Cloning and analysis of the *spc* ribosomal protein operon of *Bacillus subtilis*: comparison with the *spc* operon of *Escherichia coli*

T.M.Henkin*, S.H.Moon, L.C.Mattheakis¹⁺ and M.Nomura¹

Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, Shreveport, LA 71130 and ¹Department of Biological Chemistry, University of California—Irvine, Irvine, CA 92717, USA

Received June 21, 1989; Revised and Accepted August 16, 1989

EMBL accession no. X15664

ABSTRACT

A segment of *Bacillus subtilis* chromosomal DNA homologous to the *Escherichia coli* *spc* ribosomal protein operon was isolated using cloned *E. coli* *rplE* (L5) DNA as a hybridization probe. DNA sequence analysis of the *B. subtilis* cloned DNA indicated a high degree of conservation of *spc* operon ribosomal protein genes between *B. subtilis* and *E. coli*. This fragment contains DNA homologous to the promoter-proximal region of the *spc* operon, including coding sequences for ribosomal proteins L14, L24, L5, S14, and part of S8; the organization of *B. subtilis* genes in this region is identical to that found in *E. coli*. A region homologous to the *E. coli* L16, L29 and S17 genes, the last genes of the S10 operon, was located upstream from the gene for L14, the first gene in the *spc* operon. Although the ribosomal protein coding sequences showed 40-60% amino acid identity with *E. coli* sequences, we failed to find sequences which would form a structure resembling the *E. coli* target site for the S8 translational repressor, located near the beginning of the L5 coding region in *E. coli*, in this region or elsewhere in the *B. subtilis* *spc* DNA.

INTRODUCTION

The mechanisms for the regulation of ribosome biosynthesis have been partially characterized in *E. coli*, and include a negative feedback pathway to prevent accumulation of excess ribosomal components. In a number of ribosomal protein operons, regulation occurs at the posttranscriptional level, and is mediated by the binding of a specific ribosomal protein encoded in the operon to a target site on the messenger RNA (1,2). Although such mechanisms appear to be well-conserved among closely related *Enterobacteria*, including *Proteus vulgaris* and *Serratia marcescens* (3,4), it is not known how widely they are used in other bacterial species evolutionarily distant from *E. coli*. In the Gram-positive bacterium *B. subtilis*, many

ribosomal determinants have been mapped (5-9), but ribosomal protein gene organization and expression have not been well-characterized. Analysis of the mechanisms governing ribosomal gene expression in organisms distantly related to *E. coli* will aid in the evaluation of the evolutionary significance of particular regulatory mechanisms.

The *spc* operon has been analyzed in depth in *E. coli*, and contains genes encoding ribosomal proteins L14, L24, L5, S14, S8, L6, L18, S5, L30 and L15, as well as the *secY* gene and a gene originally designated "X" (now designated *rpmJ*, encoding ribosomal protein B; 10-12). Expression of the *spc* operon ribosomal protein genes is regulated posttranscriptionally by the levels of protein S8 (13,14). Interaction of S8 protein with a target site on the *spc* operon mRNA located near the beginning of the L5 coding region results in repression of the translation of L5, and of the downstream *spc* operon ribosomal protein genes by a translational coupling mechanism (4,16). The interaction of S8 with its target site on the mRNA also results in inhibition of the synthesis of the two proximal *spc* operon gene products, L14 and L24, by a "retroregulation" mechanism involving mRNA degradation (17).

In this paper, we report the isolation and characterization of a segment of *B. subtilis* chromosomal DNA homologous to the promoter-proximal region of the *E. coli* *spc* operon, and the promoter-distal region of the S10 operon. Our results indicate that although the structure and organization of the *B. subtilis* *spc* operon closely resembles that of *E. coli*, we find no sequences which would form a structure resembling the S8 regulatory target site on *E. coli* *spc* mRNA.

MATERIALS AND METHODS

Cloning techniques.

