Sequences implicated in the processing of *Thermus thermophilus* HB8 23S rRNA

Roland K. Hartmann, Norbert Ulbrich* and Volker A. Erdmann

Freie Universität Berlin, Institut für Biochemie, Thielallee 63, 1000 Berlin 33, FRG

Received August 10, 1987; Accepted September 11, 1987

**ABSTRACT**

Nuclease S1 mapping analyses were performed in order to detect processing intermediates of pre-23S rRNA from *Thermus thermophilus* HB8. Two processing sites were identified downstream the start of transcription and several consecutive cleavage sites are associated with the mature 5'-end. In the 3'-flanking region one "primary" site and two cleavages which generate short-living intermediates were detected. A series of successive intermediates in the region of the mature 3'-end implies the existence of - in analogy to *Escherichia coli* - a 3'-exonucleolytic activity. The data were correlated with potential secondary structures within the pre-23S rRNA, which exhibit various repeated sequence elements. M13 sequencing data support the existence of one secondary structural element associated with the strong "primary" cleavage site in the 3'-flanking region. In *T. thermophilus* we can exclude the formation of an extended base-paired and precursor-specific stem enclosing the 23S rRNA which is inferred to mediate recognition by RNase III in *E. coli*.

**INTRODUCTION**

Previous data have shown, that in the extreme thermophilic eubacterium *Thermus thermophilus* the ribosomal RNA genes exist as two copies (1) and, that the 16S rDNA is separated by at least 7.8 and 6.4 kb from the 5'-end of the 23S/5S rDNA (2,3). In subsequent experiments we identified a promoter element within 200 base pairs upstream the 23S RNA gene (4). The data prove the existence of a transcriptional unit comprising only a 23S rRNA, 5S rRNA and a tRNA*"*" gene, which is unique for prokaryotes. The promoter element shows a high degree of homology to the -35 and -10 consensus sequences for promoters described for *E. coli* (5, 6). Furthermore a G/C rich sequence similar to the one believed to be under stringent control in stable RNA and ribosomal proteins genes of *E. coli* (7) could be identified. Employing nuclease S1 protection we were able to determine the in vivo start of transcription, which is identical with the in vitro start using *E. coli* RNA-polymerase. Furthermore we detected sequences (4) in the region following the origin of transcription, which are homologous to...
sections in *E. coli* rrn promoter-leader regions responsible for anti-
termination (8).

In this communication we present sequences flanking the 23S rRNA gene, which may basepair to form a stem and loop structure in the pre-23S rRNA. We characterized the maturation of the 23S rRNA in vivo by employing nuclease S1 mapping and present potential secondary structures of the pre-
23S rRNA, which may serve as recognition sites for enzymes involved in the processing cascade.

MATERIALS AND METHODS
Enzymes were purchased from Boehringer Mannheim and Bethesda Research Laboratories. Radioactive nucleotides for DNA sequencing were from Amer-
sham.

Bacterial strains
*T. thermophilus* HB8 (ATCC 27634) cells were grown at 70-75°C as described (9). *E. coli* strain BMH 71/18 (10) was used for M13 transformation.

Primer synthesis and DNA sequencing
Oligonucleotides used for M13 DNA sequencing were produced by the phos-
phoramidite method (11), using an automated DNA synthesizer (Applied Bio-
systems, model 380A). The oligodeoxynucleotides were purified by high pres-
sure liquid chromatography (12) or preparative gel electrophoresis. DNA
sequencing was performed by employing the M13/dideoxy method (13) as well
as by the chemical method (14).

5'- and 3'-labelling of DNA fragments
The 299 base pair Sma I/Bam HI fragment derived from the recombinant
plasmid pTT 700 (2,4), harbouring a complete 23S/5S rRNA operon was label-
led at its 5'-ends by T4 polynucleotide kinase (15). The labelled fragment,
comprising approximately 100 nucleotides of 23S rDNA, was used for nuclease
S1 protection analysis of the 23S rRNA 5'-flanking region.

