The Organization and Expression of Essential Transcription Translation Component Genes in the Extremely Thermophilic Eubacterium Thermotoga maritima*

Daqing Liao and Patrick P. Dennis‡

From the Canadian Institute for Advanced Research, Program in Evolutionary Biology and Department of Biochemistry, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3

(Received for publication, April 30, 1992)

A 5789-nucleotide-long EcoRI fragment from the genome of Thermotoga maritima, identified by cross-hybridization to L11, L1, L10, and L12 ribosomal protein gene sequences from Escherichia coli, was cloned and sequenced. The fragment encodes five tRNAs (tRNA\textsuperscript{met}, anticodon complementary to AUG; tRNA\textsuperscript{asw}, AUG; tRNA\textsuperscript{waa}, ACA; tRNA\textsuperscript{wyy}, UAC; tRNA\textsuperscript{wug}, UGG), the transcription termination-antitermination factor nusG, the four 50 S subunit ribosomal proteins L11, L1, L10, and L12, and the amino-terminal portion of the RNA polymerase \(\beta\) subunit protein. The five tRNA genes, the nusG gene, and the L11, L1, L10, and L12 ribosomal protein genes form a complex transcription unit. Transcripts appear to be initiated from an upstream promoter, \(P_1\), located in front of the tRNA\textsuperscript{met} gene and from three internal promoters: \(P_2\) is located immediately in front of the tRNA\textsuperscript{met} gene; \(P_{110}\) is near the beginning of the L1–L10 intergenic space, and \(P_{112}\) is at the end of the L10 gene sequence. The tRNA sequences are excised from the leader regions of the \(P_1\)- and \(P_2\)-initiated transcripts. Three putative but potentially important regulatory sequences were identified within this operon: an L1 translational control site, a transcription attenuator, and a strong rho-independent terminator. The strong terminator located distal to the L12 gene overlaps a fifth promoter, \(P_5\), which is used to initiate transcripts of the downstream RNA polymerase \(\beta\) subunit gene. The \(T.\ maritima\) NusG protein exhibits 43% amino acid sequence identity when aligned to the \(E.\ coli\) protein; the alignment is interrupted by a large 171-amino acid-long insertion into the \(T.\ maritima\) protein after codon 45.

Living organisms derive from a common primordial ancestor and divide into three easily recognizable domains or lineages: the eubacteria, the archaeabacteria and the eucaryotes (Woese et al., 1990). In spite of this superficial understanding, our knowledge relating to the molecular features of the common ancestor and the precise origins and relationships between the three surviving domains or lineages remains obscure.

* This work was supported by a grant from the Medical Research Council of Canada (P. P. D. principle investigator). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank\textsuperscript{TM}/EMBL Data Bank with accession number(s) Z11839.

‡ A fellow in the Evolutionary Biology Program of the Canadian Institute for Advanced Research. To whom correspondence should be addressed. Tel.: 604-822-5245; Fax: 604-822-5227.
In addition, the proximal ribosomal protein transcripts contain two well characterized sites used for autogenous translational regulation. The first is a mimic of the L1 binding site in 23 S rRNA and is located immediately in front of the L11 translation initiation codon. A deficiency in 23 S rRNA production allows L1 protein to bind to the mRNA and block translation of the L11 and L1 cistrons. The long L1-L10 intergenic space contains a second control region which binds L10 (or the L10-L12 complex); protein binding is believed to switch the conformation of the mRNA to a structure which exhibits greatly reduced translational efficiency.

The region upstream of this ribosomal protein-RNA polymerase operon in E. coli is occupied by four tRNA genes, tufB (one of two genes encoding the translation elongation factor Tu), and the secEnusG operon encoding two essential proteins involved, respectively, in protein export and in transcription termination-antitermination (An and Friesen, 1980; Schatz et al., 1990; Downing et al., 1990; Sullivan et al., 1992; Linn and Greenblatt, 1992). Our analysis indicates that T. maritima lacks a tufB and secE gene in this region and that five tRNA genes and nusG are cotranscribed with the four ribosomal protein genes. The downstream RNA polymerase genes are transcribed separately.

