termination found in the eubacteria[35]. Presumably, the 5S transcript assumes its secondary conformation during transcription; the helix resulting from complementary interaction of the 5' and 3' ends of the transcript, along with the downstream T residues, may generate the requisite termination signal.

A different motif is found around the most distal 3' end of the 16S-23S transcript, which maps to around position 5584, 105 nucleotides downstream of the 23S 3' mature end. This position corresponds to a pair of C residues 1 nucleotide downstream of a postulated helical structure. Unlike the case of the 5S transcript, there is no poly-T tract in the immediate vicinity of the site. The site is unusual because archaebacterial rRNA transcript termination in the presence of upstream or downstream helical structures, or termination in the absence of local helical structures, is virtually always associated with poly-T tracts which may or may not be interspersed with C residues[29, 48, 49].

Since the level of detectable signal at this site is extremely low, any longer transcripts produced by antitermination would be undetectable. It cannot be ruled out, therefore, that the site may be generated by an endonucleolytic processing event rather than by termination.

4.02 Helical Conservation within the Universal Structure

Sequencing of rRNA operons from a variety of archaebacteria has revealed the presence of sequence as well as structural motifs conserved generally in most species, as well as specifically in certain lineages. Except for the promoter-containing helix D, present only in the extreme halophiles (Figure 1B), no function has been ascribed to any of the helices. Figure 11 shows the proposed secondary structural elements of the *S. acidocaldarius* 16S-23S rRNA primary transcript. The structures were predicted on the basis of sequence complementarity within the primary transcript according to the motifs defined by Kjems and Garrett[5, 74]. The helix nomenclature is also consistent with that of Kjems and Garrett.

Helix A, found in all archaebacteria and defined in the absence of sequence conservation by its position distal to the 5' end of the 16S gene, is formed from a pair of imperfect inverted repeat sequences at positions 714-724 and 734-745, 8 and 28 nucleotides downstream of the transcription initiation site. The sequences may generate an 11 base pair long stem structure disrupted by a single nucleotide bulge, with a 9 nucleotide apical loop.

Helix B, also found to be universally distributed within the archaebacteria, is defined by sequence at the base of the helix as well as by its position proximal to the 5' end of the 16S gene. It is formed from a pair of imperfect inverted repeat sequences at positions 763-781 and 789-807, 57 and 83 nucleotides downstream of the initiation site. The sequences may generate a 16 base pair long stem structure topped by an 8 nucleotide loop. An internal loop, observed to be a common helix B feature in halophilic and thermoacidophilic archaebacteria, is also present. The sequences surrounding the helix, 5'-AUGGA-helix-UGA-3', are identical to the partially conserved sequences 5'-RUGNN-helix-YGA-3', with the 5' A residue at position 763 and the 3' U residue at position 807 forming the terminal base pair of the *S. acidocaldarius* helix B.

Figure 11: 16S-23S Primary Transcript-Universal Structure. The secondary structure of the 16S-23S primary transcript is depicted as the *Sulfolobus*-specific variant of the universal conformation. Bold letters indicate sequences within the phylogenetically defined boundaries of the structural genes. Nucleotides are numbered relative to the sequence of Figure 7. Helices are named according to the nomenclature of Garrett[5]. The boxed numbers refer to the following processing sites: 5' = transcript start site; 1 = novel processing site 1; 2 = novel processing site 2; 3 = 16S 5' precursor processing site 3; 4 = 16S 5' maturase site 4; 5 = site 5; 6 = novel processing site 6; 7 = 23S 5' precursor processing site 7; 8 = 23S 5' maturase site 8; 9 = 23S 3' maturase site 9; 10 = novel processing site 10; 11 = 23S 3' precursor processing site; 3' = transcript end site (and putative novel processing site.

A. Conformational Variants Near the 16S 3' End. The RNA in the vicinity of the 16S 3' end is capable of folding into one of two conformations.

(1) Tangential Helix. The tangential helix, a common archaebacterial motif near the 16S structural gene arising from an inverted repeat within the intergenic 16S precursor processing stem 3' sequence, would form only at the cost of thermodynamic stability to the entire 16S precursor processing stem. Energetically less stable, its formation would be favoured in the absence of the 5' complementary sequence, a situation which exists upon the novel maturation of the 16S 5' end, as discussed in section 4.12.

(2) Contiguous Processing Stem. One half of the above-mentioned inverted repeat exists in the context of the mutually exclusive but thermodynamically favourable extended precursor processing stem. The other half exists as single-stranded RNA, with no complementarity to the 5' processing stem sequenceand no possibility of tangential helix formation in the 5' processing stem sequence.

B. Conformational Variants at the 16S Precursor Processing Site. The RNA in the vicinity of the staggered precursor processing sites, thought to be cleaved by an RNase III-like endonuclease, is capable of undergoing a minor shift in base paired interactions, giving rise to two similar forms.

(1) Energetically Favourable Helix. The base pair interactions are maximised, giving rise to an extended helix disrupted by non-symmetrical staggered bulges separated by four base pairs.

(2) Precursor Processing Recognition Motif. Here, the base pair interactions are compromised, giving rise to localized helical disruptions. By converting the two-nucleotide bulge into a three-nucleotide bulge, however, the 5' processing stem sequence is forced into the recognised conformation, allowing processing to proceed.

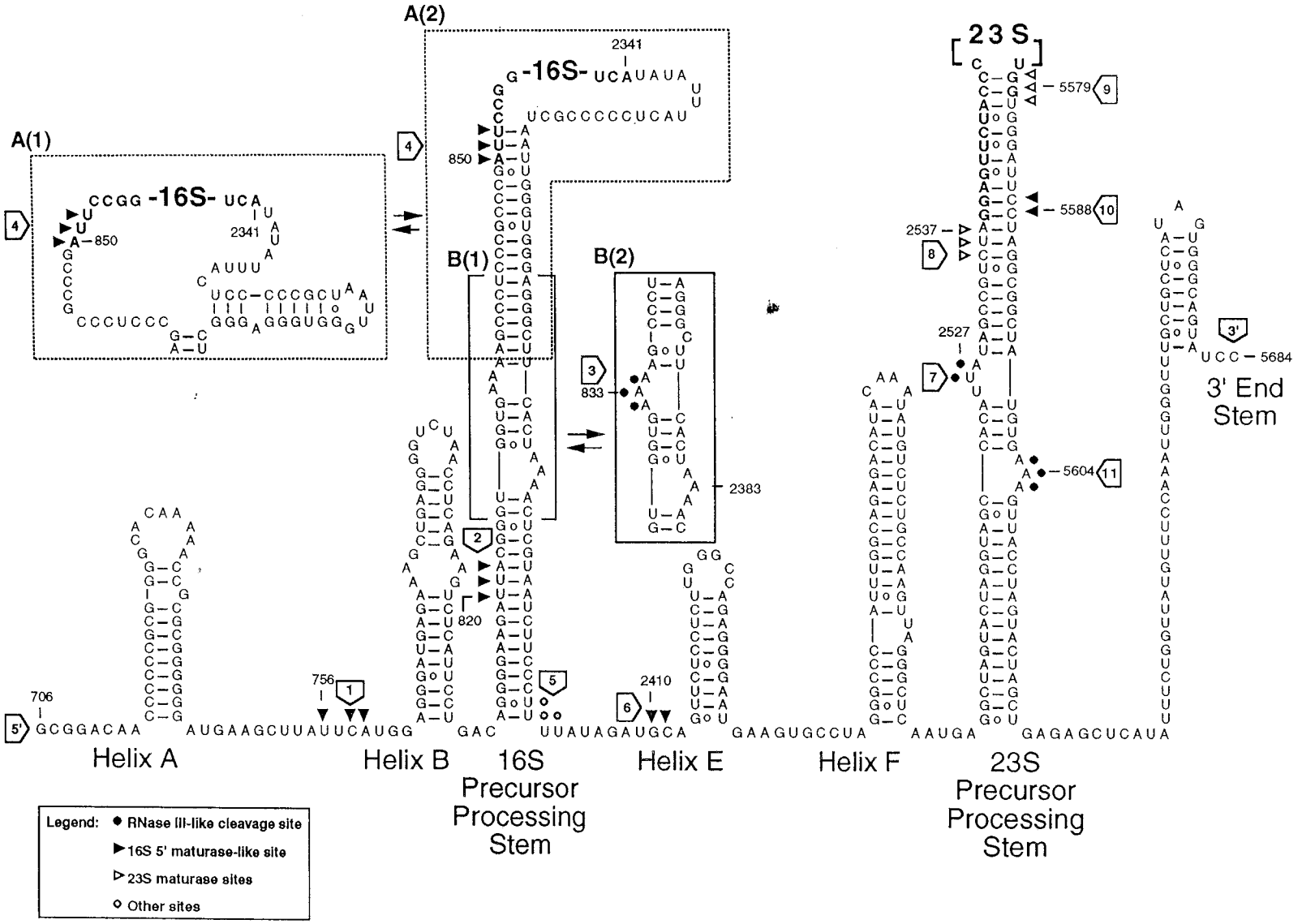

A(1)
A(2)
B(1)
B(2)
-16S-
23 S
3' End Stem
Helix A
Helix B
16S Precursor Processing Stem
Helix E
Helix F
23S Precursor Processing Stem
Legend:
RNase III-like cleavage site
16S 5' maturase-like site
23S maturase sites
Other sites

FIGURE 11

Helix E, absent from the intergenic spacer region of the thermoacidophiles exhibiting a bifurcated processing stem, is present in *S. acidocaldarius*, and is formed from a pair of perfect inverted repeat sequences at positions 2413-2422 and 2429-2438, 1717 and 1723 nucleotides downstream of the initiation site. The sequences may generate an undisrupted 10 base pair stem structure with an 6 nucleotide apical loop. The bordering sequences, AUGCA and GAAG, are again identical to the partially conserved helix E-bordering sequences RUGCA and GAAG.

Helix F, defined as the helix which directly precedes the 23S precursor processing stem, is formed from a pair of long imperfect inverted repeat sequences at positions 2449-2470 and 2475-2498, 1743 and 1769 nucleotides downstream of the initiation site. The bordering sequences, UA and AAUGA, are identical to neither, but are similar to both of the partially conserved helix F flanking sequences GA and RUGA.

The proposed 3' transcript end stem located downstream of the 23S gene and immediately preceding the transcript 3' end site is not one of the structures known to be universally conserved in either the archaebacteria generally or the thermoacidophiles in particular, but appears to be specific only to *S. acidocaldarius*. It is formed from a pair perfect inverted repeat sequences at positions 5661-5669 and 5673-5681, 5955 and 5967 nucleotides downstream of the transcription initiation site.

The 16S precursor processing stem is located between helix B and helix E. It is formed by a pair of imperfect inverted repeat sequences; the 5' sequence is located at position 811-858, 105 nucleotides downstream of the transcription start site, and the 3' sequence is located at position 2341-2401, 1635 nucleotides downstream of the transcription start site. Several conformational variants of the helix can be postulated. The simplest and thermodynamically the most stable form is a long 36 base pair linear stem disrupted by two staggered bulges separated by four base pairs, located on opposite sides of the helix. The 5' sequence contains a bulge of 2 nucleotides at position 832-833, and the 3' sequence contains a 3 nucleotide bulge at position 2382-2384 (Figure 11B(1)). This structure deviates from the conserved archaebacterial rRNA precursor processing motif consisting of two 3 nucleotide bulges staggered by four base pairs. The structure could be forced into the accepted processing motif conformation only at

the cost of thermodynamic stability; expanding the 5' bulge to 3 nucleotides entails local helical disruption, and the formation of an additional mononucleotide bulge in the 3' sequence at position 2375 (Figure 11B(2)).

Because of the inverted repeats at positions 2350-2359 and 2365-2374 within the 3' sequence of the precursor processing stem, the potential exists for the formation of a tangential 9 base pair stem-loop near the apex of the processing stem (Figure 11A(1)). Similar as well as more complex stem-loop structures have been observed at the 16S rRNA precursor processing stem apex in some, but not all, other archaebacterial species. As with the conserved rRNA transcript-encoded helices A, B, E, and F, no function has been ascribed to these structures.

In comparison to the structural variability of the 16S processing stem, the 23S processing stem is much simpler, conforming to the conserved archaebacterial motif. It is formed by a pair of imperfect inverted repeat sequences extending into the coding region of the 23S gene; the 5' sequence is located at position 2504-2548, 1798 nucleotides downstream of the transcription start site, and the 3' sequence is located at position 5478-5522, 4772 nucleotides downstream of the transcription start site. The 5' sequence contains a 3 nucleotide bulge at position 2525-2527, and the 3' sequence contains a 3 nucleotide bulge at position 5603-5605; these bulges are separated by four base pairs of double-stranded RNA and conform to the conserved archaebacterial precursor processing motif.

4.1 Nuclease S1 Sites within the Universal Structure

Based on our understanding of archaebacterial rRNA processing, the minimal requirement for maturation of each of the rRNA species consists of paired endonucleolytic cleavage events in the staggered bulges of the processing stems, liberating large precursor molecules from the polycistronic transcript, followed by an endonucleolytic cleavage at each of the four mature terminus sites, generating the mature species[49]. In *S. acidocaldarius*, these expectations are only partially fulfilled.