Chromosomal DNA was isolated from *B. subtilis* strain 168 (*trpC2*), obtained from G. Chambliss. Restriction endonucleases and DNA modifying enzymes were obtained from New England Biolabs (Beverly, MA) and were used as described by the manufacturer. Chromosomal DNA digests were analyzed by Southern hybridization, using standard techniques (18), except that the wash temperature

was reduced to 41⁰C. The *E. coli* DNA probe was the 0.49 kb HindIII-SalI fragment from plasmid pNO2749 (4), which contains most of the coding sequences for protein L5, and includes the target site for S8. The fragment was labeled by the method of Feinberg and Vogelstein (19). In HindIII-digested *B. subtilis* DNA, a single band approximately 2.5 kb in size was found to hybridize to the probe. DNA fragments of this size class were purified by preparative agarose gel electrophoresis and ligated with HindIII-digested M13 mp18 DNA (20). Plaques containing the insert were identified by hybridization with the same probe. Subclones containing smaller fragments were generated using various restriction enzymes and bacteriophage M13 vectors. *E. coli* strains JM103 and JM105 were used as the host strains (20,21). *B. subtilis* promoter-probe *cat* fusion plasmid pCPP-3 (22) was obtained from the *Bacillus* Genetic Stock Center, Ohio State University.

DNA sequencing.

DNA sequence analysis was carried out by dideoxynucleotide sequencing as described (21), using a universal synthetic oligonucleotide primer (New England Biolabs) or specific oligonucleotide primers synthesized on a Biosearch oligonucleotide synthesizer or purchased from The DNA Factory (San Diego, CA). When necessary, sequence data was confirmed using 7-deaza-dGTP (ABN, Hayward, CA) to reduce secondary structure in G+C-rich regions. DNA sequence data was analyzed using the Pustell sequence analysis system (International Biotechnologies, Inc., New Haven, CT).

RESULTS

Cloning of *B. subtilis* *spc* operon DNA.

A 2.4 kb HindIII fragment of *B. subtilis* DNA was found by Southern hybridization analysis to be homologous to an *E. coli* DNA probe containing the L5 coding region of the *spc* operon. This fragment was cloned into M13 mp18, to generate phage M13NO18171; positive clones were identified by hybridization to the *E. coli* *spc* DNA, under moderate stringency conditions. Initial DNA sequence analysis from the HindIII site adjacent to M13 priming site revealed that this region of the DNA corresponded to the amino-terminal region of the S8 coding sequence,

Figure 1. Restriction map of the *B. subtilis* S10-*spc* operon region. Restriction sites are: H, HindIII; B, BamHI; M, MluI; E, EcoRI; N, NsiI; Hp, HpaI; X, XbaI; and P, PvuII. Open reading frames identified as ribosomal protein coding regions are indicated below the restriction map. Regions corresponding to the *E. coli* S10 and *spc* operons are labeled on the bottom line.

indicated by correspondence to the published amino-terminal protein sequence of *B. subtilis* S8 (23). A restriction map of the *B. subtilis* *spc* region is shown in Figure 1. Restriction sites identified by DNA sequencing were confirmed by Southern hybridization analysis of *B. subtilis* chromosomal DNA, using nick-translated M13N018171 replicative form DNA as the probe (data not shown).

DNA sequence analysis.

The sequence of the 2.4 kb DNA fragment was determined, and is shown in Figure 2. Protein coding regions were identified by comparison with published amino-terminal amino acid sequences for *B. subtilis* 30S ribosomal proteins (23), amino acid sequences for *B. stearothermophilus* ribosomal proteins (24,25), and *E. coli* ribosomal protein gene sequences (10,26,27). The deduced amino acid sequence perfectly matched the *B. subtilis* amino-terminal amino acid sequences except for a Thr for Gly substitution at position 24 of protein S17, and Asn-Met for Asp-Asn at positions 18-19 of protein S8. It is possible that these differences reflect genetic variation between the strains of *B. subtilis* used.

For each putative ribosomal protein gene, a sequence closely resembling a *B. subtilis* ribosome binding site (GGAGG; 28) was located at an appropriate position upstream of the initiator ATG codon, although protein S14 appears to utilize a GTG initiator codon. The protein coding regions generally terminated in a TAA