**MATERIALS AND METHODS**

**Bacterial Strains and Plasmids—**E. coli strain ML101 (Δlac-proAB, supE, thi-) was used in these experiments. Escherichia coli strain B4 (lacUV5, traD80 proAB'), JM83 (recA1, supE44, endA1, hsdR17, gyrA96, relA1), triAD36, proAB', lacI, ΔlacZΔM15), and DH5αF′ (Δ lacZYA-argF’, U169, endA1, recA1, hsdR17(θ-mx'), deor, thi-1, supE44, λ', gyrA96, relA1/F′ φ80 dalZAM15) were used for cloning.

T. maritima strain MSB8 was cultured at 75 °C in MMS medium (Haberkamp et al., 1984). MMS medium contains (per liter) NaCl, 6.83 g; MgSO₄·7H₂O, 1.75 g; MgCl₂·6H₂O, 1.38 g; KCl, 0.16 g; NaBr, 25 mg; H₂BO₃, 0.75 mg; SrCl₂·6H₂O, 3.8 mg; KI, 0.025 mg; CaCl₂·0.38 g; KH₂PO₄, 0.6 g; Na₂SO₄·5 g; (NH₄)₂SO₄·2 g; trace minerals (Balch et al., 1979), 15 ml; resazurin, 1 mg; starch, 5 g; pH 6.5 (adjusted with H₂SO₄). The T. maritima strain and the recombinant plasmid pUC-TB4 which contains a portion of the T. maritima rpoB gene and about 1 kb of upstream sequence were kindly provided by W. Zillig (Max-Planck Institute, Martinsried, Germany). Plasmids pGEM-7Zf(+) and pGEM-7Zf(−) (Promega) and XGTL10 were used for cloning.

**Standard Single Letter Amino Acid and Nucleotide Nomenclature** was used. The abbreviations used are: kb, kilobase; W, nucleotide A or T; S, nucleotide G or C; PIPES, 1,4-piperazinediethanesulfonic acid.

1 The abbreviations used are: kb, kilobase; W, nucleotide A or T; S; nucleotide G or C; PIPES, 1,4-piperazinediethanesulfonic acid.

2 W. Zillig, personal communication.

**RESULTS AND DISCUSSION**

The rj region of the E. coli chromosome contains a cluster of essential genes that encode components of the transcription-translation apparatus (Lindahl et al., 1975). Included are genes for the large ribosome subunit proteins L11, L12, and L1 and L12 and the β and β′ subunits of RNA polymerase. To identify the region in the T. maritima genome that encodes the equivalent large subunit ribosomal proteins, genomic DNA was digested and probed by Southern hybridization with the 2.2-kb EcoRI fragment from E. coli (Fig. 1). At medium stringency, the probe hybridized to a single 5.8-kb EcoRI fragment.

The T. maritima 5.8-kb fragment was shown to contain a single XbaI site located 2.2 kb from one end. The same 2.2-kb XbaI-EcoRI fragment was identified as the terminal part of a larger 5.0-kb XbaI-HindIII genomic fragment present in the recombinant plasmid pUC-TB4. This 5.0-kb insert fragment was known to encode the amino-terminal portion of the RNA polymerase β subunit protein. These Southern hybridizations therefore suggested that some or all of the L11, L1, L10, and L12 equivalent ribosomal protein genes were, as in...
E. coli, located proximal to the RNA polymerase β subunit gene in the T. maritima genome.

By using the 2.2-kb XbaI-EcoRI fragment from pUC-TB4 as probe, it was possible to isolate the genomic 5.8-kb EcoRI fragment as an insert in λGT10; attempts to subclone this EcoRI fragment into a number of different plasmid vectors were unsuccessful. However, from the recombinant λGT10, the 2.2-kb XbaI-EcoRI and the overlapping 4.0-kb EcoRI-SacI fragment were isolated and subcloned to give plasmids pFD934 and pFD990. The complete nucleotide sequence of the overlapping 4.0-kb EcoRI-SacI and 2.2-kb XbaI-EcoRI fragments yielded the sequence of the entire 5789-nucleotide-long genomic EcoRI fragment. The sequence contains a cluster of five tRNA genes, a long open reading frame designated nusG, four genes encoding the equivalents of the E. coli L11, L1, L10, and L12 large ribosome subunit proteins, and, as expected, the 5′ portion of the open reading frame encoding the equivalent of the E. coli RNA polymerase β subunit protein.

