JOURNAL OF BACTERIOLOGY, Nov. 1992, p. 7458–7462
0021-9193/92/227458-05$02.00/0
Copyright © 1992, American Society for Microbiology
Vol. 174, No. 22

Sequence of the S-Layer Gene of *Thermus thermophilus* HB8 and Functionality of Its Promoter in *Escherichia coli*

M. M. FARALDO,† M. A. DE PEDRO, AND J. BERENGUER*

Centro de Biología Molecular, Universidad Autónoma de Madrid-Consejo Superior de Investigaciones Científicas, 28049 Madrid, Spain

Received 8 May 1992/Accepted 14 September 1992

The nucleotide sequence of the *slpA* gene, which is responsible for the synthesis of the S-layer protein of *Thermus thermophilus* HB8, is described. This gene is transcribed as a unit in which the coding region is preceded by a 127-base-long leader mRNA sequence. The promoter region is also recognized by the RNA polymerase of *Escherichia coli* because of the presence of homologous −35 and −10 boxes. Homologies with other promoters from *Thermus* spp. are also presented.

Crystalline surface layers (S-layers) are frequently found in archaebacteria (*Archaea*) and eubacteria (*Bacteria*) and are a common feature in bacteria that live under extreme environmental conditions, such as high temperature, extreme pH values, or high osmotic pressure (7, 19). Therefore, it has been suggested that S-layers have an important role in the adaptation of microorganisms to hostile environments. In addition to this, S-layers are often found in pathogenic bacteria, for which a direct involvement in virulence has been proposed (3, 9).

The extreme thermophilic eubacterium *Thermus thermophilus* HB8 (16) is surrounded by a hexagonal S-layer, whose main component is a protein (P_{100}) with an apparent molecular mass of 100 kDa (1). In the presence of Ca^{2+}, purified P_{100} forms highly thermostable trimers, which might represent the basic building blocks of the S-layer (4). Recently, we have cloned in *Escherichia coli* a 5.8-kbp DNA fragment from *T. thermophilus* HB8 that encodes proteins recognized by rabbit antiserum obtained against native P_{100} (5). The origin of these products was further revealed by a 2.9-kbp *Nde*I-*Xma*I fragment (see Fig. 1) which, when cloned in *E. coli* under the control of a T7 promoter, expressed high amounts of a 52-kDa protein that is also recognized by the antiserum. In fact, unless *lon* mutant hosts were used, it was not generally possible to detect products with electrophoretic mobilities similar to that of the S-layer protein (P_{100}). Interestingly, the deletion of 1.5 kbp from the 5′ region of the cloned fragment leads to the synthesis, under the control of the lactose promoter of plasmid pUC9, of a stable protein with a higher mass. This implies the existence of special features in the deleted domain that made the protein susceptible to degradation by the La protease, the product of the *lon* gene.

Sequence of *slpA*. The preliminary sequence of this 2.9-kbp *Nde*I-*Xma*I fragment from *T. thermophilus* HB8 revealed an open reading frame (ORF) starting at the ATG codon of the *Nde*I restriction site. Therefore, we decided to sequence the 5′ upstream region up to the *Mbo*I site, which was used for the cloning of the whole 5.8-kbp fragment into pUC9. Sequencing was carried out by the dideoxy termination method (18) on both strands of subcloned fragments, which were obtained either by restriction enzyme digestion or by controlled deletions with exonuclease III. The joining points between these fragments were sequenced with the appropriate synthetic oligonucleotides, with the original cloning plasmid used as the template. This allowed the identification of two *Pst*I sites separated by 75 bases, which were described as a single site by restriction mapping in our previous paper (5).

About 40 sequencing reactions were overlapped with the computer software package of the Computer Group from the University of Wisconsin (2), and we obtained the 3,416-bp sequence shown in Fig. 1. In this sequence, the ATG codon of the ORF located at the *Nde*I site (position 517) was found to continue for 917 triplets, yielding a predicted molecular mass of 96,004 Da. The sequence AAGGAGGTG, which is complementary to the 16S RNA from *Thermus* spp. (6), was found close to and upstream from this position, making it the most probable translation initiation site for the protein. The striking concentration of other ATG codons (8 of a total of 10) at the first part (200 triplets) of this ORF and the persistent resistance of the native protein to Edman degradation (5) made the definitive identification of the amino terminus of the protein difficult. However, the absence of an associated ribosome-binding site (RBS) sequence at the other ATG codons suggests that they are not used as alternative initiation codons.