AAGCTTCTTGGATTACTAACCGTCAAATCGAAGCTGCACGTATTGCGATGACTCGTTACA 60
 A S W I T N R Q I E A A R I A M T R Y
 L16...
 TGAAACGTGGCGGTAAAGTTTGGATCAAAATTTTCCCTTCTAAACCTTACACTGCTAAAC 120
 M K R G G K V W I K I F P S K P Y T A K
 CACTTGAGGTACGTATGGGATCCGGTAAAGGGGCTCCAGAAGGTTGGGTAGCTGTTGTAA 180
 P L E V R M G S G K G A P E G W V A V V
 AACCGGGCAAAGTTTATTTGAAATTTCTGGTGTGTCTGAAGAGGTTGCTCGTGAAGCGC 240
 K P G K V L F E I S G V S E E V A R E A
 TTCGCCTTGCATCTCACAAATTGCCAATTAAAACGAAGTTTCGTAAAACGTGAAGAAATTG 300
 L R L A S H K L P I K T K F V K R E E I
 GTGGTGAATCAAATGAAAGCTAATGAAATTCGTGACCTTACCACTGCTGAAATTGAACAA 360
 G G E S N E S - (end L16)
 (L29 start) M K A N E I R D L T T A E I E Q
 AAAGTAAAGTCTCTTAAAGAAGAACTTTTCAATCTTCGCTTTCAATTAGCGACAGGACAA 420
 K V K S L K E E L F N L R F Q L A T G Q
 CTTGAAAATACTGCTCGCATTCGTGAAGTGCGCAAAGCTATCGCGCGCATGAAAACCTGTG 480
 L E N T A R I R E V R K A I A R M K T V
 ATTCGTGAAAGAGAAATTGCTGCTAATAATAATCACTGAGGGGAGGTCGTGAGCATGAG 540
 I R E R E I A A N K - M S
 end L29 S17 start
 CGAACGTAAACCAACGCAAAGTTTACCAAGGCCGTGTTGTTTCTGACAAAATGGATAAAAC 600
 E R N Q R K V Y Q G R V V S D K M D K T
 CATTACTGTTGTTGTTGAGACATACAAAAACATACACTTTACGGTAAACGCGTAAAATA 660
 I T V V V E T Y K K H T L Y G K R V K Y
 CTCAAAAAATTCAAAGCACATGATGAAAATAATCAAGCTAAAATCGGTGACATCGTTAA 720
 S K K F K A H D E N N Q A K I G D I V K
 -35 -35
 GATCATGGAAACTCGTCCATTGTCTGCAACTAAGCGTTTCCGTCTAGTTGAAGTTGTCTGA 780
 I M E T R P L S A T K R F R L V E V V E
 -10 -10
 AGAAGCTGTTATTATCTAATAAAGTTCGGAATCTTTTTCCGAAGGGAGGTAACAAATAA 840
 E A V I I -
 end S17
 TGATTCAACAAGAGACTCGTTTTAAAAGTAGCTGACAACTCTGGCGCACGTGAAGTACTTA 900
 M I Q Q E T R L K V A D N S G A R E V L
 L14 start
 CTATTAAAGTTCTTGGTGGTTCTGGACGTAAAACAGCTAACATTGGTGATGTAATTGTTT 960
 T I K V L G G S G R K T A N I G D V I V
 GCACGGTTAAACAAGCAACACCAGGAGGCGTTGTCAAAAAGGTGAAGTCGTAAAAGCTG 1020
 C T V K Q A T P G G V V K K G E V V K A
 TTATCGTTCGCACTAAAAGCGGAGCGCGCAGAAGTGACGGTTCATACATCAGCTTCGATG 1080
 V I V R T K S G A R R S D G S Y I S F D

AGAATGCATGTGTTATCATCCGTGACGACAAGAGCCACGCGGAACCTCGTATCTTCGGAC 1140
E N A C V I I R D D K S P R G T R I F G

CAGTTGCTCGTGAATTGCGTGAAAACACTTTATGAAAATTGTTTCTCTAGCTCCAGAAG 1200
P V A R E L R E N N F M K I V S L A P E

TTATCTAATTGAAATAAATGCCTTACTCAAGGTGGTGCATCAGGATGCATGTAAAAAAA 1260
V I - M H V K K
end L14 L24 start
GGCGATAAAGTTATGGTTATCTCTGGTAAAGATAAAGGCAAAACAAGGAACAATCCTTGCT 1320
G D K V M V I S G K D K G K Q G T I L A

GCTTTCCTTAAAAAGGACCGCGTTTATAGTTGAAGGTGTTACATGGTAAAGAAACACTCT 1380
A F P K K D R V L V E G V N M V K K H S

AAACCAACTCAAGCTAACCTCAAGGCGGTATTTCTAATCAAGAGGCGCCAATTCATGTA 1440
K P T Q A N P Q G G I S N Q E A P I H V

TCAAACGTTATGCCGCTCGATCTAAAAACAGGTGAAGTGACTCGCGTAGGATACAAAGTG 1500
S N V M P L D P K T G E V T R V G Y K V

GAAGATGGCAAAAAAGTTTCGTGTAGCAAAAAATCTGGGCAAGTTCTAGATAAATAGTAT 1560
E D G K K V R V A K K S G Q V L D K -
end L24

TAAAAGGAAGGGAGGTCTATCTCATGAACCGCCTTAAAGAAAAGTACAATAAAGAAATTG 1620
M N R L K E K Y N K E I
L5 start
CACCTGCTTTAATGACTAAGTTCAACTATGATTTCAGTCATGCAAGTGCCTAAAAATCGAAA 1680
A P A L M T K F N Y D S V M Q V P K I E