Comparison of the content and location of genes between E. coli and T. maritima reveals both similarities and differences (Post et al., 1979; Downing et al., 1990; An and Friesen, 1980). First, the tRNA\textsuperscript{th} and tRNA\textsuperscript{ty} genes of T. maritima have the same anticodon as the thr\textsuperscript{T} and tyr\textsuperscript{U} genes of E. coli; the other tRNA genes show no correspondence (Fig. 1). Second, this region in T. maritima lacks genes or sequences related to tu\textsuperscript{B} (EF Tu) or secE of E. coli. The tufA gene of T. maritima has been located and cloned from elsewhere in the genome (Bachleitner et al., 1989). A second copy of this gene in the E. coli and T. maritima reveals both similarities and differences in the genome (Bachleitner et al., 1989). A second copy of this gene in the E. coli and T. maritima reveals both similarities and differences in the genome (Bachleitner et al., 1989).

The tRNA Gene Cluster—Four of the five tRNA genes, Met1, Thr, Tyr, and Trp, encode full length molecules that include the 3′-terminal CCA acceptor sequence and all can be folded into the universal clover leaf structure (Fig. 3). In contrast, the tRNA\textsuperscript{met2} gene encodes a 73-nucleotide-long truncated tRNA. Although the sequence ends with a CCA 3′ terminus, the two Cs are buried as part of the 7-base pair acceptor stem. It seems likely that activation of this tRNA requires the posttranscriptional addition of the terminal CCA acceptor sequence by a nucleotidylderminal transferase. Alternatively, the tRNA\textsuperscript{met2} might exhibit an atypical folding pattern where the stem of the T strand is contracted from 5 to 2 base pairs; this would result in the expansion of the variable loop from five to eight nucleotides and extend the 3′-terminal GCCA sequence above the 7-base pair acceptor stem. In this alternative configuration, the acceptor stem would contain a C-A mismatch at position 6. There is no indication as to which, if either, of the two methionine tRNAs might serve as the initiator in the translation initiation process.

The conclusions from these experiments are summarized...
Essential Transcription Translation Genes

Fig. 2. Nucleotide sequence of the T. maritima 5.8-kb EcoRI genomic fragment. The sequence of the 5789-nucleotide-long EcoRI fragment from the genome of T. maritima is illustrated. The five putative promoters P1, P2, P1(lo), P1, and P2 are indicated above the sequence; the major start sites are indicated (●). Restriction sites used for transcript mapping studies are indicated above the sequence. The position of the five tRNAs (□, anticodon □□□□□), and the predicted amino acid sequences of the proteins encoded by genes on the fragment are indicated below the nucleotide sequence. Translation initiation sequences complementary to the 3'-end of 16 S rRNA are underlined.
Fig. 2—continued

Fig. 3. Structure and processing of tRNAs. The structure of the five tRNAs encoded on the 5.8-kb EcoRI fragment are depicted. The shaded nucleotides are present in the primary transcript but removed during tRNA processing and maturation. The tRNA met is unusual; it either requires CCA addition for activation or it has a very unusual structure with a mismatch (C-A) at position six in the acceptor stem. Both possibilities are illustrated.

in Fig. 4A, and some of the experimental results are illustrated in Fig. 4, B and C. To summarize, primary transcripts appear to be initiated from two putative promoters, P1 located in front of the tRNA met gene and P2 located in front of the tRNA met gene. No transcripts could be detected from the region in front of the putative P1 promoter. Transcripts initiated from these promoter sites appear to extend through the distal tRNA gene and on into the nusG gene; rapid endonuclease processing results in the removal of the tRNA sequences from the extended leader region of these transcripts.