A 501 base pair Ava I/Bam HI fragment originating from the plasmid pTT
700 (2,3,16) was labelled by filling the recessed 3'-end of the Ava I
 cleavage site with α-[32P]-CTP utilizing Klenow-polymerase (15). The label-
led fragment, comprising approximately 85 nucleotides of 3'-23S rDNA, was
used for nuclease S1 protection analysis of the 23S rRNA 3'-flanking re-

S1 nuclease mapping
Two procedures of RNA preparation were employed. According to one proce-
dure, exponentially growing cells were immediately extracted with 65°C hot
phenol. The second method was essentially as described (17). Exponentially
growing cells were mixed with equal volumes of SDS-solution [1% SDS, 100mM
NaCl, 8mM EDTA (pH 8), 100μg/ml carrier tRNA] which had been heated to
90°C. The mixtures were incubated at 90°C for 1 min. followed by phenol
extraction.

After denaturation (15 min. at 90°C), a20 ng of the respective 3' or 5'-
labelled DNA fragment was hybridized to 100 μg RNA of *T. thermophilus* at
60°C (18). The hybridization mixture was diluted with 300 μl of 280 mM
NaCl, 50 mM sodium acetate (pH 4.6), 4.5 mM zinc sulphate and 20 μg/ml
denatured hering sperm DNA and digested with 300-400 U/ml S1 nuclease for
30 min at 37°C. Analyses of digestion products were performed by elect-
rophoresis in 5% polyacrylamide gels containing 7 M urea.
Figure 1: Schematic representation of potential secondary structures surrounding the 23S rRNA transcript. The open triangle designates the start of transcription (4). Cleavage sites of processing enzymes deduced from data shown in Figure 4 and 5 are indicated by arrows. Two direct repeats are marked, one by open boxes merging into each other, the other by hatched boxes, with one heterologous nucleotide (positions 29 and 154) denoted by an asterix at position 29. The arbitrary numbering system does not consider the remaining 23S rRNA.

RESULTS

Secondary structure of pre-23S rRNA

In E. coli rRNA operons are transcribed as long primary transcripts (230S pre-rRNA). 16S and 23S rRNA sequences are included in large loops enclosed by strongly base-paired stems (19).
We determined the sequence surrounding one 23S rRNA gene of *T. thermophilus* and present two secondary structures (Fig. 1A and B), which may constitute recognition sites for rRNA processing.

As described for the *E. coli* pre-23S rRNA we can find a stem of 13 consecutive base pairs (Fig. 1, 2 and 3) confining the 5'-and 3'-termini of the mature 23S rRNA. In contrast to the pre-23S in *E. coli* this stem is followed by an internal loop of 14 (Fig. 1A) or 15 (Fig. 1B) nucleotides, respectively, and a potential hairpin structure of low stability (dg [25°C] = -0.2 kcal, Fig. 1B) comprising 9 (Fig. 1A) or 5 (Fig. 1B) base pairs. Due to its low stability, the formation of the hairpin requires additional stabilization by protein and/or tertiary interactions. The proposed models (A and B) differ significantly in the subsequent structural elements. The internal stem around position 41/113 comprising 8 bp in structure A, is extended to 14 bp in structure B exhibiting one looped out cytidine at position 115 and encompassing nucleotides, which are constituents of an internal loop in structure A. As a consequence in structure B the subsequent nucleotides may form a G/C rich stem with two looped out bases at position 20/21, whereas in structure A a short helix, including four G/C base pairs (positions 23-26 and 128-131), and two hairpin structures on the 5'-site (positions 2-22) as well as on the 3'-site (positions 139-151) of the flanking sequences can be formed. The 5'-hairpin structure, representing the antiterminational box B element (8), is directly preceded by the start of transcription (position 1).

**Sequence elements**

It is conceivable that in *T. thermophilus* certain palindromic sequences, as well as interrupted palindromes located several base pairs away from the site of the mature 23S rRNA, may serve as recognition sequences for proteins involved in RNA processing. Such proteins may bind to these sequences and bring the sites of processing in close proximity, which might be a prerequisite for precise processing.

Analysing the pre-23S rRNA sequence (Fig. 1) for conspicuous sequence motives we identified several direct repeats. One repeat is represented by a partially overlapping heptanucleotide sequence (positions 107-113/112-118, Fig. 1). A second dodecanucleotide sequence displaying one heterologous nucleotide at position 29 and 154, respectively, is located in the 5'- and repeated in the 3'-flanking region. At both locations (positions 24-35/149-160) the three 5'-proximal guanines are constituents of a helix in model A (Fig. 1).
Figure 2: Structural elements in the pre-23S rRNA of *T. thermophilus*. The nucleotide numbers are as in Figure 1. The enclosed section of four consecutive G/C basepairs interrupts an imperfect palindrome (denoted by open and solid triangles) created by the potential secondary structure presented. One perfect repeat (positions 54-62 and 65-73) overlaps with a second imperfect repeat (positions 50-62 and 59-73), interrupted by two bases (hatched nucleotides in positions 63, 64).