AAATCGTAATCAACATGGGTGTTGGTGACGCTGTTCAAAACGCCAAAGCAATCGACAGTG 1740
K I V I N M G V G D A V Q N A K A I D S

CTGTTGAGGAATTAACGTTTATCGCAGGTCAAAAACCTGTCTGTTACTCGTGCGAAGAAAT 1800
A V E E L T F I A G Q K P V V T R A K K

CAATTGCTGGATTCCGTCTTCGTGAGGGAATGCCTATCGGTGCGAAAGTAACTCTTCGCG 1860
S I A G F R L R E G M P I G A K V T L R

GAGAGCGCATGTATGATTTTCCTTGATAAACTTATTTCTGTATCTTTACCGCGTGACGTG 1920
G E R M Y D F L D K L I S V S L P R V R

ACTTCCGCGGGGTTTCTAAAAAATCTTTTCGACGGTTCGCGGTAACACACTTGGTATCA 1980
D F R G V S K K S F D G R G N Y T L G I

AAGAACAGTTAATCTTCCCTGAAATTGACTACGATAAAGTAACAAAGGTTTCGCGGAATGG 2040
K E Q L I F P E I D Y D K V T K V R G M

ACATCGTTATCGTAACAACTGCTAATACTGACGAAGAAGCTCGTGAGCTATTAACCTCAAG 2100
D I V I V T T A N T D E E A R E L L T Q

TAGGTATGCCGTTCCAGAAATAATCAATGAAAGGGAGGCGAAATCGTGGCTAAAAAGTCT 2160
V G M P F Q K - V A K K S
end L5 S14 start

```

ATGATTGCGAAACAACAACGTACACCAAAGTTTAAAGTACAAGAGTACACACGCTGTGAA 2220
M I A K Q Q R T P K F K V Q E Y T R C E

CGCTGCGGACGTCCGCATTTCAGTCATTCGTAAATTTAAACTTTGCCGTATTTGTTTCCGT 2280
R C G R P H S V I R K F K L C R I C F R

GAGCTTGCATATAAAGGACAAATTCCTGGCGTGAAAAAGCCAGCTGGTAATCCCCAAGA 2340
E L A Y K G Q I P G V K K A S W -
                                end S14
AAGGGAAGGAGGTAATTATATAATGGTTATGACAGATCCAATTGCAGATATGCTAACTCG 2400
                                M V M T D P I A D M L T R
                                S8 start
TATTTCGCAATGCAAACATGGTACGTCATGAGAAGCTT 2437
I R N A N M V R H E K L

```

Figure 2. DNA sequence of the *B. subtilis* S10-*spc* operon region. Translation of ribosomal protein coding regions are indicated below the DNA sequence, with predicted coding region borders indicated below the amino acid sequence. Putative ribosome binding sites upstream of protein start points are underlined. Sequences resembling the consensus sequences for *B. subtilis* vegetative promoters are labeled "-35" and "-10" above the DNA sequence.

termination codon, as is also common in *E. coli* ribosomal protein genes. The distances between the termination codon of the preceding gene and the initiation codon of the subsequent gene were generally longer than in the translationally coupled *E. coli* *spc* operon.

The organization of ribosomal protein genes in this region proved to be highly similar to that of *E. coli*. The order of *spc* operon genes was found to be L14, L24, L5, S14 and S8, in accordance with the order observed in *E. coli*; the order of genes in the remainder of the *B. subtilis* *spc* operon is also in agreement with the order found in *E. coli* (J.-W. Suh and C. W. Price, personal communication). In *E. coli*, the *spc* operon is located immediately downstream from the S10 operon (10); in *B. subtilis*, sequences homologous to proteins L16, L29 and S17, the last three genes of the S10 operon (26,27), were found upstream of the L14 coding sequence.