The potential of the transcribed leader sequence, in association with the transcribed intergenic sequence, to form the double-stranded 16S precursor stem with the appropriate staggered bulges at positions 832-834 and 2382-2384 required for processing gives rise to an expectation of both cleavage events. Mapping results, however, only indicate cleavage 5' to the 16S sequence within the AAA bulge (formed in the less stable conformation, Figure 11B(2)) at position 832-834; the extremely low band intensity on the autoradiogram suggests that this precursor is either quickly processed to maturity, or that the enzyme recognizes and cleaves at this site only at relatively low levels. No 3' cleavage is observed within the AAA bulge at position 2382-2384, even upon over-exposure of the autoradiograms. This apparent lack of activity in the presence of both the structural RNA substrate and the endonuclease is unprecedented in known archaebacterial rRNA operons, and defies explanation in the context of the universally conserved primary transcript structure alone.

Nuclease S1 mapping indicates both precursor processing events occur as expected in the 23S processing stem. The cleavage 5' to the 23S sequence occurs within the UUA bulge at position 2526-2527, and the cleavage 3' to the 23S sequence occurs within the AAA bulge at position 5603-5605.

Four additional cleavage events are expected in the generation of the mature species of rRNA. As with precursor processing, maturation of the 16S rRNA does not conform to expectations. Nuclease S1 mapping using total cellular RNA indicates that the mature 5' terminus is generated as expected at position 849-852, downstream of the precursor processing stem, and conforms to phylogenetic expectations for 16S 5' ends. The mature 16S 3' terminus, expected at position 2341, is not detectable by nuclease S1 mapping using either of two different probe DNA fragments, even upon DNA:RNA optimization and extensive over-exposure. In experiments using probe DNA labelled at the 3' end at position 2235 within the 16S coding sequence, the first and only detectable signal, which indicates the shortest - and therefore presumably mature - 16S species, occurs as a series of bands around position 2400. Relative to the universal structure, this corresponds to a poly-U tract on the 3' side of the 16S precursor processing stem (site 5). This motif - a poly-U tract immediately following a helical stem - is

costructural with known eubacterial terminators, and is the same motif found at the transcriptional terminus of the *S. acidocaldarius* 5S rRNA gene. Given that, although intergenic promoters have been demonstrated in halophilic archaebacteria, there are no known examples of intergenic terminators, this result is unusual.

The experiment was repeated to remove the possibility of a methodological artefact. As well as titrating against freshly prepared total cellular RNA, rRNA extracted from 30S ribosomal subunits was titrated against the probe DNA. The RNA bound in ribosomal subunits should be in a mature and therefore functional form, and should remove possible mapping artefacts caused by an overabundance and greater stability of the processing intermediates. The products were identical for both the total RNA and ribosome-extracted RNA, and resolved at submolar concentrations of probe DNA into two closely spaced bands. When nuclease S1 protection experiments were performed using probe DNA labelled at position 2601 within the 23S RNA, the longest protected products detected were two extended series of bands corresponding to positions 2400-2405 (processing site 5) and positions 2409-2411 (processing site 6). Evidently, the nature of transcript processing at the 3' end of the 16S gene is complex.

For the 23S rRNA species, nuclease S1 mapping indicates both maturation events occur as expected on the 23S precursor processing stem. The generation of the 5' mature terminus occurs within the processing stem itself at position 2525-2527, and generates the terminus with a phylogenetically conserved 5' mature end. Since the nature of the enzymatic activity which generates the 23S 5' terminus is unknown, it is unclear whether the observed length heterogeneity is the expected result of exonucleolytic degradation from the precursor processing site 8 nucleotides upstream, or of non-specific degradation of the mature endonucleolytically-produced 5' terminus. The maturation of the 3' terminus of the 23S rRNA also occurs within the processing stem at position 5578-5580, giving rise to a phylogenetically conserved 3' end.

In addition to the signals generated by precursor processing and rRNA maturation, three other novel signals generated by endonucleolytic cleavage were detected in the nuclease S1 mapping of the primary transcript. Their locations, at positions 756-759 (novel processing site

1), 820-822 (novel processing site 2), and 5587-5588 (novel processing site 10), correspond to no known archaebacterial processing sites.

4.10 Is the Novel Endonuclease a Renegade 16S 5' Maturase?

In order to define any common sequence motifs that might help elucidate the nature of the novel processing enzyme at these unusual processing sites, RNA sequences around the cleavage sites were aligned relative to the positions of signals generated by nuclease S1 mapping. Surprisingly, sequence similarity was detected not only among the three novel processing sites 1, 2, and 10, but also with the 16S 5' maturase site 4, site 6 (the apparent 16S 3' processing intermediate site), and site 11 (the most distally identifiable 3' transcript end). Sequences around these sites are aligned in Figure 12A(1). The consensus sequence generated, GAUUCC, matches perfectly the sequence around the site of 16S 5' maturase site 4, as well as around site 9.

Since the sequence at the 16S 5' terminus is phylogenetically conserved, and since the 5' maturase may itself be a well conserved enzyme, it is a strong starting supposition that, given the extent of sequence similarity at the processing sites, the endonucleolytic activities observed at novel processing sites 1, 2, and 10, and at 16S 3' processing intermediate site 6 might be due to auxiliary processing by the 16S 5' maturase. The presence of novel sites 1 and 2 within 100 nucleotides of the 16S 5' maturase site 4 hints at a possible mechanistic interdependence of these three cleavage events, and offers the opportunity for rational modelling to resolve the events leading to maturation.

The sequence at the unusual 16S 3' processing intermediate site 6 only exhibits 4/6 nucleotide identity with the GAUUCC consensus sequence. Interestingly, the two mismatches correspond exactly to the highly conserved residues at the base of the archaebacterial helix E, suggesting an undefined functional multiplicity for the sequence. If cleavage at processing site 6 could be shown to dependent upon precedent activity at sites 1, 2, and 4, within the 5'

Figure 12: The Novel Endonuclease Site

A. Sequence Alignment.

(1) Novel Processing and 16S 5' Maturase Sites. The sequences around mapped processing sites are aligned. Bold letters indicate the hexanucleotide sequence thought to be involved in recognition and cleavage. Numbers in brackets indicate nucleotide position of the mapped cleavage event. Fractions following the nucleotide sequences indicate matches relative to the respective consensus sequence. (Numerator = number of matches; denominator = number of residues in the consensus sequence.)

(2) *Sulfolobus* tRNA Intron Processing Sites. The sequences around known tRNA intron processing sites are aligned. In all cases, cleavage occurs between the third and fourth residues of the highlighted sequence. Generally, sequence conservation is greater at the intron 3' processing site because the hexanucleotide at the 5' site is invariably constrained by the tRNA anticodon sequence.

(3) Consensus Sequences. Consensus sequences are derived from the alignment of novel processing sites and 16S 5' maturase site (con novel), of tRNA intron processing sites (con tRNA) and of all processing sites (con total). Residues present in less than 50% of the sequences are indicated by a dash, residues present in more than 50% of the sequences are indicated by lower case letters, and residues present in more than 75% of the sequences are indicated by upper case letters.

B. Structural Motif. The sequences around mapped processing sites are shown in the context of the secondary structures thought to be required for processing. Each motif is shown in both possible resonant modes. The constraint of a structural consensus imposes a weak sequence consensus (5'-uGaRG-3') in the complementary strand.

(1) A-G Double Purine Bulge. In all the structures the third base pair corresponds to the A residue in the 5'-GAUUCC-3' consensus sequence across from the second G residue in the 5'-uGaRG-3' consensus sequence.

(2) Single A Residue Bulge. In this conformation, the A residue in the 5'-GAUUCC-3' consensus sequence is unpaired, with the G residue in the 5'-uGaRG-3' consensus sequence now hydrogen-bonding to the first U rsidue of the 5'-GAUUCC-3' consensus sequence.

FIGURE 12

A (1) Novel Processing and 16S 5' Maturase:

			con novel	con tRNA	con total
Novel Processing Site	1 (756)	GAAGCU **UAUUCA** UGGAGG	4/6	2/6	2/4
Novel Processing Site	2 (820)	GGGGAA **GAUUAC** GGGUGG	5/6	3/6	3/4
16S 5' Maturase Site	4 (850)	CCGCCC **GAUUCC** GGUUGA	6/6	4/6	4/4
Novel Processing Site	6 (2411)	UUUAUA **GAUGCA** GUUCUC	4/6	3/6	3/4
Novel Processing Site	9 (5587)	UGGUGG **GAUUCC** UAGGCG	6/6	4/6	4/4
Transcript 3'End Site	11 (5692)	UGGGCA **GUAUCC** AGUAGA	4/6	3/6	3/4

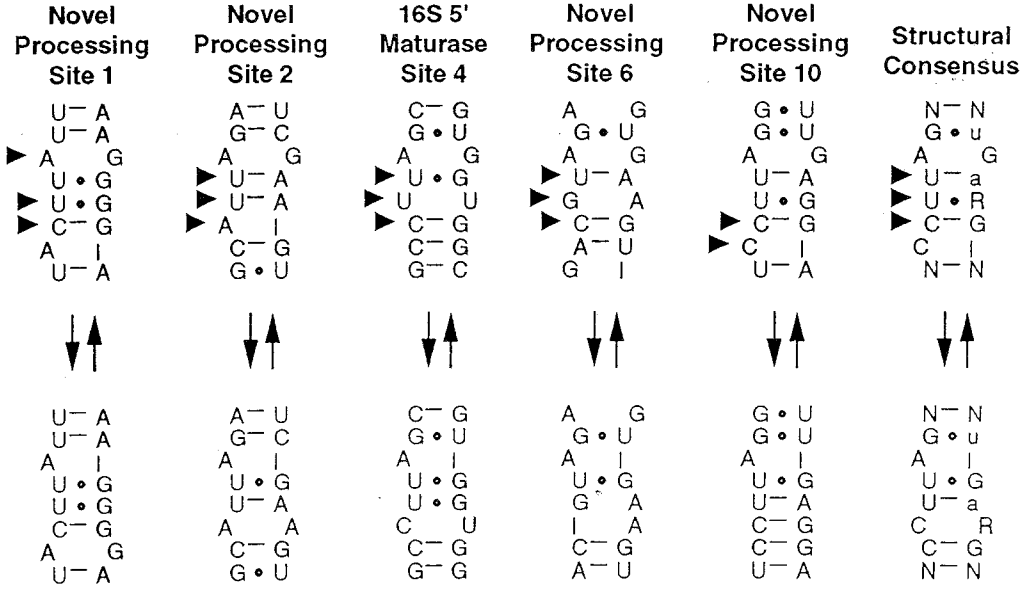
(2) *Sulfolobus* tRNA Sites Processing:

Met 5' end	CGGCTC **ATAAAC** TATGTA	1/6	2/6	1/4
Phe 5' end	CGGCTG **AAGTTT** TTAATA	1/6	1/6	1/4
Gly 5' end	GGCCTC **CCATTC** TTTAGA	2/6	1/6	1/4
Leu 5' end	GGGCTC **AGGACC** ATATGG	2/6	4/6	2/4
Ser 5' end	AGGCTC **GAGACC** ATTTCC	4/6	6/6	4/4
Met 3' end	TGTGGA **GAGACC** GGGTGG	4/6	6/6	4/4
Phe 3' end	CTCGCA **GTGACC** GGGTTG	3/6	5/6	3/4
Gly 3' end	AGAGTG **GAAAGC** CAGCAA	3/6	4/6	3/4
Leu 3' end	TGGTGA **GACCCC** CGATGG	4/6	4/6	4/4
Ser 3' end	AGGCGG **GCGACC** TACTGC	3/6	5/6	3/4

(3) Consensus Sequences:

novel con	ugGgca **GAUUCC** gG--Gg	6/6		
tRNA con	-Ggct- **gagacC** ---tg-		6/6	
total con	-Gg--- **Ga--CC** ---tg-			4/4

B (1)



transcribed leader, this would mark the first time that a biochemical correlation between the 5' and 3' maturation activities was established.

The presence of a weak match to the consensus sequence at the putative site of transcriptional termination is enigmatic. The 3' transcript end site does not conform well to any known archaebacterial termination motifs. Although two poly-T sequences are found within 180 nucleotides of the site, none are found in the immediate vicinity. If the site is in fact the result of processing, then it would at least be similar with a now-defined motif. The true site of termination would remain unknown.

4.11 Shine-Dalgarno Requirements

The 3' end of the *S. acidocaldarius* 16S rRNA is aligned with those of a variety of organisms (Figure 13). Sequence conservation of the terminal nine nucleotides is absolute, even in species where precursor processing produces a mature terminus shorter than the conserved sequence, with *S. acidocaldarius* as the only notable exception. The presence of a 58-60 nucleotide long extraneous sequence at the 3' end of the mature 16S rRNA puts in question the apparent requirement for a complementary interaction between 16S rRNA and mRNA in this archaebacterium. The problem may be resolved in one of three ways.

First, in *S. acidocaldarius*, translation initiation is not dependent upon complementary interactions between the 16S 3' terminus and the mRNA. The postulated interaction between *Sulfolobus* L12e mRNA and the 16S 3' terminus, for example, was based on four nonconsecutive complementary interactions over 8 nucleotides of the 16S 3' terminus. The probability of finding such a match in any 8 nucleotide sequence, with no limits set on the location of the complementary nucleotides in the transcribed leader sequence, is so great as to approach certainty, and renders the observation statistically insignificant.