Inspection of the region adjacent to the putative termini of mature 23S rRNA (Fig. 2) revealed one perfect repeat (positions 54-62 and 65-73) which overlaps with a second imperfect repeat (positions 50-62 and 59-73), interrupted by two bases (positions 63, 64 in Fig. 2). The perfect repeat represents a duplicated box C inferred to be involved in antitermination (4). Furthermore a section of four consecutive G/C basepairs (71-74/91-94, Fig. 2) interrupts an imperfect palindrome (*UCAAG*) created by the potential secondary structure presented.

The homology of potential helical stems (I, II in Fig. 3c and d) may reflect the existence of a bivalent protein recognition site whose partial structural interconversion is induced by protein/RNA interactions. A prerequisite for the formation of helix I is a base pairing between the guanine in position 69 and the uridine in position 47 instead of the G/C base pair in position 69/96 in helix II. In order to construct an equal length for both helices the participation of the G/U base pair at position 48/68 has been included in helix II.

**Nuclease S1 mapping**

S1 mapping analyses were performed in order to detect processing intermediates of pre-23S rRNA. A 299-bp Sma I/Bam HI fragment (4), comprising
Figure 3: Comparison of inferred intermediates in 23S pre-rRNA processing of *E. coli* (a and b) and *T. thermophilus* (c and e). Nucleotide numbers in a and b are as in (21). The nucleotide numbers in c and e are as in Figure 1. Rs's enclosed by open triangles indicate cleavage sites by RNase III (21). The arrow and filled arrowheads indicate cleavage sites of processing enzyme(s) in *T. thermophilus*. Hatched boxes in a an b denote positions of homology with sequences that appear at a comparable location relative to several RNase III cleavage sites in phage T7 pre-mRNAs (26,27,28,29) and hatched boxes in c and d indicate an extended homology (with a C to U alteration in position 66) to the pre-mRNA of the s0.3X gene of phage T7 (29). Two structural domains, which exhibit significant homologies are enclosed by double lines. Elements displaying homology are framed by a solid (I) and dashed line (II) and are depicted (d).

approximately 100 nucleotides of 23S rDNA, and a 501-bp Ava I/Bam HI fragment derived from recombinant plasmid pTT 700 (2,3,16), comprising approximately 85 nucleotides of 3'- 23S rDNA, were labelled at their 5'- and 3' -termini, respectively, denatured and hybridized to total RNA from *T. thermophilus* and digested with nuclease S1. The exact length of protected fragments was deduced by coelectrophoresis of M13 sequencing fragments of the
Figure 4: Nuclease S1 mapping of the 5'-end of the 23S rRNA primary transcript of *T. thermophilus*. A 299-bp Sma I/Bam HI fragment (2,4), comprising approximately 100 nucleotides of 23S rDNA, was labelled at its 5'-termini and hybridized to total RNA of *T. thermophilus* and digested with nuclease S1. The size of protected single stranded DNA fragments was determined by electrophoresis in 7 M urea. 1 and 2: protection by RNA extracted by the hot phenol method at different exponential growth points, lane 1 (0.3 A500), lane 2 (0.76 A500); 3,4 and 5: RNA extracted by the SDS/phenol method at 0.13 A500 (lane 3), at 0.39 A500 (lane 4) and at 1.05 A600 (lane 5); 6,7,8 and 9: M13 sequencing of the coding strand of the Sma I/Bam HI fragment (A,C,G,T). Due to the presence of M13 primer and polylinker sequence, 24 bases have to be subtracted from the size of the fragments in lane 6-9 in order to correlate the actual length to the fragments in lane 1-5. The numbers next to lane 9 indicate positions of the thymidine ladder which correspond to protected fragments according to the numbering system in Figure 3; 11: Resolution of the band pattern and intensities around position 76 of lanes 3-5; 10: Same as lane 9 (Thymidine ladder); 12: Same as lane 7 (Cytidine ladder); Cleavage sites corresponding to protected fragments are indicated by arrows in Figure 1.

coding strand of the Sma I/Bam HI fragment (Fig. 4 and 5).