In *E. coli*, a transcriptional terminator for the S10 operon and a promoter for the *spc* operon are located between the end of the S10 operon and the beginning of the *spc* operon (10,27). S1

mapping analysis of *E. coli* mRNA indicated that both the terminator and promoter function in exponentially growing cells, but there is a definite read-through of the terminator so that about 20% of S10 operon mRNA is physically contiguous to *spc* mRNA (17; F. Sor and M. Nomura, unpublished experiments). The distance between the end of the S17 coding sequence and the beginning of the L14 coding sequence is 165 bp in *E. coli*, while the corresponding region is only 40 bp in length in *B. subtilis*. No sequence resembling a rho-independent terminator, such as is found in the *E. coli* S10-*spc* region, was identified in the *B. subtilis* sequence, suggesting the possibility that transcription of the *spc* region may originate upstream, perhaps at the promoter for the S10 operon. Inspection of the DNA sequence in the region upstream of the L14 coding region (positions 745-780) revealed two overlapping sequences resembling *B. subtilis* vegetative promoter sequences, the consensus for which is identical to the *E. coli* promoter consensus (TTGACA at -35, TATAAT at -10, 16-18 bp spacing; 29). Both putative promoter sequences showed 4/6 adherence to the -35 consensus, including the most highly conserved TTG positions, and 5/6 and 4/6 adherence to the -10 consensus, as well as 3/5 adherence to a sequence in the -16 region (RTRTG) found to contribute to promoter strength in *B. subtilis* vegetative promoters (30). To test whether this region of the DNA has promoter activity, the 0.8 kb HindIII-EcoRI DNA fragment containing the putative promoter sequences was inserted into the promoter-probe *cat* fusion plasmid pCPP-3 (22), from which chloramphenicol acetyltransferase is expressed when a promoter-containing fragment is inserted in the appropriate orientation. This fusion plasmid was found to confer resistance to chloramphenicol in both *E. coli* and *B. subtilis* (data not shown), indicating that this segment of the DNA does in fact have promoter activity *in vivo*. Characterization of the *in vivo* start point of *spc* transcription is in progress.

Sequence homology between *B. subtilis* and *E. coli*.

Overall nucleotide sequence homology was approximately 50%, while certain regions showed very high homology. In general, regions corresponding to protein coding sequences were highly homologous (60% sequence identity), while noncoding intergenic

Figure 3. Secondary structure models of the S8 target site region. (a) S8 target site on *E. coli* *spc* mRNA; (b) S8 binding site on *E. coli* 16S rRNA. Arrows indicate nucleotide changes in corresponding regions of *B. subtilis*. The region of *B. subtilis* *spc* mRNA shown is from position 1555 to 1682. The *E. coli* rRNA fragment protected by S8 from pancreatic RNase digestion (42) is delineated by a broken line in (b). Sequences shared by both the S8 target site on *E. coli* mRNA and the S8 binding site on *E. coli* rRNA are boxed. A slightly different base-pairing scheme was proposed for the region with the boxed sequences by Mougél et al. (36; see also ref. 4); in this structure, three A residues, A957, A959, and A1031, of mRNA (a), and the corresponding three residues of rRNA (b), are bulged out from a base-paired stem. The numbering of nucleotide positions shown for *E. coli* *spc* mRNA is according to ref. 4, and that shown for *E. coli* 16S rRNA is according to ref. 43. The UAA stop codon for L24 is at position 934-936, and the AUG start site for L5 is at position 951-953 in *E. coli* *spc* mRNA. For *B. subtilis* *spc* mRNA, the UAG stop codon for L24 is at position 1555-1557, and the AUG start site for L5 is at position 1584-1586.

regions were not homologous and differed in length as well as sequence (for example, see the region between the L24 and L5 genes shown in Figure 3a). Of particular interest was the absence of sequences homologous to the region shown for *E. coli* to comprise the regulatory target site for S8 binding to the *spc* mRNA. This subject will be treated in detail in the Discussion section.

Ribosomal protein sequence homologies.

Alignment of the predicted amino acid sequences of the *spc* and S10 ribosomal proteins for *B. subtilis* and *E. coli* is shown in Figure 4. Comparisons with *B. stearothermophilus* protein sequences are also shown for those proteins for which amino acid sequence data is available. The percentage of identical amino acids between *B. subtilis* and *E. coli* was as high as 60%, and 88% for *B. subtilis* vs. *B. stearothermophilus*; consideration of conservative amino acid substitutions raises the homology to as high as 80% for protein L5 for *B. subtilis* vs. *E. coli*. Protein S14 showed lower homology because of the presence of a large deletion/ substitution in the central portion of protein, while the amino-terminal and carboxyl-terminal regions were conserved. Since the *B. stearothermophilus* S14 sequence shows the same deletion/substitution relative to the *E. coli* sequence, it appears that this structural difference in protein S14 is characteristic of *Bacillus* species.