In order to determine whether the sequenced fragment corresponds to the S-layer gene, we compared the properties predicted by computer for this sequence with those obtained by our group for purified P_{100} (4). The results of such a comparison are shown in Table 1. As can be observed, the theoretical mass (96 kDa), the isoelectric point (4, 63), and the overall amino acid composition (with the exception of amino acids that were difficult to identify after extensive acid hydrolysis) matched closely, supporting the identity between the product of this ORF and the S-layer protein. Furthermore, when this amino acid composition was compared with those of S-layer proteins from other bacteria (i.e., *Deinococcus radiodurans*, *Aeromonas salmonicida*, *Bacillus stearothermophilus*, *Bacillus brevis*, and *Clostridium thermosaccharolyticum*, etc.), a general similarity could be found, as all these proteins are weakly acidic and contain between 40 and 50% hydrophobic residues. The last and definitive probe for this identification was the recent isolation by homologous recombination of insertional mutants in this gene, which resulted in defective P_{100} proteins (12).

* Corresponding author.
† Present address: Instituto de Investigaciones Biomédicas, E-28029 Madrid, Spain.

```
      MboI
1     GATCCCGCACATATGCGCCTCAAAAGGCCACCGGGAAGGGGCCGAGAGCCCCCAAGGGGCACGCCTTCTTCCCGGTGGTT   80

81    TTCCTCTGGATACGCCCTGCCCCTCGGTTGTCCGCCCCTCGCAGGGCCGGCTTTCCCTCGGGGCTTTTCGCAGGCGGGCG  160

161   GGGCGGGCACACCCCGGCCGGCATCCGCGGATCCCTCGATCTTTCCGGCCGAAGCCGGTTATCCCCTTCGCGGGGTCCGC  240
                                                                                  XmaI
241   CACCTCGTGAGCCAATGGCCCCCTGCGCTTCCCGCCTTCCACCGACGCGCCTTTTCGGACCGCCTCCCACCTCCCCCCGG  320
                                          -35                -10       ▼
321   GGGGAGTATAACAGAAACCTTAAGGCCCGACCGCTTGACAAGGGCGCGTGAGGTTTTTACGATAGCGCCGGATGCGGGGA  400
                                                                        ---->--------
                                                                                  AvaI
401   AAAAGGGCTCCTTTTGGGGGGTTTTCCCCGCACCGGGCGGACCTGGGCGGAGAGGAAACGCGGCAACTCGCCCGTCTCGG  480
      -->              <--------------
                       RBS      NdeI
481   GTTCCCGCCCACGACCCTTAAGGAGGTGTGAGGCATATGAAGAAAAGGCTGGTCACACTGCTGGCAGGGCTCTTGACCGT  560
                                          M  K  K  R  L  V  T  L  L  A  G  L  L  T  V   15
561   CCTCTCCATGGGGTTCGGTCTGGCCCAGTTCTCCGACGTGCCCGCCGGGCACTGGGCCAAGGAGGCCGTGGAGGCCCTGG  640
16    L  S  M  G  F  G  L  A  Q  F  S  D  V  P  A  G  H  W  A  K  E  A  V  E  A  L  A   42
641   CGGCCAAGGGCATCATCCTGGGCTTCCCGGACGGCACCTTCCGGGGCAACGAGAACCTCACCCGCTACCAGGCGGCCCTC  720
43    A  K  G  I  I  L  G  F  P  D  G  T  F  R  G  N  E  N  L  T  R  Y  Q  A  A  L   68
721   CTCATCTACCGGCTCTTGCAGCAGATTGAGGAGGAGCTGAAGACCCAGGGCACCTCCCCCACCATGGAGGCCCTGGCCCC  800
69    L  I  Y  R  L  L  Q  Q  I  E  E  E  L  K  T  Q  G  T  S  P  T  M  E  A  L  A  P   95
         XhoI