Between the start of transcription (Fig. 4, upper arrow, position 1 in Fig. 1) and the 5'-terminus of the mature 23S rRNA, around position 76, we could identify two pre-23S rRNA intermediates (Fig. 1; Fig. 4, lane 1-5).
Figure 5: Nuclease S1 mapping of the 3'-end of the 23S rRNA primary transcript of *T. thermophilus*. A 501-bp Ava I/Bam HI fragment derived from recombinant plasmid pTT 700 (2,3,16), comprising approximately 85 nucleotides of 23S rDNA, was labelled at its 3'-termini and hybridized to total RNA of *T. thermophilus* and digested with nuclease S1. The size of protected single stranded DNA fragments was determined by electrophoresis in 7 M urea. 1: protection by RNA extracted by the hot phenol method at 0.76 A\textsubscript{260}; 2 and 3: extracted by the SDS/phenol method at 0.75 A\textsubscript{260} (lane 2) and at 0.28 A\textsubscript{260} (lane 3); 4: M13 thymidine ladder as in Figure 4 lane 9. The numbers next to lane 4 indicate positions of the thymidine ladder which correspond to protected fragments according to the numbering system in Figure 1.

One scission, around position 31 (second arrow in Fig. 4), is located within the imperfect dodecanucleotide repeat sequence and can be found in both secondary structure models within an internal loop. The second cleavage site around position 54 (third arrow in Fig. 4) is located within the stem of the potential hairpin structure (dG [25°C] = -0.2 kcal, Fig. 1 B) and may result from an RNase III equivalent activity.
Figure 6: M13 sequencing of both strands comprising positions 128 - 161 in Fig. 1. 1: non-coding strand, the cytidine marked by an arrow represents the nucleotide preceding the region of band compression; 2: coding strand.

Protected fragments corresponding to mature 23S rRNA 5'-termini are barely visible in lanes 1 and 2 (Fig. 4), due to the method of directly extracting cells with 65°C hot phenol, which is inefficient for isolating RNA from ribonucleoprotein particles. In contrast, RNA used in lanes 3-5 (Fig. 4) was isolated from cells preincubated at 90°C in SDS-buffer followed by phenol extraction. The resolution of the protection analysis concerning the mature 23S rRNA 5'-termini is shown in lane 11 (Fig. 4). We observe two protected fragments between positions 71 and 73 (Fig. 1) which would coincide with the mature 5'-terminus suggested for *E. coli* 23S rRNA (20). The most prominent bands of protection are slightly shorter and correspond to cleavage sites between positions 76-81 (Fig. 1).

An additional fragment migrating above the band protected by the primary transcript (lanes 1 and 2 in Fig. 4, top arrow) could be detected, which may represent a further transcriptional initiation site.

In the 23S rRNA 3'-flanking region a strong primary cleavage site could be identified (Fig. 1, positions 147/149; Fig. 5, top arrow), which is associated with a stable hairpin formed by five G-C base pairs. Evidence for the existence of the hairpin structure (dG [25°C] = -8.4 kcal, positions
Nucleic Acids Research

139-151 in Fig. 1) comes from M13 sequencing data concerning complementary strands (Fig. 6) since the sequence ladders of either strand display band compression as soon as the hairpin can be formed at the termini of single stranded fragments. This hairpin may represent a cryptic terminator followed by an uridine stretch typical for rho-independent termination signals.

Two further cleavages in the 3'-flanking region (Fig. 5, lane 1, second and third arrow) generate short-living intermediates and lead to a further shortage of the transcript. A series of successive intermediates in the region of the mature 3'-end (between positions 89-98 in Fig. 1 and 5) implies the existence of - in analogy to E. coli - a 3'-exonucleolytic activity (21).