Second, translational initiation is dependent on the interaction of the 16S pyrimidine-rich sequence, with no strict requirement for the sequence being at the terminus. The tertiary conformation of 16S rRNA within the 30S ribosomal particle conserves the spatial location of

FIGURE 13

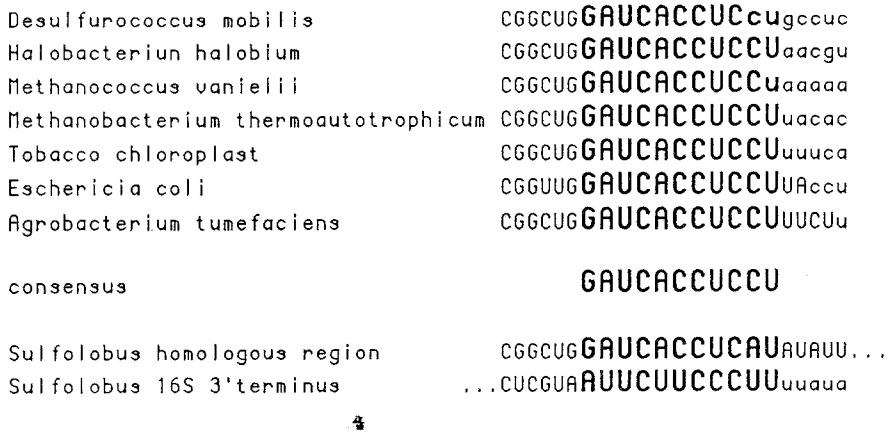


Figure 13: Alignment of 16S 3' Termini. The sequences around the 16S 3' terminus of a variety of organisms and organelles are aligned. Upper case letters indicate sequence found within the mature 16S molecule. Bold letters indicate the 11 nucleotides at the 16S 3' terminus that are not known to be in intramolecular secondary structure, but are postulated to be involved in the intermolecular complementary interactions required for mRNA transcriptional initiation in the eubacteria and perhaps the archaebacteria. Lower case letters indicate sequence in the intergenic spacer region, removed upon maturation. For Sulfolobus acidocaldariu, two sequences are shown. The first is from the conserved location relative to the rest of the 16S coding region (*Sulfolobus* homologous region), the second is from the conserved location relative to the 3' end of the mature molecule (*Sulfolobus* 16S 3' terminus).

the UCACCUCAU sequence via some as-yet unexplained mechanism, allowing it to interact with mRNA during translation initiation.

Third, translational initiation is dependent not only on the interaction of the 16S pyrimidine-rich sequence with the mRNA, but also on the sequence being present at the 16S 3' terminus. The additional 3' terminal sequence, which would normally be non-functional and result in a lethal phenotype, is functional by virtue of an encoded pyrimidine-rich sequence at the observed terminus (Figure 13). While notionally preserving the requirement for a 3' terminal pyrimidine-rich sequence, the 60 extra nucleotides would disrupt its spatial location. Inverted repeats within the additional 3' sequence, by forming a helical structure, may alleviate the problem somewhat (Figure 11A(1)).

4.12 Alternative Processing Pathways

Clearly, the nature of processing and maturation of *S. acidocaldarius* rRNA is unusual in its departure from previously determined archaebacterial norms. The apparent conservation of a sequence motif at the unique processing sites hints at the existence of a novel processing and maturation pathway but the observation is one of a mere correlation. A causal connection between structure and function needs to be established in order to validate these observations. Alternative processing pathways for rRNA maturation have been postulated and partially elucidated in *E. coli* by mutational inactivation of known processing enzymes[23]. The formation of mature termini is generally unaffected, but the transient accumulation of novel intermediate products leading to maturation indicate the activity of a battery of as-yet undefined enzymes.

4.120 Alternative Folding Model

The capacity for structural multiplicity of RNA has been implicated in its function in a variety of phenomena, including autocatalysis[90], termination-antitermination[40], and autogenous translational regulation[42]. The initial impetus for exploring structural multiplicity

in the *S. acidocaldarius* rRNA primary transcript was concern about the potential compromise to RNA stability during transcription. The 5' and 3' strands of the 23S precursor processing stem, for example, are separated by over 3000 nucleotides of transcript. Given the typical eubacterial RNA transcription rate of 50 nucleotides per second, the 5' strand would remain in a single-stranded conformation for a full minute before the 3' strand was transcribed. In a distinctly hostile thermophilic environment, unprotected RNA might be subject to extensive degradation.

As a result of the search for hidden structural potential, extensive capacity for alternative folding was discovered in the transcript. Virtually all of the flanking transcribed RNA could exist in more protected double-stranded forms. These conformational variants not only protect the RNA from non-specific degradation, but also offer interesting perspectives into the nature of the alternative processing and maturation pathway suggested by sequence conservation at the cleavage sites.

4.121 5' Flanking Region

The 5' flanking transcribed region from position 706 to position 860 can assume one of two conformations. In the context of the universal structure, the 5' flanking transcript would contain, apart from helices A and B, extensive single-stranded domains during the transcription of the 16S coding region (Figure 14B(1)). An alternative cloverleaf conformation can be modelled on the basis of internal complementary interactions (Figure 14B(2)). This hypothetical structure, consisting of four stem structures (labelled a, b, c, and d), is reminiscent of tRNA. The 10 nucleotides embedded within the 3' end of the proposed cloverleaf structure are part of the mature 16S 5' terminus. Support for this model, although circumstantial, is as follows:

1. In the cloverleaf conformation, all of the transcribed 5' flanking RNA is embedded within secondary structure. Single-stranded RNA is restricted to loops at the apices of three of the helical stem structures. This conformation also protects that part of the 16S rRNA 5' terminus which does not take part in intramolecular base pairing in the mature molecule. The

Figure 14: Alternate Conformations of the Operon Leader Sequence.

A. Schematic of the Operon Transcript Secondary Structure. The primary transcript is shown folded into the universal secondary structural conformation. The complex structures of the rRNA molecules have been excluded for the sake of visual clarity. The helical nomenclature is that of Garrett. The shaded region represents the intermediate stage when the operon's leader sequence as well as the 16S have been transcribed.

B. Schematic of the Leader Sequence Secondary Structures. The leader sequence is shown fluctuating between two possible conformations. The shaded regions are included for the sake of visual orientation between the two forms, and correspond to the structural features of the cloverleaf conformation.

(1) Universal Structure. The helical nomenclature in the universal structural context is that of Garrett.

(2) Cloverleaf Conformation. The helical nomenclature in the cloverleaf context is that used in the text. Limited structural correspondence exists between helices within the two form.

C. Features in an Alternate Context. The base pair interactions responsible for conferring stability to the cloverleaf structure are indicated. Bold letters indicate sequence from within the 5' end of the mature 16S molecule. Shading indicates regions of high sequence and structural similarity between the two species.

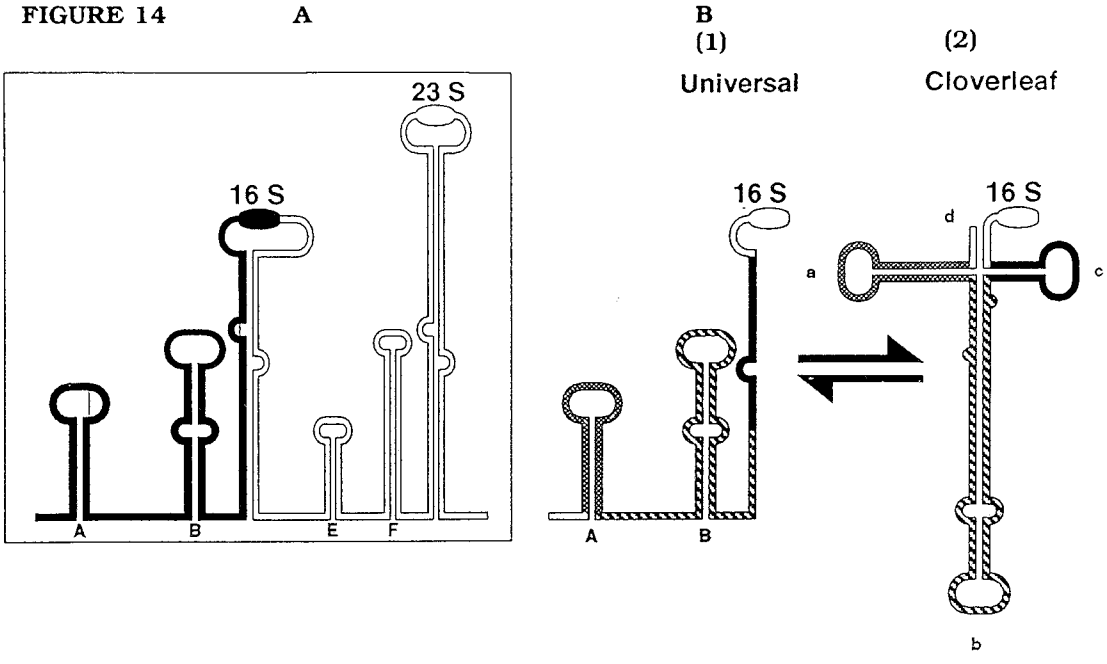
(1) *S. acidocaldarius*. The structure shown is derived from the transcribed leader of *S. acidocaldarius*. The filled triangles correspond to the locations of cleavage sites indicated from mapping experiments. The boxed numbers refer to the following processing sites: 1 = novel processing site 1; 2 = novel processing site 2; 4 = 16S 5' maturase site 4. Transcription start sites for the PCGene deletion series Δ4 to Δ8 are indicated relative to the alternate structure.

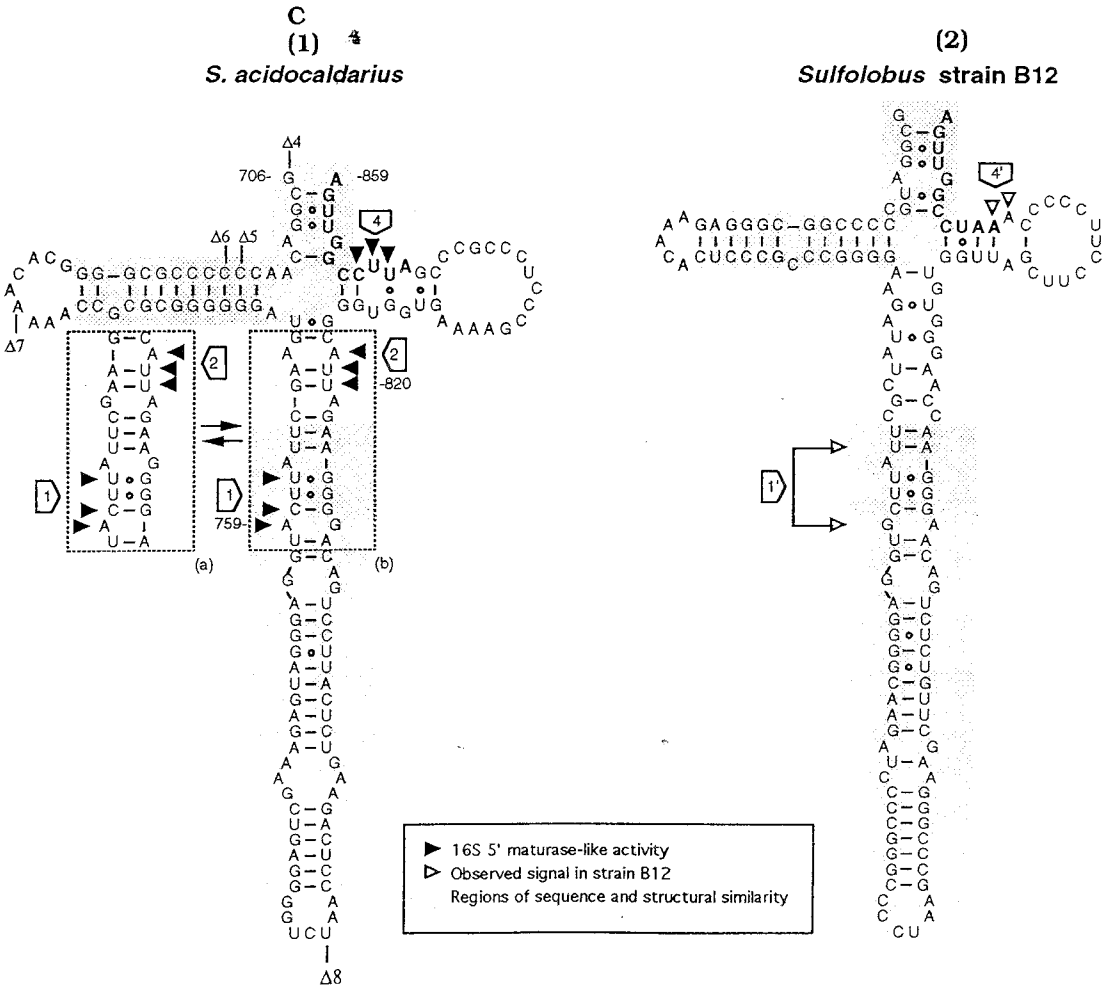
a. A-G Double Purine Bulge. In both processing sites, the 5'-GAUUCC-3' consensus sequence A residue is across from a G residue on the other strand. Refer to Figure 12(B) for details

b. Single A Residue Bulge. In this conformation, the GAUUCC consensus sequences' A residues are unpaired.