801   CGAGGACCTCGAGGCGATGATCGCCGAGCTCAAGGCCCAGCCCATGCCTGAGCCCGGCATGGACCAGGCGGCCCTTAAGG  880
96    E  D  L  E  A  M  I  A  E  L  K  A  Q  P  M  P  E  P  G  M  D  Q  A  A  L  K  D  122
881   ACCTCATGGACCGGGTGGAGGCCGCCTCCATCGCCGCCGACACCGCGCTTGCCCAGGCCCAGCAGCTCGCCGAGCGGCTG  960
123   L  M  D  R  V  E  A  A  S  I  A  A  D  T  A  L  A  Q  A  Q  Q  L  A  E  R  L   148
961   GACGCCCTGGCCCAGGACGTGGAGGGCGTGAAGGGCGACCTGGCCGGGCTTAGGAGCCAGGTGGAGGCCAACGCCGACGC 1040
149   D  A  L  A  Q  D  V  E  G  V  K  G  D  L  A  G  L  R  S  Q  V  E  A  N  A  D  A   175
1041  CATCCAGGCCCTGAACGAGCTCGCCGTCCTCCTGAACCAGGACGTCCTCTCCCTCCAGGACCGGGTCACCGCCCTGGAGA 1120
176   I  Q  A  L  N  E  L  A  V  L  L  N  Q  D  V  L  S  L  Q  D  R  V  T  A  L  A  K  202
1121  AGATGGTCTCCGGCGGCCAGGAGCTCCCCGACCTGGAGCAGTTCGCCACCAAGGAGGACGTGGCCGCCGTCCAGGAGTTC 1200
203   M  V  S  G  G  Q  E  L  P  D  L  E  Q  F  A  T  K  E  D  V  A  A  V  Q  E  F   228
1201  GCCGCCGCCCTGCGCTCCGACCTGGTGGGCCTCTCCGAGAAGGTTTCCAAGCTGGAGGGCACCGTGGGCGACCTTTCTGG 1280
229   A  A  A  L  R  S  D  L  V  G  L  S  E  K  V  S  K  L  E  G  T  V  G  D  L  S  G  255
             PstI
1281  GAAGGTCGCTACCCTGCAGCGCAACGCCTTCACCATCAGCGGGAGCCTGAGCCTCAACTACAGCGTGTATCGGGCCTGGG 1360
256   K  V  A  T  L  Q  R  N  A  F  T  I  S  G  S  L  S  L  N  Y  S  V  Y  R  A  W  G  282
                 PstI
1361  GCCCCGACGCCTCGGCTGCAGGGCCGGGCACGGCCAACACTTTTGACATTGACCGCCTCTTCTCCAGCAAGTTCAGCACT 1440
283   P  D  A  S  A  A  G  P  G  T  A  N  T  F  D  I  D  R  L  F  S  S  K  F  S  T   308
1441  GGGGACGGGAACGGGAATGGCTCTGTGGGCGACGAGGCCGACCTGGGCAAGAACACGGAGGGCGTCACCAACGCCACCCT 1520
309   G  D  G  N  G  N  G  S  V  G  D  E  A  D  L  G  K  N  T  E  G  V  T  N  A  T  L  335
1521  AAGCGTCAGCTTCTCCACGGGGAAGTTGGATGCGGCTTCCGATCCGGGCAAGCTGAACAGCTACCCCGGCCTGGTCCAGT 1600
336   S  V  S  F  S  T  G  K  L  D  A  A  S  D  P  G  K  L  N  S  Y  P  G  L  V  Q  F  362
1601  TCAGCCTGCGCGCGAAGCTGACCAACCCCGGCAAATACGACCCCAGCACGGGTGCGCCCACCTACCCCATCAACCTCACC 1680
363   S  L  R  A  K  L  T  N  P  G  K  Y  D  P  S  T  G  A  P  T  Y  P  I  N  L  T   388
1681  CTGGATGAGTTTTCCTCCACGCTGGCGGTGGCCAAGGACCAGACCCTCTCCTTTAGCTTCGGTCGGTCGGTGAGGTCCAA 1760
389   L  D  E  F  S  S  T  L  A  V  A  K  D  Q  T  L  S  F  S  F  G  R  S  V  R  S  K  415
1761  GTTCACCGAGTACGTCTTTGACAACGACTACAACAGCCGCGGGCACGGTTTCGTGGCCACCTACAAGCCCGGCCTCCTGG 1840
416   F  T  E  Y  V  F  D  N  D  Y  N  S  R  G  H  G  F  V  A  T  Y  K  P  G  L  L  G  442
1841  GGCCACGCTGACCGGGGTTTACGGCTCCAAGGGGGCCAACAATGGTGACTTCACGTACTTCCGCGGCGCCCGCCTGGCC  1920
443   A  T  L  T  G  V  Y  G  S  K  G  A  N  N  G  D  F  T  Y  F  R  G  A  R  L  A   468
1921  TTGAGCCCTGTTGAAGGCATCGCCTTGGGTGGCTCCTTTGTCCAGGAGGGCCTGGACGCCAACCAAGGGACCACGAGTGC 2000
469   L  S  P  V  E  G  I  A  L  G  G  S  F  V  Q  E  G  L  D  A  N  Q  G  T  T  S  A  495
2001  CTCCTTCCCCGCTCCTACCACCGTGTACGGGGTGGATGCCAGCGTGAAGCTCGGTCCTGTGGGCCTCGCCGGTGAGTACT 2080
496   S  F  P  A  P  T  T  V  Y  G  V  D  A  S  V  K  L  G  P  V  G  L  A  G  E  Y  F  522

2081  TCAATTCCGATGCCGCTCCCAACGCGAACGGGTACTACGTGAAGGCCGACGTGGCCTTGGGGTCTATCTCCGTGGCGGGG 2160