tRNA Processing—The results of S1 nuclease protection experiments using the 5'-end-labeled 173-nucleotide-long EcoRI-AvaI fragment, the 308-nucleotide-long AvaI fragment, the 508-nucleotide-long AvaI fragment and the 280-nucleotide-long MspI fragment are illustrated in Fig. 4B. The two protected products of the 173-nucleotide-long probe (Fig. 4Bi) are 72 and 66 nucleotides in length and correspond, respectively, to protection by (i) the primary transcript initiated at the P1 promoter and the (ii) transcript that has been processed to generate the mature 5'-end of tRNA. Clearly, the amount of mature tRNA detected is much greater than the amount of primary transcript. The two most visible protection products of the 308-nucleotide-long probe are about 200 and 300 nucleotides in length (Fig. 4Bii). These correspond, respectively, to protection by (i) the trailer sequence that is liberated following cleavage of the primary transcript at or near the 3'-end of the tRNA met sequence, and (ii) by primary transcripts or processed intermediates with a 5'-end at or near the beginning of the tRNA met sequence. The three major protection products obtained with the 508-nucleotide-long
probe were about 435, 260, and 180 nucleotides in length (Fig. 4Biii). These correspond, respectively, to protection by trailer sequences liberated by processing at or near (i) the 3'-end of the tRNA\text{met}^\text{2} gene, (ii) the 5'-end of the tRNA\text{met}^\text{2} gene, and (iii) the 3'-end of the tRNA\text{met}^\text{r} gene. Finally, using the 5'-end-labeled 280-nucleotide-long MspI fragment as probe, it was possible to demonstrate that transcripts exiting the tRNA\text{met}^\text{r} gene are extended well into the nusG gene. The observed products, 280 and 270 nucleotides in length, result, respectively, from (i) full length protection of the probe by the primary transcript, and (ii) partial protection by the trailer liberated following processing at the 3'-end of the tRNA\text{met}^\text{r} sequence (Fig. 4Bic). No other abundant transcripts with either 3'- or 5'-ends within the tRNA\text{met}^\text{r}-nusG intergenic space were detected. Thus, the five tRNA genes are processed from the leader region of the mRNA that extends into the nusG gene. These results were confirmed and extended using the corresponding 3'-end-labeled AvaI fragments and other 5'- and 3'-end-labeled MspI fragments as probes in S1 nuclease protection assays (data not shown).

Some of the 5'-transcript ends detected by S1 nuclease protection were confirmed and precisely positioned using primer extension analysis (Fig. 4C). The primer oD10 is complimentary to a sequence within the tRNA\text{met}^\text{1} gene. The major extension product terminating at the G residue at position 101 is five nucleotides in front of the tRNA\text{met}^\text{1} gene (Fig. 4Ci). This is the position where transcripts are initiated from the putative P1 promoter. Less abundant products with end sites at nucleotides 98, 99, and 106 were also apparent. The first two positions probably correspond to minor transcription initiation sites, and the third corresponds to the 5'-end of the mature tRNA\text{met}^\text{1}. It is likely that this oligonucleotide primes more efficiently on the precursor than mature tRNA.

Fig. 4. Mapping of transcript end sites in the tRNA-nusG region. A, a detailed genetic map of the 1.2-kb region is illustrated with the five tRNA and the nusG genes indicated. The positions of the two putative promoters, P1 and P2, are indicated (vertical bar with right-pointing arrow) along with restriction sites used for making S1 nuclease protection probes: E, EcoRI; M, MspI; A, AvaI. The primary transcript is depicted below the map: ±, putative transcript initiation sites; △ and ▼, respectively, the positions of detectable 5'- and 3'-ends generated during the excision and processing of the tRNA sequences. B, the structures of four 5'-end-labeled DNA fragments used as S1 protection probes are illustrated as rectangles on the left. Below each are the major protection products illustrated as lines (i–iv). The position in the nucleotide sequence (from Fig. 2) used for end labeling at the ends of the minus strand DNA probes, and the ends of the protected products are indicated in parentheses. The length of the protected products in nucleotides (n) corresponds to the visible autoradiographic bands. The autoradiograms are illustrated at the right: S, molecular length standard; T, S1 protection using T. maritima RNA. For simplicity, the controls using E. coli RNA and the DNA probe without RNA are not illustrated. C, the autoradiograms of the primer extension experiments are illustrated. The primers used were (a) oD10, complementary to position 155–136 within the tRNA\text{met}^\text{1} gene; (b) oD1, complementary to position 248–231 in the Met1-Met2 intergenic space; and (c) oD12, complementary to position 455–417 within the tRNA\text{met}^\text{2} gene. The major extension products are indicated and their position within the complementary DNA (+) strand nucleotide sequence is illustrated: †, strong stop; O, weak stop. The ladder (G, A, T, C) depicts the DNA (−) strand sequence; EP designates the lane containing the primer extension products.