(2) *Sulfolobus* strain B12. The structure shown is derived from sequence of the transcribed leader of Sulfolobus strain B12. The open triangles correspond to the locations of cleavage sites identified by reverse transcript mapping of the *Sulfolobus* strain B12 rRNA operon. The boxed numbers refer to the following sites: 1' = postulated processing site; 4'= postulated 16S 5' maturation site[34].

FIGURE 14


A
B
(1)
Universal
(2)
Cloverleaf
23 S
16 S
16 S
16 S
A
B
E
F
A
B
a
b
c
d


C
(1)
S. acidocaldarius
(2)
Sulfolobus strain B12
706-
-859
-820
759-
(a)
(b)
16S 5' maturase-like activity
Observed signal in strain B12
Regions of sequence and structural similarity

structure's 3' terminal A residue (position 859) precedes the first helical domain located within the mature 16S rRNA.

2. In the cloverleaf conformation, novel processing sites 1 and 2 are brought into spatial proximity on opposite sides of stem-loop b, with one base pair separating the staggered consensus sequences. The local helical irregularities, more prominent at the processing sites than in the rest of the helix, exhibit a high degree of structural similarity around the axis of symmetry. In both cases, an internal helical loop marks the conserved A residue across from a G residue in the complementary strand, and a second bulge exists downstream near the end of the consensus sequence. For both structures, a variant conformation can be drawn where the conserved A residue exists in a bulge, and the internal helical loop exists near the end of the conserved sequence (Figures 14C(1) and 12B).

3. The 16S 5' maturase site is located on stem-loop c, allowing for the formation of an internal loop consisting of the conserved A residue across from a G residue, as well as for the variant conformation containing the A residue bulge. If, as sequence conservation indicates, the 16S 5' maturase site 4 as well as the novel processing sites 1 and 2 are recognized and cleaved by the same enzyme, the proximity of the three sites to the cloverleaf crux would increase the local concentration of the substrate.

4. Sequence of the 16S 5' region of *Sulfolobus* strain B12, obtained for the purpose of promoter analysis[34], exhibits a similar potential for folding into a cloverleaf conformation (Figure 14C(2)). High levels of both sequence homology and structural similarity exist, especially in stem-loop a (corresponding to helix A in the universal structure), the portion of stem-loop b from the position of novel processing site 1 down to the apical loop (corresponding to helix B in the universal structure), and stem-loop d. Both sequence and postulated structure appear to differ most at the top of stem-loop b and in stem-loop c. These regions of dissimilarity coincide with the observed positions of novel processing site 2 and the 5' maturation site 4, respectively, in *S. acidocaldarius*. Interestingly, the nuclease S1 mapping promoter analysis of *Sulfolobus* strain B12 indicated, aside from the sites of transcriptional initiation and 5' maturation, the presence of very weak bands at a position equivalent to that of the novel processing site 1.

Both the proposed sequence and structure of the novel processing site 1 are present in the *Sulfolobus* strain B12 cloverleaf conformation of the 16S 5' transcript.

4.122 Intergenic Region

The intergenic region from the poly-U tract comprising the 16S 3' maturase site 5 at position 2400 to the mature 23S 5' terminus at position 2537 can also assume resonant conformations. Unlike the 5' flanking sequence, whose alternative folding is postulated to create a static cloverleaf structure, the intergenic sequence may generate a dynamic structure dependent on transcript length. Upon transcription of 90 nucleotides of intergenic sequence, RNA folding could create the initial shortform (Figure 15B(1)), consisting of stem-loop g at positions 2423-2455 and stem-loop f (equivalent to helix F of the universal structure) at positions 2456-2489. The transcription of an additional 8 nucleotides would create the subsequent form (Figure 15B(2)) by extending the base pairing of stem-loop f at the sacrifice of stem-loop g. Only then would stem-loop e (equivalent to helix E of the universal structure) form at positions 2407-2443 to give rise to the longform.

Several features of this conformational dynamic should be pointed out. First, stem-loops g and e, by sharing sequence at positions 2423-2443, are structurally exclusive in a manner reminiscent of the *trp* antiterminator. Second, even though stem-loop g would be thermodynamically more stable than stem-loop e, the additional transcribed nucleotides, by extending stem-loop f, disrupting stem-loop g, and driving the formation of stem-loop e, would increase the total stability of the proposed subsequent structure. Finally, only upon the forced formation of stem-loop f would the sequence at the processing site 6 exist in a double stranded RNA context. The proposed double-stranded RNA exhibits the internal A-G loop and the potential for the A bulge-containing conformation postulated in the novel processing sites 1 and 2 and in the 16S 5' maturase site 4 (Figure 15C(2); see also Figure 12B).

Figure 15: Alternate Conformations of the Operon Intergenic Spacer Sequence.

A. Schematic of the Operon Transcript Secondary Structure. The primary transcript is shown folded into the universal secondary structural conformation. The complex structures of the rRNA molecules have been excluded for the sake of visual clarity. The helical nomenclature is that of Garrett. The shaded regions are a visual aid representing the sequence under consideration.

B. Schematic of the Intergenic Spacer Sequence Secondary Structures. The intergenic spacer sequence is shown fluctuating between two possible conformations. The shaded regions are the same as those in the universal structural context, and are included for the sake of visual orientation between the two forms. The helical nomenclature is that used in the text.

(1) The Shortform. This structure represents the intermediate found after transcription of the 16S 3' terminal sequence when the polymerase is transcribing around position 2489.

(2) The Longform. This structure represents the intermediate formed when the transcription of an additional 8 nucleotides causes stem-loop f extension at the cost of stem-loop g stability, resulting in the formation of stem-loop e containing the novel processing site 6.

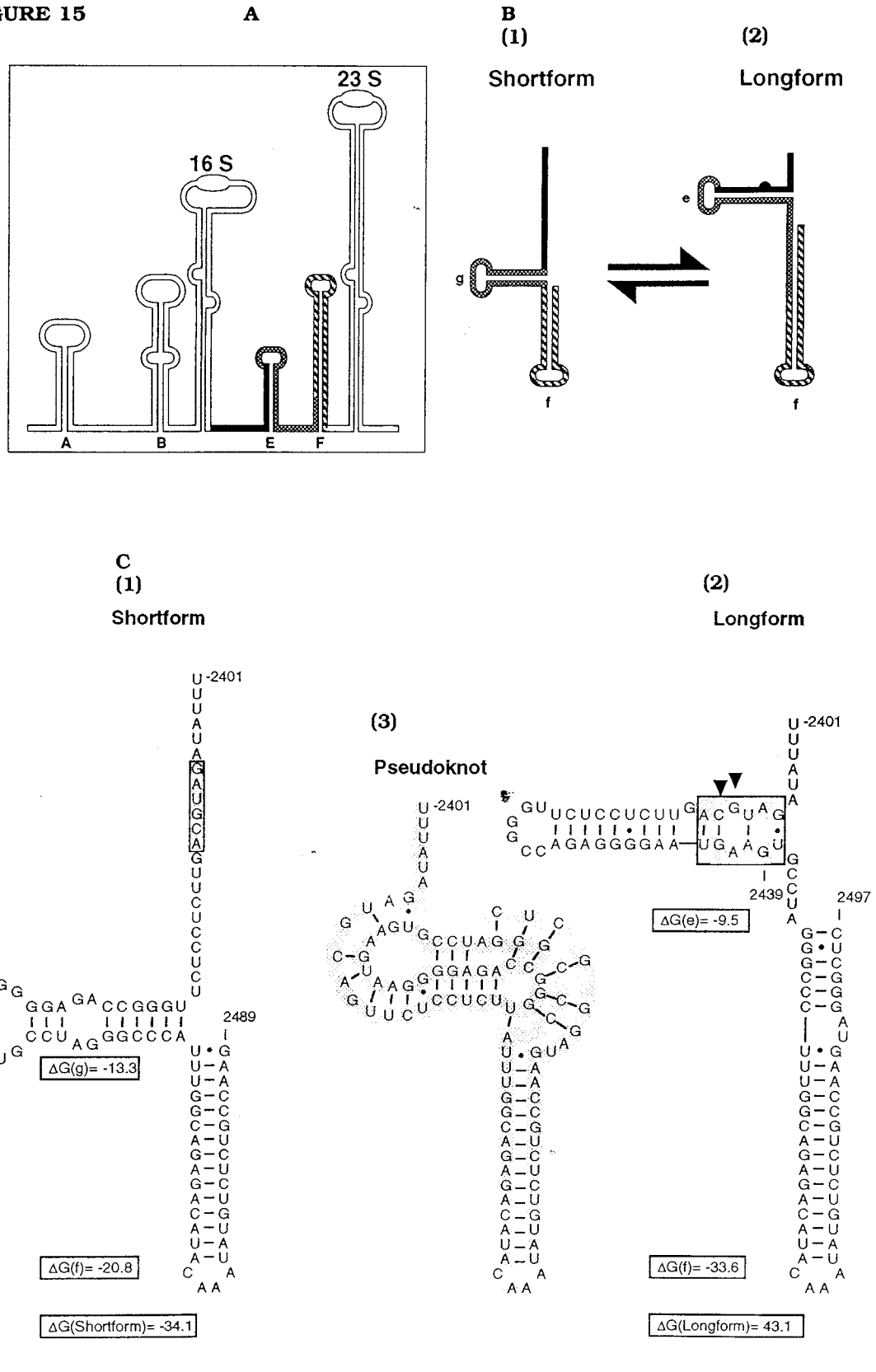
C. Features in the Alternate Context. The base pair interactions responsible for transferring stability from the shortform to the longform are indicated. The Gibb's free energy values for the helices, calculated from textbook values valid at 25°C are meant only to indicate relative helical stabilities. Shading indicates the phylogenetically conserved sequences at the base of helix E.

(1) The Shortform. The boxed region indicates the novel processing site 6 sequence motif.

(2) The Longform. The boxed region indicates the novel processing site structural motif in the A-G double purine bulge conformation. The alternative conformation of the novel processing site structural motif can be found in Figure 12(B).

(3) The Pseudoknot. An attempt to reconcile the base pair interactions of both the shortform and the longform simultaneously results in the pseudoknot shown.

FIGURE 15


A
23 S
16 S
A
B
E
F
B
(1)
Shortform
(2)
Longform
g
f
e
f
C
(1)
Shortform
(2)
Longform
(3)
Pseudoknot
ΔG(g)= -13.3
ΔG(f)= -20.8
ΔG(Shortform)= -34.1
ΔG(e)= -9.5
ΔG(f)= -33.6
ΔG(Longform)= 43.1

A structural alternative to the mutual exclusivity of the shortform and longform involves the formation of a pseudoknot between stem-loops e and g, where the mutual exclusivity is reconciled by both structures existing simultaneously.

4.2 NATURE OF 16S 5' MATURASE

The association of the conserved GAUUCC consensus sequence as well as the proposed irregular double-stranded helical RNA structure with the mapped sites of RNA processing is strongly correlative, but not causal. Preliminary attempts at elucidating the structure-function relationship between substrate and enzyme involved assaying the *in vitro* cleavage of the synthetically transcribed PCGene RNA substrate with enzyme-containing cell-free extract.

4.20 Cleavage Product Identification

The Δ3PCGene RNA transcript is 264 nucleotides long, and consists of the following sequences: positions 1-26 and positions 247-264 correspond to positions 69-43 and positions 23-5 within the pGEM-3Zf(+) multiple cloning site, positions 27-31 correspond to transcription of 5 nucleotides preceding the *S. acidocaldarius* 16S-23S rRNA transcription start site, positions 31-174 correspond to the 144 nucleotide 5' flanking region preceding the 16S sequence, and positions 175-246 correspond to the first 72 nucleotides of 16S sequence. Relative to the Δ3PCGene RNA transcript, novel processing site 1 is at position 80-85, novel processing site 2 is at position 144-149, and the 16S 5' maturase site 4 is at position 175-180 (Figure 10).

All three cleavage reactions were shown to proceed on the synthetic substrate, generating the four expected fragments. The 78-81 nucleotide long fragment *5'-1* (5 'transcribed leader sequence to novel processing site *1*) ran alongside the 45 nucleotide long size marker. This large deviation from its expected motility may be explained by the presence of the sturdy stem-loop a (helix A), consisting exclusively of 11 G-C base pairs. That this fragment truly is the 5' product, generated by processing at the novel processing site 1, was demonstrated by cleavage

of transcripts from other deletion clones of similar but not identical length; the mobility of the other three fragments remained unaffected, whereas the relative migration rates of the *5'-1* equivalent fragments was determined by the presence or absence of additional sequence. Fragment *1-2* (novel processing site *1* to novel processing site *2*) of length 61-66 ran as expected, as did fragment *2-m4* (novel processing site *2* to the 16S 5' *m*aturase site *4*) of length 28-32. No characterization other than motility was used to identify fragment *1-2*. Fragment *2-m4* was identifiable by monovalent ion-mediated cleavage inhibition at the 16S 5' maturase site giving rise to the disappearance of fragment *2-m4* and the concomitant appearance of a larger fragment *2-3'*, consisting of *2-m4* and *m4-3* '(16S 5' *m*aturase site *4* to the run-off transcript *3'* end). Fragment *m4-3'*, of length 84-91, was identified unambiguously by 3' end labelling of the substrate precursor.