523   N  S  D  A  A  P  N  A  N  G  Y  Y  V  K  A  D  V  A  L  G  S  I  S  V  A  G   548
2161  AACTACCGGAACATCGGCGCCGGGGTTACCGGGGCGAACATGCTCTCCGGGGACGCCACCAGCACCCTGGACCAGGGCGG 2240
549   N  Y  R  N  I  G  A  G  V  T  G  A  N  M  L  S  G  D  A  T  S  T  L  D  Q  G  G  575
2241  CTGGGGCGGCGTTGATTCCAGCGGTAACGTCATCAACGGGGCGCCCTTCCGCTCTAACCGCCAGGGCTTTGGGGTGAGCG 2320
576   W  G  G  V  D  S  S  G  N  V  I  N  G  A  P  F  R  S  N  R  Q  G  F  G  V  S  A  602
2321  CTTCCGCGGGGCTTGGACCCATCACGGTCAAGGGCTACTACGACAGCTCCACGGTCCTTGCCAACGAGACCATTACCAAC 2400
603   S  A  G  L  G  P  I  T  V  K  G  Y  Y  D  S  S  T  V  L  A  N  E  T  I  T  N   628
2401  TCCTACGGCGCCTTCAACTACAGTGCAAACAACCAGCTCGTGGCCTACGGCGGCCAGGCGGACCTCGCCTTCGGCGGCTT 2480
629   S  Y  G  A  F  N  Y  S  A  N  N  Q  L  V  A  Y  G  G  Q  A  D  L  A  F  G  G  F  655
2481  CACGCTGAGCGGCTTCTACCGCATCGCCCAGCTCAATGGCTCCACCACCCGTTACATCCTCACCGAGAAGCCCGCGGAAG 2560
656   T  L  S  G  F  Y  R  I  A  Q  L  N  G  S  T  T  R  Y  I  L  T  E  K  P  A  E  A  682
2561  CGGTGTACGCCAGCGAGTACGGGGCCAAGCTGGCCCACGATGGGGCCTCTAAGGACGCTCTGGTGCCCAAGCTCAACTTC 2640
683   V  Y  A  S  E  Y  G  A  K  L  A  H  D  G  A  S  K  D  A  L  V  P  K  L  N  F   708
                                                                       AccI
2641  ACGGCCGCTTACACCCAGAAGTACGACAACGCCACCAGCGGCTTCACCACCCAGGACATCGCCGTCTACGGTTCCTACGA 2720
709   T  A  A  Y  T  Q  K  Y  D  N  A  T  S  G  F  T  T  Q  D  I  A  V  Y  G  S  Y  E  735
2721  GCTGGCCCTCGGGCCTCTGACCCTCAAGCCCATGGGCCGGTACCACACCCAGGACGCCGCGGCGGCGAGCACTTCTTCCG 2800
736   L  A  L  G  P  L  T  L  K  P  M  G  R  Y  H  T  Q  D  A  A  A  A  S  T  S  S  D  762
2801  ACTACACCACGGTGAAGTACGGGGTGGCGGCTTCTATTGCCCTGGACCTTCCCTTCAAGCCCAGCCTTTCCGGGGAGTAC 2880
763   Y  T  T  V  K  Y  G  V  A  A  S  I  A  L  D  L  P  F  K  P  S  L  S  G  E  Y   788
2881  TACGCCCGCTCCACCCAGGTAACCTCGGCCAACAGTGGTTCTTCCGCCACGGGCACCATCTCCGAAAGCAAGTACGCTGT 2960
789   Y  A  R  S  T  Q  V  T  S  A  N  S  G  S  S  A  T  G  T  I  S  E  S  K  Y  A  V  815
2961  GGGGCTGAAGCTGGGCGAGTTCCTCTTCAAGAACTCCTCGGTGGAAGCCAAGTACGCCTCCTACACGGGCAGCGGCCTCA 3040
816   G  L  K  L  G  E  F  L  F  K  N  S  S  V  E  A  K  Y  A  S  Y  T  G  S  G  L  N. 842
3041  ACGCCCCCATTCTCCTGGGTGTCGCTGACGCCGCCAGCTCCACCACCTCCGACTACCTCTACAACAATGCGGTGAGCTCG 3120
843   A  P  I  L  L  G  V  A  D  A  A  S  S  T  T  S  D  Y  L  Y  N  N  A  V  S  S   868
3121  GTGGGGAGCAACCGGGGCAGCGTGACCGGGTGGTACTTCACCTGGACGTACTGGGACCTCACCTTCGCCTACGTGGAAGC 3200
869   V  G  S  N  R  G  S  V  T  G  W  Y  F  T  W  T  Y  W  D  L  T  F  A  Y  V  E  A  895
3201  CGACGTGAACAACAACGGCAACCAGACCCACGGTCAGGCCTTCAAGATCAGCTACACCGTAAAGTTCTAACCCCTGGCGG 3280
896   D  V  N  N  N  G  N  Q  T  H  G  Q  A  F  K  I  S  Y  T  V  K  F                917

3281  TCCCGCCCCCGCCCCCTTTGGGGGGGCGGGGGTTTTGTTCTCCCCCTTCCTTTGGGGAGGAGAGGGGCCCCTATACTT 3360
         ---------->      <----------
                                                  XmaI
3361  GGGTAAGGATGACTTCCGCTACCGGGGCCTGCACCTAAGGGGGACCAGCCCCGGG                          3416
```