The second primer oD1 is complementary to a region within the primary transcript between the mature tRNA\text{met}^\text{1} and tRNA\text{met}^\text{2} sequences. Four extension products with end sites at nucleotide positions 182, 164, 134, and 101 were detected (Fig. 4Ci). The product with an end at position 182 most probably corresponds to priming on the trailer intermediate released following endonuclease cleavage at or immediately adjacent to the 3'-end of the tRNA\text{met}^\text{1}. This result implies that endonuclease incision occurs precisely at the end of the mature tRNA sequence and that there is no extensive exonuclease trimming required to produce the mature 3'-tRNA end. Alternatively, the product may be due to termination of extension caused by secondary structure of the tRNA within the primary transcript. The next two products corresponding to reverse transcription stops at position 164 and 134 within the tRNA\text{met}^\text{1} structural sequence are presumably caused by impediments to elongation. The first site is in the Yψ loop, and the second is near the base of the descending portion of the anticodon stem. The longest product has an end corresponding to the transcription initiation site of the putative P1 promoter at position 101. The absence of detectable product with an end site corresponding to the mature tRNA 5'-end (position 106) may indicate that 3'-end processing normally precedes 5'-end processing.

The third primer oD11 is complementary to a sequence within the tRNA\text{met}^\text{2} (Fig. 4Ci). Two extension products with ends at positions 279 and 281 were evident; these correspond to the 5'-ends of transcripts initiated at the putative P2 promoter immediately in front of the tRNA\text{met}^\text{2} gene. By using other primers, it has been possible to show that endonuclease processing at the 3'-ends of the tRNA\text{met}^\text{1} and tRNA\text{met}^\text{2} appear to occur immediately adjacent to the CCA-terminal sequence; extension products resulting from priming of the Thr and Tyr trailer sequences exhibited stops at nucleotide positions 460 and 554, respectively (data not shown). In S1 nuclease pro-
tection experiments, 3'-end sites were detected in approximately the same positions.

**Characterization of Transcripts Derived from Protein Encoding Genes**—Transcripts entering the nusG gene were efficiently extended through the L11 and L1 ribosomal protein genes and into the L1-L10 intergenic space (Fig. 5A). Both nuclease S1 and primer extension assays failed to reveal significant levels of transcripts with either 5'- or 3'-ends in or between these genes (data not shown); this implies that the region between nucleotides 820 and 3300 is devoid of internal promoters, terminators and major mRNA processing sites (that upon cleavage produce transiently stable intermediates) and that the nusG, L11, and L1 cistrons are sequestered on a large polycistronic mRNA.

In contrast, both read-through transcripts and transcripts with 3'- or 5'-ends within the L1-L10 intergenic space have been identified (Fig. 5, Bi, Bii, and Ci). These 3'- and 5'-ends were not generated by an endonuclease cleavage event because the 3'-transcript end site at position 3426 is located 112 nucleotides downstream from the 5'-transcript end site at position 3314. The 5'-transcript end probably results from transcription initiation at a putative internal promoter, Pm, used to augment the expression of the downstream L10 and L12 genes. The 3'-end site at position 3426 presumably results