Cleavage at the site of the mature terminus, as monitored by the release of the 84-91 nucleotide long *m4-3'* fragment, proves that maturation of the 5' end proceeds in a manner independent of the universal structure and probably consistent with the proposed alternative folding structure.

4.21 Optimal Activity

All four expected cleavage products are observed under a wide range of assay conditions. Cleavage appears to proceed efficiently over a wide range of temperatures and buffer concentrations. The production of all four cleavage products is not equal under all conditions, however. In Pharmacia One-4-All buffer concentrations greater than 1X, the disappearance of fragment *2-m4* is concomitant with the appearance of an extended fragment corresponding in length to *2-3'* (*2-m4* uncleaved from *m4-3'*); thus, cleavage at the 16S 5' maturase site 4 is inhibited, although intact *m4-3'* product is present, presumably via specific exonucleolytic trimming through the *2-m4* sequence appended to the mature terminus. This exonucleolytic trimming would represent another potential 16S 5' maturation pathway. In buffer

concentrations less than 1X, however, the *m4-3'* fragment appears to experience nonspecific terminal degradation. Whether this is due to differential affinity of the enzyme for the two substrates, or to two separate activities is unclear. Most of the subsequent assays were carried out in 2X buffer, with improved *m4-3'* integrity at the cost of *2-m4* visualization.

Enzymatic activity appears to be inhibited by monovalent cations. Concentrations greater than 40 mM of Na^+, K^+, or a combination of both result in total cleavage inhibition, yet salt concentrations lower than 20 mM result in fragment length heterogeneity. Whether this is due to non-specific degradation of cleavage products or relaxed enzyme specificity for the cleavage site is unclear. In the case of *E. coli* RNaseIII, it has been shown that substrate specificity is crucially dependent on ionic strength[75].

4.22 Cleavage Requirements

4.220 Double Stranded RNA Substrate

In the proposed alternative folding model, all three cleavage sites reside within irregular helical RNA. By deleting sequence through the novel processing site 1 to downstream position 790, the sequence coding for one side of stem loop b (helix B) in the vicinity of novel processing site 2 was also deleted, even though novel site 2 was itself unaffected. The transcript produced from Δ8PCGene was cleaved at a slower rate at the 16S 5' maturase site 4. Cleavage inhibition at novel site 2 was shown by the accumulation of a 65 nucleotide long fragment corresponding to the remainder of fragment *1-2* uncleaved from fragment *2-m4*. Two conclusions flow from this observation. First, stem a in the proposed alternative structure is not strictly required, but may increase the rate of processing at maturase site 4. Second, the paired novel processing sites 1 and 2 appear not to be necessary for the generation of the mature 16S 5' terminus.

Cleavage inhibition at site 2 of the Δ8PCGene transcript indicates that the enzyme is a double-strand specific endonuclease, although it is still unclear whether the double stranded context is required for substrate recognition or for substrate cleavage. Clearly, Δ7PCGene

transcripts (lacking the proposed stem-loop a and stem d, and by extension the proposed cloverleaf structure of the transcript, but retaining all three processing sites and their complementary sequences) exhibited robust processing at all three sites. If the double stranded context is necessary for recognition by the enzyme, that recognition is probably restricted to local rather than global structure. Also unclear is the nature of the two bands, corresponding to no expected products, present early in the incubation of Δ8PCGene transcript but disappearing upon the appearance of the novel 65 nucleotide-long fragment. The possibility of additional enzymatic activities associated with the putative maturase cannot be ruled out.

4.221 NTPαS Thio-Derivative Inhibition

The second observed mode of inhibiting enzymatic function involves incorporating nucleotide analogs into the substrate transcript. The analogs of UTP and CTP, alone or in combination, appear to inhibit all but the maturation reaction which produces fragment *m4-3'*. No other combinations of incorporated analogs produced any of the four expected cleavage products. Notably, the analog of GTP alone was sufficient to inhibit all three processing events.

Cleavage inhibition may take place by one of at least four mechanisms. First, incorporation of nucleotide analogs may disrupt the RNA transcript structure, shown to be vital in cleavage, sufficiently to mask recognition by the endonuclease. Second, the enzyme may be unable to cleave a thiophosphoester bond. This form of inhibition is well documented. Both exonuclease III and a set of restriction endonucleases are incapable of cleaving DNA containing thiophosphoester bonds. Third, total inhibition by the GTP analog hints at the possibility that incorporation into the complementary strand at the conserved G residue inhibits cleavage, perhaps by disruption of a necessary transition state intermediate, in a manner analogous to that of the *Tetrahymena* rRNA self-splicing intron[95]. Finally, if the cleavage reaction requires the binding of nucleotide cofactors, there may be a possibility of analog carry-over from the SP_6 polymerase reaction to the cleavage reaction and a subsequent inhibition of the enzyme.

4.222 RNP Complex Requirement for RNA Substrate Cleavage

The most interesting set of observations regarding the novel processing enzyme relate to its composition. When the cell-free extract was preincubated for 15 minutes with 400 pg of RNase A, cleavage of the substrate RNA was completely inhibited. Only the full length transcript was observed by autoradiography. Substrate cleavage therefore seems to be critically dependent on the presence of a non-substrate RNA moiety in the crude cell-free extract. When the cell-free extract was replaced by total cellular RNA or with free nucleotides, the cleavage reaction was not observed. Thus, there is a strong suggestion that cleavage is dependent upon the presence of a ribonucleoprotein complex. Neither RNA nor protein in isolation is capable of initiating cleavage; the reaction proceeds only in the presence of both components.

This apparent requirement for a ribonucleoprotein in *S. acidocaldarius* rRNA processing is somewhat reminiscent of the situation in some eukaryotes in at least three respects[89]. First, it has been shown that U3 snRNP - a 220 nucleotide nucleolus-localized RNA polymerase II transcript - plays a role in the rRNA maturation of certain species. Crosslinking experiments suggest and sequence complementarity supports an interaction between the External Transcribed Spacer - the eukayyotic equivalent of the archaebacterial 5' transcribed spacer - and the U3 snRNA. Furthermore, nuclease treatment of the U3 snRNP reduces or obliterates cleavage at a site within the External Transcribed Spacer. Second, as in *S. acidocaldarius*, this cleavage within the External Transcribed Spacer appears not to be universal to all eukaryotes, being freported in mammalian, but not yeast or *Xenopus* transcripts. Finally, in both organisms, pre-rRNA processing occurs without the structural requirement of a helix formed by the presence of inverted repeats flanking the coding sequence of the gene.

Despite the experimental evidence presented here, the possibility of RNA-catalyzed cleavage in the absence of protein cannot be excluded. It is possible that the reaction conditions for RNA alone differ from those of the ribonucleoprotein. Previous observations of *Tetrahymena* rRNA catalyzed cleavage in the presence of magnesium ions and nucleotides alone

does not preclude a requirement for additional or novel constituents in the *S. acidocaldarius* reaction[91, 92].

4.23 Is the Novel Endonuclease a Renegade Intronase?

Short introns have been found in a number of *Sulfolobus* tRNA genes[37, 38]. These introns vary in length between 15 and 25 nucleotides, and are located one nucleotide downstream of the anticodon. Sequences within the 3' ends of the introns are complementary to sequences within the anticodon loop, giving rise to symmetrically located staggered processing bulges on the precursor molecule, separated by anticodon-containing base pairs. The processing sites are located 11 nucleotides downstream of the start of the anticodon stem on the 5' side, and 6 nucleotides upstream on the 3' side.

The similarity between the cloverleaf structure of tRNA transcripts and the postulated cloverleaf conformation of the 16S 5' transcribed sequence is striking. As in the case of the tRNA, the novel processing sites 1 and 2 are symmetrically located on the descending central stem-loop b. Although this putative stem structure is significantly longer than in the tRNA, the locations of the processing sites relative to the central crux are the same. On the 28 base pair stem-loop b, the processing sites are located 10-14 nucleotides downstream of the start of the stem on the 5' side, and 4-6 nucleotides upstream on the 3' side.

Sequences around both the 5' and 3' processing sites of 5 intron-containing *Sulfolobus* tRNA molecules are aligned relative to the processing site (Figure 12A(2)). The consensus sequence gagacC (with cleavage reported between the second G residue and the second A residue) is similar to the GAUUCC consensus sequence observed around the processing sites in the 16S 5' transcript.

Finally, the catalytic role of RNA in intron splicing is well documented in the self-splicing class I introns (exemplified by the *Tetrahymena* rRNA intron), in the self-splicing class II introns (exemplified by the yeast mitochondrial lariat-forming intron of the *oxi 3* gene) and in the eukaryotic mRNA intron splicing ribonucleoprotein complexes[95]. The discovery that an RNA

moiety may be required for both precursor processing and 16S 5' maturation in *S. acidocaldarius* strengthens the suggestion of an intron processing-like complex being involved.

All of the evidence for the activity of an intronase is merely circumstantial, however, and each of the phenomena, taken in isolation, may have an alternative explanation:

1. The proposed requirement for a tRNA-like conformation of the 16S 5' flanking transcript is not absolute. Deletion of most of stem-loop a and stem d did not inhibit cleavage; even though the tRNA-like nature of the conformation had been destroyed, the potential for disrupted helical structure at the cleavage sites was maintained.

2. Despite elements of sequence similarity at all the cleavage sites, the consensus derived from the cleavage sites examined in this work differs from the consensus derived from an alignment of the tRNA cleavage sites. Using either consensus against the other data set results in a reduced level of identity (Figure 12A).

3. The implication of RNA involvement in catalysis exists in areas outside of intron removal. In the rRNA operon precursor processing of *E. coli*, the known ribonucleoprotein RNase P, is involved in tRNA 5' maturation; no introns are involved[77].

4.3 Dual Processing Pathway Model

4.30 Argument against Empirically Identified Artefacts

As stated earlier, our understanding of transcriptional regulation and processing of rRNA in archaebacteria is in its infancy. The full range of regulatory signals is not yet known or is not agreed upon, and work on the enzymatic processes associated with the signals has barely begun. Two additional factors have restricted the scope of the research done in the field. Because of the significance of archaebacteria in reshaping our understanding of evolution and phylogeny, and because of the lack of good transformation techniques with which to dissect genetic systems, a large part of the effort has focused on identifying important elements from the perspective of their similarity or dissimilarity to those of the bacteria and the eukaryotes. The

significance of structure is seen as paramount, and functional analysis, if addressed at all, is often assigned a lesser significance. Thus, the literature is resplendent with promoter consensus sequences and terminator motifs, and processing bulges in precursor rRNA stem structures are proclaimed significant pointers attesting to the phylogenetic unity of all archaebacteria. Kjems and Garrett go so far as to uphold the unity of archaebacteria by their identification of the highly conserved helices A to G, along with their conserved flanking sequence motifs - embedded in the many species-specific variants of the universally conserved primary transcript structure - without attempting to address the question of function[74]. There must be a functional purpose to these empirically observed elements other than that of providing significant phylogenetic landmarks from which taxonomists can take their bearing.

A summary of the present work in the context of what little is known about the functional details of transcriptional regulation and rRNA processing in *Sulfolobus* leaves some questions begging to be answered.

4.300 Presence of Signals on 16S; Presence of Site on 23S

When the primary transcript is folded into its universal configuration, staggered three nucleotide bulges on both the 16S and the 23S rRNA precursor processing stems - postulated to be the substrates for a processing endonuclease[49] - are found in *S. acidocaldarius*. The partially characterized enzymatic activity responsible for processing at these bulges is also present in *S. acidocaldarius*, as evidenced by the observed cleavages in the 23S precursor stem. The absence of the expected cleavage on the 3' side of the 16S precursor processing stem is therefore perplexing. A partial explanation may rest in the fact that an alternative mode of processing, postulated on the basis of alternative configurations and supported by *in vitro* cleavage analysis, may exist in *S. acidocaldarius*.

A dilemma arises in formulating an evolutionarily justifiable rationale for the presence of this apparently dysfunctional and therefore artefactual structure.

1. If the precursor processing stem is cleaved as predicted, then cleavage products should be observed in nuclease S1 mapping experiments. They are not.

2. Since the 16S precursor processing stem (known to be required in the standard archaebacterial processing pathway) appears to be dysfunctional, and given that there exists an alternative means of precursor processing, the helical structure should be actively selected against, given its genetic burden on the cell. At the very least, the structure should be subject to evolutionary drift, discernible by an accumulation of mismatches within the two complementary sequences. It appears to be subject to neither.

3. No strong argument can be made for the proposition that the observed novel processing pathway has been acquired so recently in evolutionary time that the effects of genetic drift are not yet discernible. The proposed cloverleaf structure shares sequence with the precursor processing stem. In fact, the novel processing site 2 is located in a part of the sequence that can fold either into stem b of the cloverleaf or into the ascending 5' side of the 16S precursor processing stem in the universal structure. Unless the 16S 5' flanking transcript had the pre-existing capability to fold into the cloverleaf conformation, that capability must have evolved through selection for a more stable structure. This would inevitably change the sequence of the precursor processing stem's 5' flanking sequence. That the complementarity of the 16S processing stem is undisrupted testifies to the existence of a strong selective pressure for compensatory changes in the 16S precursor processing stem's descending 3' sequence. The only reasonable conclusion is that the processing stem is functional, but the true function has not been observed under the experimental conditions outlined.