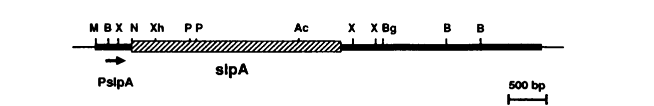

FIG. 1. Sequence of *slpA*. The proposed RBS sequence is underlined. Palindromes in the transcription promoter and terminator regions are labeled with dashed arrows. The vertical arrowhead indicates the transcriptional start point, as deduced from Fig. 2 (see text). Promoter sequences with high levels of homology to the −35 and the −10 box consensus sequences for *E. coli* are also underlined. Only those restriction sites described in the text are indicated. At the bottom of the figure is the restriction map of the 5.8-kbp DNA fragment from *T. thermophilus* in which the *slpA* gene was identified. In this, the structural part of the gene (striped) and the promoter (PslpA) (solid) are also indicated. Restriction enzymes: Ac, *Acc*I; B, *Bam*HI; Bg, *Bgl*II; M, *Mbo*I; N, *Nde*I; P, *Pst*I; X, *Xma*I, *Sma*I; XH, *Xho*I.

TABLE 1. Comparison of the properties deduced for the product of *slpA* and those of purified P_{100}

Property or component	Value	
	P_{100}	*slpA*
Molecular mass (kDa)	100	96
Isoelectric point	4.7	4.63
Amino acid composition[a]		
Alanine	16.4	12.9
Arginine	3.2	2.8
Aspartic acid	13.1[b]	5.6
Asparagine		5.1
Cysteine	ND[c]	0.0
Glutamic acid	11.9[b]	4.4
Glutamine		3.7
Glycine	5.2	10.1
Histidine	0.7	0.8
Isoleucine	2.2	2.5
Leucine	10.2	10
Lysine	4.2	4.3
Methionine	0.7	1.1
Phenylalanine	4.2	4.1
Proline	ND	3.4
Serine	8.3	9.3
Threonine	6.1	7.5
Tryptophan	ND	0.6
Tyrosine	5.6	4.8
Valine	7.3	6.4

[a] Molar fraction of each amino acid.
[b] Molar fractions of Asp plus Asn and Glu plus Gln obtained after acid hydrolysis.
[c] ND, not detected.

Analysis of the amino acid sequence and the hydropathy profile revealed the existence of a signal peptide at the amino terminus of the protein, which has been described for the other S-layer genes sequenced so far (7). This signal peptide is formed by three positively charged residues (Met-Lys-Lys-Arg) and a sequence of hydrophobic and uncharged amino acids. Following this sequence, the first charged amino acid appears at position 27 (Asp), which could indicate the end of the exportation signal. However, as mentioned above, the inability to sequence the native protein impeded the determination of whether this signal peptide is processed in the thermophile, although we have demonstrated recently that it could be used in *E. coli* to direct the secretion of other proteins (13).

Despite these similarities, a comparison between the sequences of *slpA* and those of other S-layer genes did not show significant homology, even at the level of hydropathy profiles. This finding was especially surprising for the S-layer gene (*hpl*) of *D. radiodurans*, an organism phylogenetically

related to *Thermus* spp. (21). Furthermore, there exists a striking difference in the distribution of negative charges along P_{100} with respect to other S-layer components. In P_{100}, the negative charges are concentrated within the first third of the protein, whereas in the other predicted S-layer proteins, they occur randomly or are concentrated in the carboxy-terminal domain (7).

The absence of significant homology between the S-layer genes of related bacteria, which is reinforced by the presence of tetragonal instead of hexagonal S-layers in other *Thermus* isolates (e.g., ATCC 25105 and ATCC 27737) (our unpublished results), suggests that S-layers are subjected to strong selective pressures (phages and enzymes) that result in a high frequency of amino acid changes. However, their peripheral nature forces the conservation of a similar overall composition by physicochemical constrictions.

Codon usage of *slpA*. As expected from the high G+C content of *slpA* (65.5%), which is similar to those of other *Thermus* genes, there is an extreme preference for G or C in the third position of the codons (89%). In addition, when C and G are synonymous, a preference for C over G was found (64%). The C+G content values of the first and the second base positions for *slpA* are 59.5 and 47.5%, respectively, which is similar to the codon preferences found in other genes of *Thermus* spp. (14, 15, 20) and in the gene encoding the S-layer protein of *D. radiodurans* (17).

Transcription of *slpA* in *T. thermophilus* HB8. Preliminary RNA blot experiments, in which total RNA of *T. thermophilus* HB8 was hybridized with radioactive probes derived from the *slpA* gene, demonstrated that *slpA* was transcribed as a single species of about 3 kb (data not shown). This result strongly suggested that *slpA* was transcribed as a monocistronic unit, with its own promoter and terminator. In support of this, a transcription terminator sequence was found at nucleotide position 3268 (Fig. 1), about 15 nucleotides downstream from the translational stop codon (TAA). This sequence consists of a stem-loop motif, with a predicted energy of −24.2 kCal/mol (ca. −101 kJ/mol) for the 11-base-long inverted repeat, and is followed by a T-rich stretch (8 Ts of a total of 10 bases), which is very similar to the ρ-independent terminators of *E. coli*. As no other putative terminator sequences were found at a reasonable distance downstream, we assumed that the transcription of *slpA* terminates here.