**Fig. 5. Characterization of transcripts from the NusG and ribosomal protein encoding genes.** A, the genetic map illustrates the position of the protein encoding genes (solid boxes). Restriction sites used to generate Si probes are: M, MspI; F, Fnu4HI; X, XbaI; P, PstI; and H, HinfI. The vertical arrows indicate the positions of putative regulatory signals on the DNA (or mRNA below): P, promoter; A, attenuator; T, terminator. B, the structures of several 5'- and 3'-end-labeled DNA probe fragments used in S1 nuclease protection assays are illustrated as rectangles. Under each probe are the protection products (lines). Nucleotide positions corresponding to 5' or 3' sites of end labeling on the minus DNA strand (in parentheses) and protected fragment lengths in nucleotides (n) for each of the probes and the corresponding protection products are indicated. The autoradiograms are illustrated below: S, molecular length standard; T, S1 protection using *T. maritima* RNA. For simplicity, the controls using *E. coli* RNA and the end-labeled DNA probe alone are not illustrated. The probes used are as follows: (i) 3'-labeled Fnu4HI-PstI; (ii) 5'-labeled XbaI-Fnu4HI; (iii) 3'-labeled MspI-MspI; (iv) 5'-labeled MspI-MspI; (v) 3'-labeled HinfI-HinfI; (vi) 5'-labeled HinfI-HinfI. C, primer extension was used to locate the transcription initiation sites for the putative L10(i), L12(ii), and β(iii) promoters. Positions of major (\(\star\)) stops on the (+) DNA strand sequence are indicated. The primers used were (i) oD15 complementary to position 3464-3447, (ii) oD16 complementary to position 4025-4008, and (iii) oD5, complementary to position 4610-4593. The ladder (G, A, T, C) depicts the DNA (−) strand sequence, and PE depicts the products of the primer extension reaction. D, total RNA was separated by electrophoresis and probed with the XbaI-SmaI fragment (nucleotide position 3567-4526) spanning the L10 and L12 genes. The fragments hybridized to 0.4- and 1.0-kb RNA and to larger RNA molecules.
from transcript attenuation. The end site is located within a poly T stretch and is preceded by overlapping sequences with inverted repeat symmetry. The results from S1 protection experiments indicate that this structure mediates the termination of about 50% of the mRNA transcripts during exponential phase growth. Together, the L1–L10 intergenic promoter and attenuator elements probably play an important role in modulating expression of the downstream genes (see below).

The L10–L12 intergenic space is only 19 nucleotides in length. In the ribosomes of E. coli and other organisms, the L12 protein is present in four copies per 50 S subunit, whereas all other proteins including L11, L1, and L10 are stoichiometric and present in single copy (Denissen, 1974; Subramanian, 1975; Hardy, 1975). In E. coli this 4-fold excess of L12 is achieved through an ill defined translational control mechanism (Downing and Dennis, 1987; Petersen, 1990). In contrast to E. coli, a major 5′-transcript end was mapped near the end of the L10 gene (position 3972–3974) and probably results from a transcription initiation event from a promoter element buried within the T. maritima L10 gene (Fig. 5, Bio and Cii). Transcripts from this promoter represent between one-third to one-half of the total L12 mRNA.

Analysis of the L12-β intergenic space indicates that few if any transcripts extending the L12 gene are extended into the downstream RNA polymerase β subunit gene (Fig. 5, Bu, Bui, and Cii). Rather, the transcripts are efficiently terminated in a poly T stretch centered around position 4446 that is preceded by a region of inverted repeat symmetry. Expression of the RNA polymerase β subunit gene requires transcript reinitiation at a downstream promoter. The 5′-ends of these reinitiated transcripts have been located by primer extension to position 4566 and 4571.

The above nuclease protection and primer extension results suggest the presence of two internal promoters within the ribosomal protein gene cluster; these are used to augment the expression of downstream genes. One of these, P1L0, is located in the L1–L10 intergenic space, and the other, P1L0, is located immediately in front of the L10–L12 intergenic space. The presence of the efficient transcription terminator immediately after the L12 gene results in the production of monoo-, bi-, and pentacistronic L12 cistron containing transcripts with lengths of about 400, 1000, and greater than 3500 nucleotides, respectively. By probing Northern RNA blots with an L10-L12 probe, the 400- and 1000-nucleotide-long transcripts along with a heterogenous large transcript were identified (Fig. 5D).

mRNA Secondary Structure and Function—Fig. 6 summarizes and contrasts the transcription patterns and regulatory features of the nusG, L11, L1, L10, L12 regions of the E. coli and T. maritima genomes. In E. coli there are three nonoverlapping transcription units: the trnA-tufB operon, the secE-nusG operon, and the L11, L1, L10, L12, β, β′-operon (An and Friesen, 1980; Downing et al., 1990; Downing and Dennis, 1987). Because of internal promoters, terminators, and attenuators, a number of different primary transcripts are produced from the ribosomal protein-RNA polymerase operon. In addition, the transcripts from all three operons contain potential endonuclease cleavage sites which increase further the number of detectable mRNA species.