4.301 Terminator Motif and Cryptic Promoter

As was shown, nuclease S1 mapping indicates the presence of two sets of bands at the 3' base of the 16S processing stem, and suggests the existence of a complex intergenic regulatory mechanism. The novel processing site 6, located around position 2411, was shown to map to a sequence similar to the GAUUCC motif involved in novel 5' processing. Site 5,

FIGURE 16

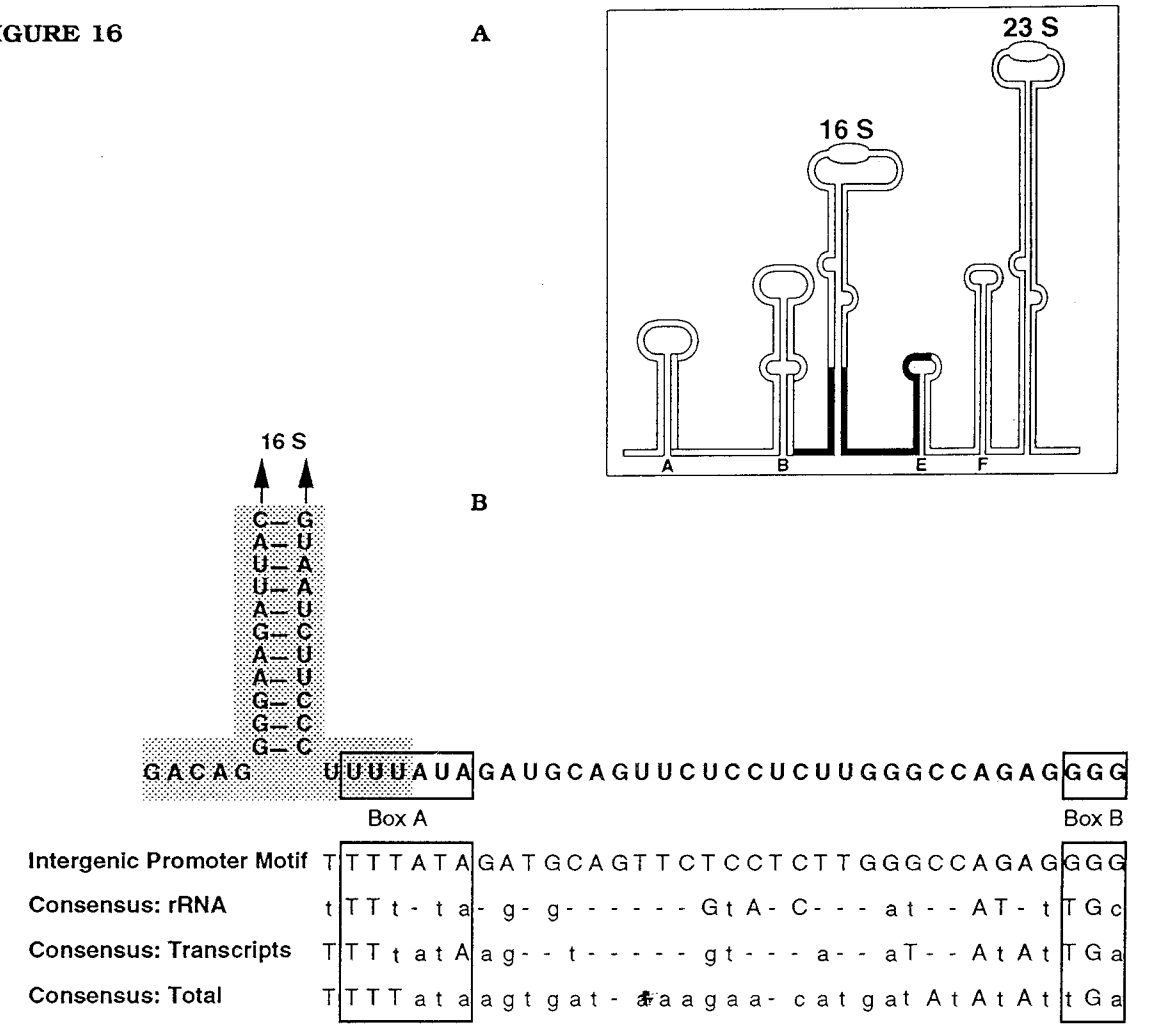


Figure 16: Putative Intergenic Overlapping Promoter-Terminator Motif.

A. Schematic of the Operon Transcript Secondary Structure. The primary transcript is shown folded into the universal secondary structural conformation. The complex structures of the rRNA molecules have been excluded for the sake of visual clarity. Shading represents the location of the putative intergenic promoter-terminator motif. Note that termination in the intergenic region is postulated to depend on the structure formed in part from the transcribed 5' leader sequence.

B. Structure and Sequence of the Putative Intergenic Overlapping Promoter-Terminator Motif. Shading highlights the putative terminator, consisting of the helical structure at the base of the 16S precursor processing stem and the poly-U tract immediately downstream of it. The labelled boxes indicate the locations of the box A and box B elements of the putative promoter. Directly below the putative promoter are the derived promoter consensus sequences from Figure 10(D).

located at positions 2400-2405, maps to a poly-U tract on the 3' side of the 16S processing stem and is costructural with putative *S. acidocaldarius* terminator motifs. Since the site 5 signal was detected both in experiments using probe DNA labelled either at its 3' end within the 16S coding region (which could detect termination signals) and those using probe DNA labelled at its 5' end within the 23S coding region (which could not), the signal cannot be generated by termination alone. Most if not all of the signal must be generated by an endonucleolytic mechanism, although it is impossible to rule out a termination-derived component in the experiment using 3' labelled probe DNA.

The presence of a 5'-stem-poly U-3' terminator-like motif in a highly structured RNA transcript might be meaningless, and the transcript mapping of a signal to the motif might be entirely fortuitous, were it not for one additional observation: downstream of the terminatormotif, and overlapping it slightly, is a perfect box A promoter element (Figure 16B). No bands corresponding to the putative site of transcription initiation at an appropriate distance downstream of the box A motif were observed in transcript mapping experiments.

4.31 Structure-Function Relationships

4.310 Hidden Determinants in Choice of Processing

S. acidocaldarius is cultured in the laboratory under what are assumed to be ideal conditions: the temperature is maintained at 75° C, the pH is continually monitored and adjusted to 4.0, the medium is kept well aerated, and cell density is kept below A_{600}=1.0. While maximizing cell yield, these ideal conditions may in fact bring with them methodological artefacts. Any analysis of protein or of RNA content will yield results that, while accurate, may not be complete. Growth state-dependent metabolic controls may activate some pathways and repress others, so that the overall picture of transcriptional and translational control is masked by the context-dependent particulars. Sequence and structural motifs associated with processing pathways induced under other sets of growth conditions would be present in the

transcript but would not be associated with any signals generated by regulatory or processing events in cells grown under laboratory conditions.

With no supporting data but for the presence of these motifs, the reconstruction of other processing pathways is not possible. However, by drawing together known information from related systems, models based on best estimate approximations may be proposed to give direction to future research.

The model proposed, based on the results of this work, consists of two processing pathways for the rRNA operon. In this model, the presence of the novel endonuclease determines the nature of transcript processing which in turn affects downstream transcriptional regulation.

4.311 The Observed Processing Pathway

The observed processing pathway is characterized by processing events independent of long-range RNA interactions. Once the polymerase has transcribed the 16S 5' flanking region and has moved on to transcribing the 16S gene itself, the flanking region folds into the cloverleaf configuration. The novel endonuclease, induced under defined growth conditions, recognizes and cleaves the nascent transcript at the novel processing sites 1 and 2 and at 16S 5' maturase site 4. The ramifications of this are twofold. First, the full length transcript does not accumulate in the cell. Second, formation of the universally observed 16S precursor processing stem is rendered impossible, since the 5' strand is not only dissociated from the transcript, but has been fragmented by the endonuclease.

The polymerase continues transcribing RNA through the sequence homologous to the 3' terminus of the 16S rRNA, through the 16S precursor processing stem 3' sequence, and into the intergenic space. The intergenic sequence folds either at first into the shortform (Figure 15C(1)) and then with additional transcription into structurally exclusive longform (Figure 15C(2)), or into the proposed pseudoknot structure (Figure 15C(3)). At this point, the novel endonuclease recognizes and cleaves at the novel processing site 6, liberating an unusual but nonetheless

functional 16S rRNA molecule. The polymerase continues until the 23S gene and its 3' flanking sequence have been transcribed. Here, cleavage within the precursor processing stem proceeds as expected.

4.312 The Postulated Processing Pathway

The postulated processing pathway is characterized by processing events dependent upon long-range interactions of RNA. Transcription of the 16S 5' flanking sequence results in the formation of the cloverleaf, followed by the subsequent transcription of the 16S gene. However, the novel endonuclease is either repressed or its cytosolic concentration is greatly reduced, so that little if any endonucleolytic cleavage takes place at the three 5' processing sites. The polymerase continues transcribing RNA through the sequence corresponding to the 3' terminus of the 16S rRNA and into the 16S precursor processing stem 3' sequence. The thermodynamically stable precursor processing stem begins to form at the expense of its mutually exclusive cloverleaf conformation.

Upon the completed transcription of the processing stem, a contingent transcriptional regulatory mechanism may be activated. In the presence of specific cytosolic factors, the 16S precursor processing stem along with its 3' poly-U tract acting as a signal for transcriptional termination, releases a 16S rRNA precursor replete with processing and maturation sequence and structural motifs. In the absence of these postulated cytosolic factors, transcription continues through the intergenic region and the 23S gene.

Since the competition between the shortform and the longform of the intergenic transcript is structurally independent of the 16S precursor processing stem sequences, processing at the novel processing site 6 is contingent only upon the presence of the novel endonuclease. The choice made at the base of the 16S precursor processing stem determines whether the polymerase transcribes only the 16S gene or both the 16S and the 23S genes together. Furthermore, an added level of transcriptional control involves the putative promoter-

like element directly downstream of the proposed terminator motif, allowing for the transcription of the 23S gene alone, independent of the 16S gene.

4.313 Is the Novel Endonuclease a Polycistronase?

In the model outlined, the novel endonuclease functions as a digital control switch. By cleaving the 16S 5' flanking sequence at three clustered processing sites in the cloverleaf conformation, it destroys a putative termination motif downstream of the 16S gene and ensures continued transcription of the polycistronic transcript. Termination in the intergenic region can only occur in the absence or repression of the processing enzyme, whereby the resultant formation of the proposed termination structure, along with the existence of a promoter-like sequence immediately downstream, effectively unlinks transcription of the 23S gene from the 16S gene. Notionally, polycistronic transcription is replaced by the potential for individual transcription of two proximal but unlinked genes. This combination of gene organization and transcriptional control occupies the conceptual fine line between the paradigms of polycistronic operon and of monocistronic transcription units.

4.4 Future Prospects

The field of archaebacterial RNA processing is not unlike an empty canvas waiting to be filled with new forms generated through creative insight. The present work was an initial attempt in that direction, albeit with a limited palette. The model of multiple processing pathways, copied from the bacterial school but with relevant modifications, attempts to sketch in the rough details of the archaebacterial system. Undoubtedly, in refining the details, the original sketch will be greatly modified, the palette expanded, and causal connectivity layered until a well-balanced picture is obtained.

Hopefully, some of the results of this work as well as some of the techniques developed will be incorporated into future explorations.

4.40 Answering "Is the Endonuclease a Renegade Maturase?"

Cleavage inhibition studies using the incorporation of the GTPαS thio-derivative into run-off transcripts can be further utilized to dissect the nature of the endonuclease substrate. Transcripts of an appropriate PCGene 3' deletion could be used to structurally complement Δ8PCGene transcripts. With the irregular helical RNA at both processing sites 1 and 2 thus dissected on reciprocal transcripts, selective incorporation of nucleotide analogs into one of the transcripts may be used to assess whether the analog must be present on the strand cleaved or on the strand containing the conserved semi-complementary G residue. This modified cleavage specificity would address whether the opposite strand is directly involved in catalytic activity.

The sensitivity of the 16S 5' maturase site 4 to the same general reaction conditions as the novel processing sites 1 and 2 has been demonstrated, but the identity of the RNA-enzyme complex in both cases has not. Cleavage inhibition studies using the incorporation of NTPαS thio-derivatives of nucleotides into transcripts has opened the door to the possible isolation and purification of the endonuclease complex. Competitive inhibition studies between labelled RNA substrate and unlabelled analog-containing RNA substrate should clarify whether processing is inhibited through the disruption of substrate recognition by the complex or of covalent bond cleavage in complex-bound substrate, and whether the mechanism is the same for both the 16S 5' maturase site 4 and the novel endonuclease sites 1 and 2.