In order to localize the 5′ terminus of this mRNA, we used S1 nuclease digestion assays, with a labeled *Xma*I-*Ava*I fragment (complementary to nucleotides 321 to 480 of the positive strand) as the probe. Electrophoretic analysis of the S1 digestion mixture revealed the formation of fragments, the most intense of which corresponds to a digestion of 70 nucleotides from the 3′ terminus of the probe (Fig. 2, lane 2). These results demonstrate that the transcription of the *slpA* gene most probably starts at guanine 390, which is labeled with a vertical arrowhead in Fig. 1.

The results described above indicate that *slpA* is transcribed with a 127-base-long leader mRNA, unlike other *Thermus* genes. As we could not identify potential start codons with adequate RBSs within this transcribed upstream region, we propose that this leader region has regulatory functions. In support of this, we have identified a 13-base-long inverted repeat at the 5′ terminus of this mRNA (underlined with convergent arrows in Fig. 1), which could form a hairpin structure with a single mismatch when transcribed into RNA. Hypothetically, such structures could protect the 5′ terminus from digestion by nucleases, thereby increasing the stability and the half-life of the mRNA of one of the most abundant proteins of *T. thermophilus* HB8. A second possibility is that the inverted repeat is a binding site for a dimeric ancillary protein. This possibility is also supported by the presence within the inverted repeats of a group of five A bases followed by two groups of four T bases, separated by approximately a helix-turn distance, which could form a bend in this DNA region (11).

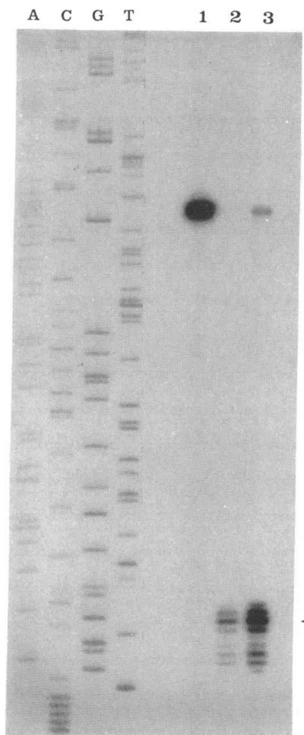

FIG. 2. Identification of the transcription origin of *slpA* in *T. thermophilus* HB8 and in *E. coli*. Total RNA preparations from *T. thermophilus* HB8 (lane 2) and from *E. coli* transformed with plasmid pMF42 (lane 3) were hybridized to a ^{32}P-end-labeled *Xma*I-*Ava*I fragment (complementary to nucleotides 321 to 480 in Fig. 1) for 16 h at 42°C. After digestion with S1 nuclease, the samples were subjected to electrophoresis in a 6% acrylamide gel. The major protected band is indicated by an arrowhead. As a size marker for the degradation of the probe, an unrelated sequence is shown (lanes A through T). Lane 1 corresponds to the undigested probe.

Transcription of *slpA* in *E. coli*. In a previous paper, we demonstrated that *slpA* could be expressed in *E. coli*, even when cloned in a promoterless vector (pMF42) (5), although the sizes of the peptides thus obtained were much smaller than that of the native P_{100}. To demonstrate the promoter activity of the 5′ region of *slpA* in *E. coli*, the *Xma*I-*Ava*I fragment used for the S1 nuclease mapping was inserted into the *Eco*RI site of the promoter-probe vector, pJAC4 (10), and the ligation mixture was plated on Luria-Bertani agar plates containing 30 μg of kanamycin (the resistance marker for the vector) per ml and 100 μg of ampicillin per ml. Under these conditions, only those clones harboring hybrid plasmids with inserts conferring promoter activity should be able to produce β-lactamase and to form colonies. Analysis of the transformants revealed three kinds of plasmids. One type bore one copy, and another carried two in tandem copies of the promoter region of *slpA*. Surprisingly, the third kind of plasmid carried the promoter region of *slpA* in an inverted orientation. The production of β-lactamase by these plas-

TABLE 2. Comparison of the promoter regions of various genes

Gene[a]	Promoter region[b]
Consensus *E. coli*	TTGACa[c] TAtAat[d] AAGGAGGTG[e]
icdh	TTACAAGGCCTCAAGCCCCGTGGTGTAGTGTAAAGGGGGCGATTCCGCCCCCGGAGGTGTGACCCATG
trpAB	TTTACCGGGAGGCCCCTCCGGGGTAGGATGGGAGTTGTCTTGGCGCGAGGCGCCTTTAGGGAGCGAAGCATG
aquaI	TTGACACCCGGGCATCCTTAGGGTTAGCTTTGCCCTCGTGAAATCCACAAAGGAGCGTATG
scs	CCGTTTCCACGGCCCAAGGCCTATATAATGCCCTTGACCGGGGTCGCCCGGCAAGGGAGGTGGGTCTTG
23S-5S RNA	TTGACAAAGGCCATGCCTCCTTGGTATCTTCCCTTTTGCGCT. . .
4.5S RNA	TAGCCTCAGGGCTTCCATGGGTGCTATACTACCCGAGCCCCC. . .
16S RNA	TTGACAAAAAGGAGGGGGATTGATAGCATGGCTTTTCTGCGCGGCGAAA. . .
slpA	TTGACAAGGGCGCGTGAGGTTTTTACGATAGCGCCGGATGCG(100 b)CCTTAAGGAGGTGTGAGGCATATG