The ribosomal protein mRNAs of E. coli possess well characterized translational control elements (reviewed by Lindahl and Zengel, 1986; Jinks-Robertson and Nomura, 1987). The site controlling L11 and L1 synthesis is a mimic of the 23 S rRNA binding site for protein L1 and is located immediately preceding the L11 cistron on the mRNA (Baughman and Nomura, 1983). Because of translational coupling, translation of the downstream L1 cistron is also blocked (Thomas and Nomura, 1987).

The E. coli L10 regulatory site located in the middle of the long L1–L10 intercistronic space is more complex and less well understood (Fii et al., 1980; Johnsen et al., 1982; Christiansen et al., 1984). The L10 protein (or a 1:4 L10-L12 complex) has been shown to bind to a segment of the mRNA about 100 nucleotides in length near the middle of the intercistronic space. A region of interrupted inverted repeat symmetry immediately adjacent to the L10 protein binding site has been shown to be extremely sensitive to nucleotide substitution (see Fig. 6) and is believed to be a crucial component in the on/off switching of mRNA translation; both translation defective and translation constitutive mutants have been characterized (Fii et al., 1980; Christiansen et al., 1984).

In T. maritima, the nusG gene is cotranscribed with the ribosomal protein gene cluster, and the tRNAs are processed from the leader region of this polycistronic mRNA transcript. The distal RNA polymerase subunit genes would appear to form a separate operon, transcribed from a promoter, P0, located in the L12-β intergenic space.

Potential sites relevant to those cleaved by RNaseIII and RNaseE in E. coli have not yet been identified in the T. maritima nusG-ribosomal protein transcripts. It is possible, nonetheless, that such endonuclease sites do occur and are used to trigger rapid degradation of mRNA sequences. If the products formed by an endonuclease cleavage are rapidly degraded, they would escape detection in the nuclease protection and primer extension assays used here. In many of our S1 protection experiments, autoradiographic bands of low intensity are apparent. These bands representing minor protection products with 5′- and 3′-ends falling within generally nondescript sequences probably represent transiently stable degradation intermediates and have for simplicity not been emphasized in this study.

Examination of the T. maritima nucleotide sequence of the nusG, L11, L1, L10, L12, and β genes and intergenic spaces has revealed a number of potentially important regions that could form regulatory structures within an RNA transcript (Fig. 6B). The first is in the short nusG-L11 intergenic space and forms a bipartite helical structure immediately preceding the L11 translation initiation site. The region exhibits primary sequence and secondary structural similarity to the L1 binding site within 23 S rRNA. By analogy with E. coli, this site is probably utilized to mediate translational regulation by protein L1.

A direct comparison of the L1–L10 intergenic space of E. coli and T. maritima failed to reveal any nucleotide sequence similarity. The transcription termination signal that has been identified is referred to as an attenuator, because (i) it functions at about 50% efficiency during exponential phase growth, and (ii) it possesses structural features which suggest that the termination frequency can be modulated. The second inverted repeat within this element is characteristic of eubacterial rho-independent terminators; when this structure is allowed to form in the nascent transcript, termination almost certainly occurs. The first inverted repeat within this element exhibits structural and possibly functional similarity to an E. coli repeat implicated in the on/off switching of mRNA translation (Fii et al., 1980; Christiansen et al., 1984; see Fig. 6B). If the on/off switch exists in T. maritima, it would appear to control transcript termination by either allowing or preventing the formation of the second terminator hairpin in the mRNA.