DISCUSSION

The ribosomal RNA genes in E. coli are transcribed as a single unit which is subsequently cleaved and modified to yield the three mature rRNA species. The observation of a large precursor molecule was made employing a RNase III deficient E. coli mutant (22). RNase III is involved in the early steps of the processing pathway of the ribosomal RNA transcript. The ribonuclease is dispensable for 16S rRNA maturation but is obligate for 23S rRNA maturation in E. coli (23,24). The enzyme cleaves a 7,500 bp long primary transcript of phage T7 encoding early functions to five different mRNA species (25,26). Analysis of regions surrounding the five cleavage sites of the phage T7 early mRNA precursor revealed little sequence homology, but comparable secondary structures could be predicted. In each region the cleavage occurs within an internal loop flanked by a hairpin (26,27,28, 29). In contrast, RNase III-cleavages of pre-16S rRNA and pre-23S rRNA in E. coli are located within precursor-specific stems of long-range base pairing, albeit the pre-16S rRNA cleavage site (19) and the pre-23S site at position 3492/3493 (21) are in immediate vicinity of a looped out, helix-destabilizing nucleotide. The manner in which RNase III chooses its exact cleavage site(s) in any RNA stem remains unclear. It seems improbable that in E. coli the mere length of the helical region could align the enzyme, but the existence of a helical region adjacent to RNase III cleavage sites is obviously one criterion of substrate specificity.

In T. thermophilus pre-23S rRNA several processing sites (positions 31/32 in Fig. 1 A and B, 121/122 in A, 53/54 in B) are associated with potential loop structures flanked by base paired segments. Other cleavage sites are
located within helical (positions 115-118 and 139-151, Fig. 1 A; 71-81, Fig. 1 A and B; 121-122 in Fig. 1 B) or partially destabilized helical (positions 115-118, Fig. 1 B) regions. Thus it is conceivable that all these cleavages originate from a RNase III equivalent activity.

There are however limited sequence homologies proximal to RNase III cleavage sites in the *E. coli* 23S pre-rRNA and the T7 early mRNA precursor (Fig. 3a and b), which may facilitate RNase III selection of the exact phosphodiester bond to be hydrolysed. It is noteworthy that five consecutive base pairs located in vicinity of the RNase III cleavage site within the pre-mRNA of the +0.3a gene of phage T7 (29) occur – with a C to U alteration in position 66 – in the *T. thermophilus* pre-23S rRNA (positions 49-53/64-68, Fig. 3c and e). Mutational alterations in this region drastically reduce the susceptibility of the T7 pre-mRNA to RNase III (30).

Consequently, the processing site at position 53/54 of the *T. thermophilus* pre-23S rRNA (Fig. 3c and e) probably represents a RNase III equivalent scission.

In the 13 bp stem enclosing the 23S rRNA we observed two cleavage products (positions 71-73 in Fig. 1), which would be equivalent to the mature 5'-terminus suggested for *E. coli* 23S rRNA by earlier investigations (20). However, the most prominent 5'-termini are slightly shorter and correspond to cleavage sites between positions 76-81 (Fig. 1). This observation is consistent with reported data derived from *in vitro* studies utilizing *E. coli* 23S RNA precursor molecules as RNase III substrate (21). It was observed that at least 10% of the 5'-termini generated by RNase III coincided with a position four nucleotides shorter than the anticipated mature 5'-terminus. However, we cannot exclude, that the shortened 5'-termini of *T. thermophilus* pre-23S rRNA are a consequence of hybrid instability in the A/U (A-T) -rich region between position 75-81 (Fig. 1). Under the hybridization conditions employed they could have resulted in an increased accessibility to the S1 nuclease. We exclude the possibility of »nibbling« activity in S1 nuclease reactions, since no significant bands of protection can be observed which correspond to cleavage sites between positions 73-76 (Fig. 1).

Furthermore it was reported that in an RNase III - deficient *E. coli* mutant all ribosomes are provided with unprocessed pre-23S rRNA (21). This unprocessed pre-23S rRNA contains in addition to the mature molecule a double stranded stem of variable length with up to 100 bases at each terminus. The precursor structural elements are even stable enough to be
visualized by electron microscopy under conditions under which the secondary structure of the remaining ribosomal RNA is denatured (24,31). It is clear that the incorporation of extended pre-23S rRNA molecules into the ribosome does not lead to severe structural perturbations, since the aberrant subunits are able to participate in protein biosynthesis. Thus the data indicate the existence of a short stem involving the mature 3'- and 5'-termini in wild-type ribosomes.

In this context, our observation of microheterogenous 5'-termini in pre-23S rRNA of *T. thermophilus* apparently reflects - in analogy to *E. coli* - a relaxed stringency of accuracy in 5'-terminal 23S rRNA maturation.

ACKNOWLEDGEMENT

The financial support of these studies from the Deutsche Forschungsgemeinschaft (DFG-SFB 9/85), the Fonds der Chemischen Industrie e.V. and the FGS "Regulatorische Elemente" (Freie Universität Berlin) is acknowledged.

REFERENCES