[a] Isocitrate dehydrogenase gene (*icdh*) of *T. thermophilus* HB8; tryptophan synthetase genes (*trpAB*) of *T. thermophilus* HB27; succinyl-coenzyme A synthetase operon (*scs*) of *T. flavus* AT62; aqualysin I gene (*aquaI*) from *Thermus acquaticus*; 23S-5S, 4.5S, and 16S RNA genes; and S-layer gene (*slpA*) of *T. thermophilus* HB8.

[b] The conserved bases of the promoter motifs and the RBS, as well as the first codon (end of each sequence), are boldface. The 5′ terminus of the mRNA is underlined. 100 b, 100 bases.

[c] Underlined groups below the consensus correspond to the hypothetical −35 region.

[d] Underlined groups below the consensus correspond to the hypothetical −10 region.

[e] The RBS sequence compared is complementary to the 16S RNA of *T. thermophilus* HB8 (6). Boldface groups below the consensus are RBS.

mids was assayed as the rate of nitrocefin (Oxoid) hydrolysis, and the in tandem construction was the most active (100% reference), with the single (85%) and the inverted (40%) forms next in order.

The unexpected β-lactamase activity detected in *E. coli* cells when the promoter region of *slpA* was cloned in an inverted orientation suggested the existence of a promoter-like sequence in the minus strand. Upon further inspection, we could detect the sequence TTGTCA (complementary to nucleotides 361 to 356), which is very similar to the −35 box consensus for *E. coli* promoters. Furthermore, two putative −10 box regions were detected in the same strand at different distances (13 and 25 bases) from this, the best of which (TATACT) was identified at positions 330 to 327. These data, which explain the transcriptional activity in *E. coli* of the inverted *slpA* promoter, also suggest the existence in vivo of two overlapping and divergent promoters. However, the absence of a significant ORF in the minus band indicates that this promoter activity is not coupled to protein synthesis in *T. thermophilus* HB8.

Although these data demonstrated that the promoter region of *slpA* could function in *E. coli*, we still did not know whether both organisms used exactly the same signals. In order to determine this, we repeated the S1 nuclease protection experiment of protection against S1 digestion with total RNA obtained from *E. coli* cells transformed with plasmid pMF42. The results of this experiment, shown in lane 3 of Fig. 2, clearly demonstrate the identical sizes of the protected fragments and support the use of the same promoter signals by the RNA polymerases of *E. coli* and *T. thermophilus* HB8.

Comparison of the promoter sequence with those from other *Thermus* genes. As far as we know, there are only four genes of *Thermus* spp. for which an analysis of the transcription start points has been published (6, 15). This complicates the comparison of promoter-conserved motifs in these organisms. However, we could align the *slpA* promoter with the promoter of these genes (the *rrn* genes of *T. thermophilus* and the succinyl-coenzyme A synthetase of *Thermus flavus* AT62) and with three other genes for which the identification of −35 and −10 motifs has been proposed (*icdh*, *aqua*I, and *trpAB*) (15, 20). This alignment has been made by using a putative −10 box that shows homology to that of *E. coli* promoters as a reference point. The results, shown in Table 2, demonstrate the presence of −35 regions, identical to the consensus described for *E. coli*, in the promoter of *slpA*, as previously found (6) in other strong promoters of this thermophile (the 23S-5S and 16S RNAs and the gene for aqualysin I). Interestingly, the gene that shows the lowest homology (*scs*) in this region belongs to a very different *Thermus* isolate (8). In contrast, the −10 region was less conserved in these promoters, and only the gene of *T. flavus* showed complete homology.

When the sequences of more *Thermus* genes become available, it will be possible to generate a consensus for the promoter regions of this extreme thermophile. However, the data deduced from the comparison presented here suggest a high degree of conservation of the −35 box between strong promoters of *Thermus* spp. and the consensus for those of *E. coli*, while the −10 region could most frequently include C or G bases in its central part.

Nucleotide sequence accession number. The EMBL-GenBank-DDBJ accession number for the 3,416-bp sequence we obtained is X57333.

We are very grateful to J. Ayala for his help in computer analyses. The excellent technical assistance of J. de la Rosa is greatly appreciated.

This work has been supported by grant number BIO91-0523 from the CICYT and by an institutional grant from the Fundación Ramón Areces. M. M. Faraldo is the holder of a fellowship from the Fondo de Investigaciones de la Seguridad Social.