The efficient terminator in the L12-β intergenic space was identified and defined by nuclease S1 protection assays. It is

1 L. Achenbach-Richter, personal communication.
Fig. 6. Structure and features of E. coli and T. maritima RNA transcripts. A. The genomic maps and transcripts produced for E. coli (top) and T. maritima (bottom) are illustrated. The positions of promoters (P), terminators (T), and attenuators (A) are indicated. Where transcription termination is substantially less than 100% at terminators or attenuators, the (percent) read-through is indicated. The 5’-transcript ends resulting from initiation events are indicated (e-), along with sites of mRNA processing by known endonucleases (R3, RNase III; RE, RNase E) (X·X). Protein binding (PB) autogenous translational control sites are boxed on the mRNAs. B, regions of RNA secondary structure that presumably serve a regulatory function in the T. maritima mRNA are illustrated. For comparison, the putative L1 protein binding site in T. maritima 23 S rRNA is presented. Also illustrated for comparison are a portion of the L10 autogenous translational control region and the β attenuation structure from the E. coli RNA transcripts; mutational substitutions resulting in L10, L12 translation defective (>) or translation constitutive (>) phenotypes are indicated. The designation (T) is represented by U in RNA transcripts.

a typical eubacterial rho-independent terminator consisting of a single inverted repeat followed by a stretch of T residues. In E. coli, the analogous structure is more complex and functions as an attenuator, terminating approximately 80% of the transcripts that exit from the L12 gene and extend into the intergenic space (Downing and Dennis, 1991). Recent in vitro transcription experiments suggest that antitermination at this attenuator may be stimulated by E. coli NusG protein (Linn and Greenblatt, 1992). The termination frequency at this E. coli attenuator is adjustable and functions to control the level of production of the β and β’ subunits of RNA polymerase. In T. maritima, it seems that transcription of the upstream ribosomal protein genes and the downstream β and β’ RNA polymerase subunit genes are dissociated; transcription of the RNA polymerase β gene requires reinitiation at the P3 promoter which partially overlaps with the termination sequence of the ribosomal protein operon.

Transcription and Translation Initiation Signals—A total of five putative transcription initiation sites were identified by primer extension and nuclease protection assays within the tRNA, nuG, ribosomal protein, RNA polymerase gene cluster. The sequences preceding these five sites were examined in order to identify conserved features which might constitute elements of a T. maritima promoter (Fig. 7A). The following
PROMOTER ALIGNMENTS

<table>
<thead>
<tr>
<th>Protein</th>
<th>Common positions</th>
<th>Internal gaps</th>
<th>Identical amino acids</th>
</tr>
</thead>
<tbody>
<tr>
<td>NusG</td>
<td>175</td>
<td>5</td>
<td>75 (43%)</td>
</tr>
<tr>
<td>L11</td>
<td>141</td>
<td>1</td>
<td>87 (62%)</td>
</tr>
<tr>
<td>L1</td>
<td>252</td>
<td>1</td>
<td>116 (50%)</td>
</tr>
<tr>
<td>L10</td>
<td>147</td>
<td>8</td>
<td>65 (44%)</td>
</tr>
<tr>
<td>L12</td>
<td>119</td>
<td>4</td>
<td>78 (66%)</td>
</tr>
</tbody>
</table>

Features emerge: (i) All promoters exhibit one or more start sites located within a region up to eight nucleotides in length. (ii) Centered about ten nucleotides upstream from the major start site is an AT-rich sequence (TTGACA) which clearly distinguishes them from the equivalent eubacterial promoters. The T. maritima consensus derived here is TAWAAT (Hoopes and Clutter, 1987). (iii) A second conserved element possibly corresponding to the e. coli -35 sequence (TTGACA) was also identified; the T. maritima consensus was TGGAC.
state of the L10 and L12 protein was euabacterial and that the domain rearrangements and duplications now apparent in the archaeobacterial-eucaryotic proteins occurred early in the lineage leading to these now well separated domains. This proposal differs from a previous model which suggested that the archaebacterial-eucaryotic lineage leading to eubacteria (Shimmin et al., 1989).

Acknowledgments—We thank W. Zillig for introducing us to Thermotoga maritima and providing us with genomic DNA and the recombinant clone pUC-TB4, Deidre de Jong-Wong for technical assistance, and Jack Greenblatt (University of Toronto) for kindly providing polyclonal antibody against E. coli Nus G.

REFERENCES


Johnsen, M., Christiansen, T., Dennis, P., and Fils, N. (1982) EMBO J. 1, 999–1004