If the complex can be shown to recognize but not cleave the analog-containing RNA substrate, a one-step purification protocol for the isolation of the endonuclease can be envisioned. The Δ3PCGene DNA fragment could be recloned into the Promega pSP64 (poly A) vector in front of the encoded $(dA:dT)_{30}$ sequence. Transcription driven from the SP6 promoter of the *Eco* RI linearized plasmid in the presence of nucleotide analogs would yield Δ3PCGene(polyA) RNA. The RNA substrate analog could be bound to oligo(dT)-cellulose columns in a modified binding buffer where divalent cations replace the standard monovalent cations. Under optimal conditions, cell-free extract passed over the column should wash through, leaving behind column-bound complex. A slight variant of this theme would use

biotinated oligo(dT) to bind the analog-containing substrate and streptavidin-coated magnetic beads to purify the endonuclease-substrate-oligo(dT) complex. Since both the oligo(dT) columns and the synthetic substrate are relatively inexpensive, large quantities of the complex could be isolated. Furthermore, since these methods are generally non-disruptive and the complex, derived from a thermoacidophile, is quite thermotolerant, enzymatic activity should be retained through the procedure. The question of whether the same RNA-protein complex is responsible for processing at both the 16S 5' maturase site 4 and the novel processing sites 1 and 2 could be answered directly.

Sequence from both the purified protein and the purified RNA component would allow for the construction of probes to screen the λGEM-11 *S. acidocaldarius* genomic library for the two genes. Cloning, sequencing and expression would naturally follow.

4.41 Answering "Is the Endonuclease a Renegade Intronase?"

Sequence similarity around cleavage sites of *Sulfolobus* intron-containing tRNA genes and of *S. acidocaldarius* rRNA transcript has raised the question of whether the two activities responsible for processing reside in the same RNP complex[37, 38]. *In vitro* transcripts of *Sulfolobus* intron-containing tRNA genes could be generated by a method identical to the generation of Δ3PCGene transcript, and could be subject to the same assays by the purified endonuclease complex. Observed cleavage would answer the question in the affirmative immediately. An absence of cleavage would require additional studies. The known intron-containing tRNA genes were probably isolated from *S. solfataricus*, a related species. The absence of *in vitro* processing could be a reflection of species differences in substrate recognition due to sequence differences at the processing site. Competitive inhibition studies between unlabelled tRNA transcripts and labelled Δ3PCGene transcripts could clarify whether the absence of processing is due to lack of substrate recognition or an inability to cleave bound RNA. Alternatively, cleavage of tRNA transcripts could be tested with native *S. solfataricus* cell-free extract to verify that *in vitro* cleavage is in fact possible.

Since an RNA moiety has been implicated in the endonucleolytic activity, the simplest mechanistic explanation for differences in the cleavage site sequence motif involves changes in the specificity of RNA-mediated recognition through sequence complementarity. Using the cloned RNA moiety gene from *S. acidocaldarius* as a probe for Southern hybridization of *S. solfataricus* DNA should allow for the cloning of the *S. solfataricus* gene. Reconstitution assays using the previously obtained protein and the newly obtained RNA could be attempted.

Furthermore, sequence differences between the two catalytic RNA molecules, in the context of conserved structural motifs, should point out sites responsible for substrate recognition as well as aid in the understanding of the underlying processing mechanism. Site directed mutagenesis would naturally follow.

4.42 Answering "Is the Endonuclease a Polycistronase?"

The most difficult point to affirm or refute, but perhaps the most important, is whether the endonuclease has the proposed role in the regulation of transcriptional termination. Its postulated role in processing at intergenic site 6 can be examined by techniques fundamentally identical to those used in defining its role in the 5' flanking region. Partial deletions of the cloned intergenic sequence can be used to produce transcripts corresponding in length and in sequence to the intergenic transcript's shortform and longform. These can then be assayed to confirm endonucleolytic cleavage in the second, but not the first, construct. The results generated would not distinguish between the exclusive shortform and longform, and the inclusive pseudoknot.

Recent work on the characterization of *Sulfolobus* RNA polymerase describes a system which uses the native enzyme for *in vitro* transcription from cloned DNA containing *Sulfolobus* promoters[39]. Implementing this technique on an intergenic clone containing the putative intergenic promoter-like element could help in defining its function.

Confirming transcriptional termination within the 5'-helix-poly(U)-3' termination motif at the base of the 16S precursor processing stem will be a more difficult matter. The multiplicity

of apparent events at the base of the stem renders difficult the task of isolating only the termination signal for analysis. An appropriate construct would consist of three linked sequences: DNA from the operon transcription initiation site to just short of the novel processing site B sequence within the 16S precursor processing stem at position 820, intergenic sequence corresponding to the cleavage-resistant shortform (these transcripts should not be processed by the endonuclease), and a nonspecific trailer sequence to distinguish termination from transcriptional run-off. This construct would contain the 5'-helix-poly(U)-3' putative termination motif ending at position 2405, but should not be amenable to the novel endonuclease cleavage reaction within the transcribed novel processing site 6 at position 2407-2412. Transcribed RNA would either terminate at position 2405, or would continue through the non-specific trailer sequence. A completely different approach to the elucidation of intergenic control could consist of exploring the effect of altered media conditions on detectable nuclease S1-mapped signals. If, as postulated, activation of the different processing pathways is dependent on the growth conditions of the culture, altered growth conditions might be reflected in altered transcript mapping signals.

5 BIBLIOGRAPHY

1. Woese, C.R. (1982) "Archaebacteria and Cellular Origins: An Overview" Zbl. Bakt. Abt. Orig. **3**: 1-17.

2. Dennis, P.P. (1986) "Molecular Biology of Archaebacteria" J. Bacteriol. **168**: 471-478.

3. Gropp, F., Reiter, W. D., Sentenac, A., Zillig, W., Schnabel, R., Thomm, M., and Stetter, K. O. (1985) "Homologies of Components of DNA-dependent RNA Polymerases of Archaebacteria, Eukaryotes and Eubacteria" Syst. Appl. Microbiol. **7**: 96-102.

4. Woese, C. R., and Olsen, G. J. (1985) "Archaebacterial Phylogeny: Perspectives on Urkingdoms" Syst. Appl. Microbiol. **7**: 161-178.

5. Garrett, R.A., Dalgaard, J., Larsen, N., Kjems, J., and Mankin, A.S. (1991) "Archael rRNA operons" TIBS. **16**: 22-26.

6. Zillig, W., Palm, P., Reiter, W., Gropp, F., Puhler, G.,and Klenk, H., (1988) "Comparative Evaluation of Gene Expression in Archaebacteria" Eur. J. Biochem. **173**: 473-482.

7. Lake, J.A. (1984) "Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes" Proc. Natl. Acad. Sci. USA **81**: 3786-3790.

8. Stoffer, G. (sp) (1986) "Electron microscopy of archaebacteria ribosome" Syst. Appl. Microbiol. **7**: 123-130.

9. Lake, J.A., Henderson, E., Clark, M.W., Scheinman, AA., and Oakes, M.I. (1986) "Mapping Evolution with Three Dimensional Ribosome Structure" Syst. Appl. Microbiol. **7**:131-136.

10. Pribnow, D. (1975) "Nucleotide Sequence of an RNA Polymerase Binding Site at an Early T7 Promoter" Proc. Natl. Acad. Sci. USA **72**: 784-789.

11. Pribnow, D. (1975) "Bacteriophage T7 Early Promoters: Nucleotide Sequences of Two RNA Binding Sites" J. Mol. Biol. **99**: 419-443.

12. Ishihama, A. (1988) " Promoter Selectivity of Prokaryotic RNA Polymerases" Trends Genet. **4**: 282-286.

13. Takanami, M., Sugimoto, K., Sugisaki, H., and Okamoto, T. (1976) "Sequence of Promoter for Coat Protein of Bacteriophage fd" Nature **260**: 297-302.

14. McClure, W. R., (1985) "Mechanism and Control of Transcription Initiation in Prokaryotes" Ann. Rev. Biochem. **54**: 171-204.

15. Helmann, J. D., and Chamberlin, M. J. (1988) "Structure and Function of Bacterial Sigma Factors" Ann. Rev. Biochem. **57**: 193-195.

16. McClure, W.R. (1985) "Mechanism and Control of Transcription Initiation in Prokaryotes" Ann. Rev. Biochem. **54**:171-204.

17. Schmitz, A., and Galas, D. (1979) "The Interaction of RNA Polymerase and *lac* repressor with the *lac* Control Region" Nucl. Acids Res. **6**: 111-137.

18. vonHippel, P. H., Bear, D. G., Morgan,W. D., and McSwiggen, J. A. (1984) "Protein-Nucleic Acid Interactions in Transcription" Ann. Rev. Biochem. **53**:389-446.

19. Gilbert, W. (1976) In *RNA Polymerase*, ed. R. Losick, Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

20. Maxam, A. M., and Gilbert, W. (1977) "Sequencing End-labelled DNA with Base Specific Chemical Cleavages" Proc. Natl. Acad. Sci. USA **74**: 560-564.

21. Reznikoff, W.S., Siegele, D.A., Cowing, D.W. and Gross, C.A. (1985) "The Regulation of Transcription Initiation in Bacteria" Ann. Rev. Genet. **19**: 355-387.

22. Menzel, R., and Gellert M. (1983) "Regulation of the Genes for *E. coli* DNA Gyrase: Homeostatic Control of DNA Supercoiling" Cell **34**: 105-113.

23. Gegenheimer, P. Watson, N., and Apiron, D. (1977) "Multiple Pathways for Primary Processing of Ribosomal RNA in *E. coli*" J. Biol. Chem. **252**: 3064-3073.

24. Leslie, A. G. W., Arnott, S., Chandrasekaran R., and Ratliff, R. L. (1980) "Polymorphism of DNA Double Helices" J. Mol. Biol. **143**: 49-55.

25. Reiter, W.D., Palm, P., and Zillig, W. (1988) "Analysis of Transcription in the Archaebacterium *Sulfolobus* Indicates that Archaebacterial Promoters are Homologous to Eukaryotic Pol II Promoters" Nucl. Acids Res. **16**:1-19.

26. Bucher, P., and Trifonov, E.N. (1986) " Compilation and Analysis of Eukaryotic Pol II Promoter Sequences" Nucl. Acids Res. **14**:10009-10027.

27. Reiter, W., Hudepohl, U., and Zillig, W. (1990) "Mutational Analysis of an Archaebacterial Promoter: Essential Role of a TATA Box for Transcription Efficiency and Start Site Selection *In Vitro*" Proc. Natl. Acad. Sci. USA **87**: 9509-9513.

28. Chant, J., and Dennis, P. (1986) "Archaebacteria: Transcription and Processing of Ribosomal RNA Sequences in *Halobacteruim cutirubrum*" EMBO J. **5**: 1091-1097.

29. Hui, I., and Dennis P.P. (1985) "Characterization of the Ribosomal RNA Gene Clusters in *Halobacterium cutirubrum*" J. Biol. Chem. **260**: 899-906.

30. Mankin, A.SS., Teterina, N.L., Rubtsov, P.M., Baratova, L.A., and Kagramanova, V.K. (1984) "Putative Promoter Region of rRNA Operon from Archaebacterium *Halobacterium halobium*" Nucl. Acids Res. **12**:6537-6547.

31. Mankin, A.S., Skripkin, E. A., and Kagramanova, V.K. (1987) "A Putative Internal Promoter in the 16S/23S Intergenic Spacer of the rRNA Operon of Archaebacteria and Eubacteria" FEBS Lett. **219**:269-273.

32. Mankin, A.S., and Kagramanova, V.K. (1988) "Complex Promoter Pattern of the Single Ribosomal RNA OPeron of an Archaebacterium *Halobacterium halobium*" Nucl. Acids Res. **16**:4679-4692.

33. Kjems, J., and Garrett, R.A. (1987) "Novel Expression of the Ribosomal RNA Genes in the Extreme Thermophile and Archaebacterium *Desulfurococcus mobilis*" EMBO J. **6**: 3521-3530.

34. Reiter, W., Palm, P., Voos, W., Kaniecki, J., Grampp, B., Schulz, W., and Zillig, W. (1987) "Putative Promoter Elements for the Ribosomal RNA Genes of the Thermoacidophilic Archaebacterium *Sulfolobus* sp. strain B12" Nucl. Acids Res. **15**: 5581-5595.

35. Rosenberg, M., and Court, D. (1979) "Regulatory Sequences Involved in the Promotion and Termination of RNA Transcription" Ann. Rev. Genet. **13**:319-353.

36. Bauer, C. E., Carey, J., Kasper, L. M.,Lynn, S. P., Waechter, D. A. Gardner, J. F. (1983) in *Gene Function in Eukaryotes*, ed. J. Beckwith, J. Davies, Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

37. Kaine, B. P., Gupta, R., and Woese, C. R. (1983) "Putative Introns in tRNA Genes of Prokaryotes" Proc. Natl. Acad. Sci. USA **80**: 3309-3312.

38. Kaine, B. P. (1987) "Intron-containing Genes of *Sulfolobus solfataricus*" J. Mol. Evol. **25**: 248-254.

39. Hudepohl, U., Reiter, W. D., and Zillig, W. (1990) "*In Vitro* Transcription of two rRNA genes of the Archaebacterium *Sulfolobus* ssp. B12 Indicates a Factor Requirement for Specific Initiation" Proc. Natl. Acad. Sci. USA **87**:9509-9514.