REFERENCES

1. **Castón, J. R., J. L. Carrascosa, M. A. de Pedro, and J. Berenguer.** 1988. Identification of a crystalline surface layer on the cell envelope of the thermophilic eubacterium *Thermus thermophilus*. FEMS Microbiol. Lett. **51**:225–230.
2. **Devereux, J., P. Haeverli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
3. **Dooley, J. S. G., and T. J. Trust.** 1987. Surface protein composition of *Aeromonas hydrophila* strains virulent for fish: identification of a surface array protein. J. Bacteriol. **170**:499–506.
4. **Faraldo, M. M., M. A. de Pedro, and J. Berenguer.** 1988. Purification, composition and Ca^{2+}-binding properties of the monomeric protein of the S-layer of *Thermus thermophilus*. FEBS Lett. **235**:117–121.
5. **Faraldo, M. M., M. A. de Pedro, and J. Berenguer.** 1991. Cloning and expression in *Escherichia coli* of the structural gene coding for the monomeric protein of the S-layer of *Thermus thermophilus* HB8. J. Bacteriol. **173**:5346–5351.
6. **Hartmann, R. K., and V. A. Erdmann.** 1989. *Thermus thermo-*

philus HB8 16S rRNA is transcribed from an isolated transcription unit. J. Bacteriol. **171:**2933–2941.

7. **Hövmoller, S., A. Sjögren, and D. N. Wang.** 1988. The structure of crystalline bacterial surface layers. Prog. Biophys. Mol. Biol. **51:**131–163.
8. **Hudson, J. A., H. W. Morgan, and R. M. Daniel.** 1989. Numerical classification of *Thermus* isolates from globally distributed hot springs. Syst. Appl. Microbiol. **11:**250–257.
9. **Ishiguro, E. E., W. W. Kay, T. Ainsworth, J. B. Chamberlain, R. A. Austen, J. T. Buckley, and T. J. Trust.** 1981. Loss of virulence during culture of *Aeromonas salmonicida* at high temperature. J. Bacteriol. **148:**333–340.
10. **Jaurin, B.** 1987. A promoter probe vector (pJAC4) that utilizes the ampC β-lactamase gene of *Escherichia coli*. Nucleic Acids Res. **15:**8567.
11. **Koo, H., and D. M. Crothers.** 1988. Calibration of DNA curvature and unified description of sequence-directed bending. Proc. Natl. Acad. Sci. USA **85:**1763–1767.
12. **Lasa, I., J. R. Castón, L. A. Fernández-Herrero, M. A. de Pedro, and J. Berenguer.** 1992. Insertional mutagenesis in the extreme thermophilic eubacteria *Thermus thermophilus* HB8. Mol. Microbiol. **6:**1555–1564.
13. **Lasa, I., M. de Grado, M. A. de Pedro, and J. Berenguer.** 1992. Development of *Thermus-Escherichia* shuttle vectors and their use for expression of the *Clostridium thermocellum celA* gene in *Thermus thermophilus*. J. Bacteriol. **174:**6424–6431.
14. **Miyazaki, K., H. Eguchi, A. Yamagishi, T. Wakagi, and T. Oshima.** 1992. Molecular cloning of the isocitrate dehydrogenase gene of an extreme thermophile, *Thermus thermophilus* HB8. Appl. Environ. Microbiol. **58:**93–98.
15. **Nishiyama, M., S. Horinouchi, and T. Beppu.** 1991. Characterization of an operon encoding succinyl-CoA synthetase and malate dehydrogenase from *Thermus flavus* AT-62 and its expression in *Escherichia coli*. Mol. Gen. Genet. **226:**1–9.
16. **Oshima, T., and K. Imahori.** 1974. Description of *Thermus thermophilus* (Yoshida and Oshima), comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. Int. J. Syst. Bacteriol. **24:**102–112.
17. **Peters, J., M. Peters, F. Lottspeich, W. Shäfer, and W. Baumeister.** 1987. Nucleotide sequence analysis of the gene encoding the *Deinococcus radiodurans* surface protein, derived amino acid sequence, and complementary protein chemical studies. J. Bacteriol. **169:**5216–5223.
18. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
19. **Sleytr, U. B., and P. Messner.** 1983. Crystalline surface layers on bacteria. Annu. Rev. Microbiol. **37:**311–339.
20. **Terada, I., S. Known, Y. Miyata, H. Matsuzawa, and T. Ohta.** 1990. Unique precursor structure of an extracellular protease, aqualysin I, with NH_2 and COOH-terminal pro-sequences and its processing in *E. coli*. J. Biol. Chem. **265:**6576–6581.
21. **Weisburg, W. G., S. J. Giovannoni, and C. R. Woese.** 1989. The *Deinococcus-Thermus* phylum and the effect of rRNA composition on phylogenetic tree construction. Syst. Appl. Microbiol. **11:**128–134.