40. Li, S.C., Squires, C.L., and Squires, C. (1984) "Antitermination of *E. coli* rRNA Transcription is caused by a Control Region Segment containing Lambda *nut*-like Sequences" Cell **38**:851-860.

41. Schmeissner, U., Ganem, D., and Miller, J. H. (1977) "Genetic Studies of the *lac* Repressor: Fine Structure Deletion Map of the *lacI* Gene" J. Mol. Biol. **109**: 303-326.

42. Miozzari, G. F., and Yanofsky, C. (1978) "Translation of the Leader Region of the *E. coli* Tryptophan Operon" J. Bacteriol. **133**: 1457-1466.

43. Post, L. E., Arfsten, A. E., Reusser, F.,. and Nomura, M. (1978) "DNA sequences of Promoter Regions for the *str* and *spc* Ribosomal Protein operons in E. coli" Cell **15**: 215-229.

44. Kjems, J., and Garrett, R. A. (1985) "An Intron in the 23S Ribosomal RNA Gene of the Archaebacterium *Desulfurococcus mobilis*" Nature **318**: 675-677.

45. Hamilton, P. T., and Reeve, J. N. (1985) "Structure of Genes and an Insertion Element in the Methane Prodicing Archaebacterium *Methanobrevibacter smithii*" Mol. Gen. Genet. **200**: 47-59.

46. Bollschweiler, C., Kuhn, R., and Klein, A. (1985) "Non-repetitive AT-rich Sequences are Found in the Intergenic Regions of *Methanococcus Voltae* DNA" EMBO J. **4**: 805-809.

47. Muller, B., Allmanberger, R., and Klein, A. (1985) " Termination of a Transcription Unit Comprising Highly Expressed Genes in the Archaebacterium *Methanococcus voltae*" Nucl. Acids Res. **13**: 6439-6445.

48. Larsen, N., Leffers, H., Kjems, J., and Garrett, R.A. (1986) "Evolutionary divergence Between the Ribosomal RNA Operons of *Halococcus morrhuae* and *Desulfurococcus mobilis*" Syst. Appl. Microbiol. **7**: 49-57.

49. Wich, G., Hummel, H., Jarsch, M., Bar, U., and Böck, A. (1986) "Transcription Signals for Stable RNA Genes in Methanococcus" Nucl. Acids Res. **14**: 2459-2479.

50. Ostergaard, L., Larsen, N., Leffers, H., Kjems, J., and Garrett, R. (1987) "A Ribosomal RNA Operon and its Flanking Region from the Archaebacterium *Methanobacterium thermoaoutotrophicum*, Marburg Strain: Transcription Signals, RNA Structure and Evolutionary Implications" Syst. Appl. Microbiol. **9**: 199-209.

51. Wich, G., Sibold, L., and Bock, A. (1986) "Genes for tRNA and their Putative Expression Signals in *Methanococcus*" Syst. Appl. Microbiol. **7**:18-25.

52. Stahl, D. A., Luehrsen, K. R., Woese, C. R., and Pace, N. R. (1981) "An Unusual 5S RNA from *Sulfolobus solfataricus*, and its Implications for a General 5S rRNA Structure" Nucl. Acids Res. **9**: 6129-6137.

53. Luehrsen, K. R., Nicholson, D. E., Eubanks, D. C., and Fox, G. E. (1981) "An Archaebacterial 5S rRNA contains a Long Insertion Sequence" Nature **293**: 755-756.

54. Muller, B., Allmansberger, R., and Klein, A. (1985) "Termination of a Transcript Unit Comprising Highly Expressed Genes in the Archaebacterium *Methanococcus voltae*" Nucl. Acids Res. **13**: 6439-6445.

55. Kjems, J., Leffers, H., Garrett, R.A., Wich, J., Leinfelder, W., and Bock, A. (1987) "Gene Organisation, Transcription Signals and Processing of the Single Ribosomal RNA Operon of the Archaebacterium *Thermoproteus tenax*" Nucl. Acids Res. **15**: 4821-4835.

56. Grummt, I., Maier, U., Ohrlein, A., Hassouna, N., and Bachellerie, J. (1985) " Transcription of the Mouse rRNA Terminates Downstream of the 3' End of the 28S RNA and Involves Interaction of Factors with Repeated Sequences in the 3' Spacer" Cell **43**: 801-810.

57. Reiter, W.D., Palm, P. and Zillig, W. (1988) "Transcription Termination in the Archaebacterium *Sulfolobus*: Signal Structures and Linkage to Transcription Initiation" Nucl. Acids Res. **16**:2445-2459.

58. King, T.C., and Schlessinger, D. (1986) "Processing of RNA Transcripts" J. Bacteriol. **169**: 703-718.

59. Bedrook, J.R., Kolodner, R., and Bogorad, L. (1977) "*Zea mays* Chloroplast Ribosomal RNA Genes are part of a 22,000 Base Pair Inverted Repeat" Cell **11**:739-749.

60. Orozco,E.M., Gray, P.W., and Hallick, R.B. (1980) "*Euglena gracilis* Chloroplast Ribosomal RNA Transcription Units: I Gene Location" J. Biol. Chem. **255**:10991-10996.

61. Orozco,E.M., Rushlow, K.E., Dodd, J.R., and Hallick, R.B. (1980) "*Euglena gracilis* Chloroplast Ribosomal RNA Transcription Units: II Sequence Homology" J. Biol. Chem. **255**: 10997-11003.

62. King, T.C., Sir deskmuhk, R., and Schlessinger, D. (1986) "Nucleolytic Processing of Ribonucleic Acid Transcripts in Procaryotes" Microbiol. Rev. **50**: 428-451.

63. Tu, J., and Zillig, W. (1982) "Organization of rRNA Structural Genes in the Archaebacterium *Thermoplasma acidophilum*" Nucl. Acids Res. **10**: 7231-7245.

64. Hountonji, C., Blanquet, S., and Lederer, F. (1985) "Methiomyl-tRNA Synthetase from Eschericia coli: Primary Structure at the Binding Site for the 3" End of $tRNA^{Met}$" Biochem. **24**: 1175-1180.

65. Rould, M. A., Perona, J. J., Soll, D., and Steitz, T. A. (1989) "Structure of *E. coli* Glutaminyl tRNA Synthetase Complexed with $tRNA^{Gln}$ and ATP at 2.8 Angstrom Resolution" Science **246**: 1135-1141.

66. Fersht, A. R., Shi J. P., Knill-jones, J., Lowe D. M. Wilkinson, A. J., Blow, D. M., Brick P., Carter, P., Waye, M. M. Y., and Winter, G. (1985) " Hydrogen Bonding and Biological Specificity Analysed by Protein Engineering" Nature **314**: 235-238.

67. Achenbach-Richler, L., and Woese, C.R. (1988) "The Ribosomal Gene Spacer Region in Archaebacteria" Syst. Appl. Microbiol. **10**: 211-214.

68. Bram, R.J., Young, R.A., and Steitz, J.A. (1980) "The Ribonuclease III Site Flanking 23S Sequences in the 30S Ribosomal precursor RNA of *E. coli*" Cell **19**: 393-401.

69. Altamura, S., Caprini, E and Londei, P. (1991) "Early Assembly Proteins of the Large Ribosomal Subunit of the Thermophilic Archaebacterium *Sulfolobus*" J. Biol. Chem. **266**: 6195-6200.

70. Krych, A., Sirdeshmukh, R., Gourse, R., and Schlessinger, D. (1987) "Processing of *Escherichia coli* 16S rRNA with Bacteriophage Lambda Leader Sequences" J. Bacteriol. **169**: 5523-5529.

71. Wireman, J.W., and Sypherd,P.S. (1974) "*In Vitro* Assembly of 30S Ribosomal Particles from Precursor 16s RNA of *Escherichia coli*" Nature **247**: 552-554.

72. Dams, E., Londei, P., Cammarano, P., Vandenberghe, A., and DeWachter, R. (1983) "Sequences of the 5S rRNAs of the Thermo-acidophilic Archaebacterium *Sulfolobus solfataricus* (*Calderiella acidophila*) and the thermophilic eubacteria *Bacillus acidocaldarius* and *Thermus aquaticus*" Nucl. Acids Res. **11**:4667-4676.

73. Smith, J. D. (1974) In *Processing of RNA*, ed. J. J. Dunn, Brookhaven National Laboratory, Upton, New York.

74. Kjems, J., and Garrett, R.A. (1990) "Secondary Structural Elements Exclusive to the Sequences Flanking Ribosomal RNAs Lends Support to the Monophyletic Nature of the Archaebacteria" J. Mol. Evol. **31**: 25-32.

75. Birenbaum, M., Schlessinger, D., and Hashimoto, S. (1978) "RNase III Cleavage of *Escherichia coli* rRNA Precursors: Fragment Release and Dependence on Salt Concentration" Biochem. **17**: 298-307.

76. Krinke, L., and Wulff, D.L. (1990) "The Cleavage Specificity of RNase III" Nucl. Acids Res. **18**: 4809-4815.

77. Darr, S. C., Brown, J. W., and Pace N. R. (1992) "The Varieties of Ribonuclease P" TIBS Lett. **17**: 178-182.

78. Sirdeshmukh, R., and Schlessinger, D. (1986) "Why is Processing of 23S Ribosomal RNA in *Escherichia coli* not Obligate for its Function?" J. Mol. Biol. **186**: 669-672.

79. King, T.C., Sirdeshmukh, R., and Schlessinger, D. (1984) "RNase III Cleavage is Obligate for Maturation but not for Function of *Escherichia coli* Pre-23S rRNA" Proc. Natl. Acad. Sci. USA **81**: 181-185.

80. Srivastava, A., and Schlessinger, D. (1989) "Escherichia coli 16S rRNA 3'-end Formation Requires a Distal Transfer RNA Sequence at a Proper Distance" EMBO J. **8**: 3159-3166.

81. Viera, J., and Messing, J. (1982) "The pUC Plasmids, an M13mp7-derived System for Insertion Mutagenesis and Sequencing with the Synthetic Universal Primers" Gene **19**: 259- 268.

82. Yanisch-Perron, C., Viera, J., and Messing, J. (1985) "Improved M13 Phage Cloning Vectors and Host Strains: Nucleotide Sequence of the M13mp18 and pUC19 Vectors" Gene **33**: 103-119.

83. Chung, C. T., and Miller, R. H. (1988) "A Rapid and Convenient Method for the Preparation and Storage of Competent Bacterial Cells" Nucl. Acids Res. **16**: 3580-3581.

84. Maniatis, T., Fritsch, E. F.,and Sambrook, J. (1982) in *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

85. Sanger, F., Nicklen, S., and Coulsen, A. R. (1977) "DNA Sequencing with Chain Terminating Inhibitors" Proc. Natl. Acad. Sci. USA **74**: 5463-5467.

86. Dente, L., Cesareni, G., and Cortese, R. (1983) "pEMBL: A New Family of Single Stranded Plasmids" Nucl. Acids Res. **6**: 1645-1655.

87. Fienberg, A. P., and Volgenstein, B. (1982) "A Technique for Radiolabelling DNA Restriction Endonuclease Fragments to High Specific Activity" Anal. Biochem. **132**: 6-13.

88. Olsen, G. J., Pace, N. R., Nuell, M., Kaine, B. P., Gupta, R., and Woese, C. R. (1985) "Sequence of the 16S rRNA Gene from the Extreme Thermoacidophile Sulfolobus solfataricus and its Evolutionary Implications" J. Mol. Evol. **22**: 301-307.

89. Gerbi, S. A., Savino, R., Stebbins-Boaz, B., Jeppeson, C., and Rivera-Leon, R. (1990) "A Role for U3 Small Nuclear Ribonucleoprotein in the Nucleolus?" in *The Ribosome: Structure, Function, and Evolution.* ed, W. E. Hill, American Society for Microbiology, Washington, D.C.

90. Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A. and Cheng, H. L. (1986) "A Self-splicing RNA Excises an Intron Lariat" Cell **44**: 213-223.

91. van der Horst, G. and Tabak, H. F. (1985) "Self-splicing of Yeast Mitocondrial Ribosomal and Messenger RNA Precursors" Cell **40**: 759-766.

92. Chu, F. K., Maley, G. F., West, D. K., Belfort, M., and Maley, F. (1986) "Characterization of the Intron in the Phage T4 Thymidilate Synthase Gene and Evidence for its Self-excision from the Primary Transcript" Cell **45**: 157-166.

93. Savino, R. and Gerbi, S. A. (1990) "In vivo Disruption of Xenopus U3 snRNA Affects Ribosomal RNA Processing" EMBO J. **9**: 2299-2308.

94. Gerbi, S. A., Savino, R., Stebbins-Boaz, B., Jeppeson, C., and Rivera-Leon, R. (1990) "A Role for U3 Small Nuclear Ribonucleoprotein in the Nucleolus?" in *The Ribosome: Structure, Function, and Evolution.* ed, W. E. Hill, American Society for Microbiology, Washington, D.C.

95. Kim, S.H. and Cech, T.R. (1987) "Three Dimensional Model of the Active Site of the Self-splicing rRNA Precursor of Tetrahymena" Proc. Natl. Acad. Sci. USA **84**:8788-8792.