DISCUSSION

In this paper we report the isolation and characterization of a segment of the *B. subtilis* genome corresponding to the promoter-proximal region of the *E. coli* *spc* ribosomal protein operon. A high degree of conservation of ribosomal protein gene structure and organization was found, despite the evolutionary distance between *E. coli* and *B. subtilis*. Sequences homologous to the promoter-distal region of the S10 ribosomal protein operon were found immediately preceding *spc* operon sequences, while alpha operon sequences are located downstream from the *spc* operon (31), indicating that the order of ribosomal protein operons in the major ribosomal protein gene cluster in *B. subtilis* (S10-*spc*-alpha) is identical to that of *E. coli*. Classi-

		Homology (%)
BSU L16:	..ASWITNRQIEAARIAMTRYMKRGGKVWIKIFPSKPYTAKPLEVRMGSGKGAP 	
ECO L16:	..RGRLTARQIEAARRAMTRAVKRQGIWIRVFPDKPITEKPLAVRMCKGKGNV 	
BSU L16:	EGWVAVVKGPKVLFEISGVSEEVAREALRLASHKLPKTKFKVKREEIGGESNES 	59%
ECO L16:	EYWVALIQPGKVLYEMDGVPEELAREAFKLAACKLPKTTTFVKTVM 	
BST L29:	MKAKEIRELTAEIEQKIKALKEELFNLRFLATGQLENTARIRQVRKDI 	
BSU L29:	MKANEIRDLTAEIEQKVKSLKEELFNLRFLATGQLENTARIREVRKAI 	
ECO L29:	MKAKELREKSVEELNTELLNLLREQFNLRMQAASGQLQSHLLKQVRDV 	
BST L29:	ARMKTIIREREIAANK 	88%
BSU L29:	ARMKTVIREREIAANK 	38%
ECO L29:	ARVKTLLNEKAGA 	
BST S17:	-SQRNQRKVYVGRVVDKMDKXXXVLV.. 	
BSU S17:	MSERNQRKVYQGRVVDKMDKTITVVVETYYKHTLYGKRVKYSKKFKAHD 	
ECO S17:	MTDKI-RTL-QGRVVDKMEKSIVVAIERFVKHPYIGKFIKRTTKVHVHD 	
BSU S17:	ENNAQIGDIVKIMETRPLSATKFRFLVEVVEEAVII 	53%
ECO S17:	ENNECGIGDVVEIRECRPLSKTKSWTLVRVVEKAVL 	
BST L14:	MIQQESRLKVADNSGAREVLVIKVLGGSGRRYANIGDVVVATVKDADPGG 	
BSU L14:	MIQQETRLKVADNSGAREVLTIKVLGGSGRKTANIGDVIVCTVKQATPGG 	
ECO L14:	MIQEQTMLNVADNSGARRVMCIKVLGGSHRRYAGVGDIKIITIKEAIPRG 	
BST L14:	VVKKGQVVKAVVVRTKRGVRRPDGSYIRFDENACVIIIRDDKS-PRGTRIF 	
BSU L14:	VVKKGEVVKAVIVRTKSGARRSDGSYISFDENACVIIIRDDKS-PRGTRIF 	
ECO L14:	KVKKGDVLKAVVVRTKKGVRRPDGSVIRFDGNACVLLNNNSEQPIGTRIF 	
BST L14:	GPVAREKRDKDFMKIISLAPEVI 	84%
BSU L14:	GPVARELRENNFMKIVSLAPEVI 	63%
ECO L14:	GPVTRELRSKFMKIIISLAPRVL 	

```

BST L24:  MHVKKGDKVQVI--SGKDKGKQGVILAAFPKKNRVIVEGVNIVKKHAKPSQAN
          ||||| || ||||| ||||| || ||||| ||||| |||||
BSU L24:  MHVKKGDKVMVI--SGKDKGKQGTILAAFPKKDRVLVEGVNMVKKHKSPTQAN
          || || ||||| ||||| ||||| ||||| ||||| |||||
ECO L24:  MAAKIRRDDEVIVLTGKDKGKRGKVKNVLSSG-KVIVEGINLVKKHQKVPVAL

BST L24:  PQ-GGIEKEAPIHVSVMPLDPKTGEPTRIGYKIVDGKKVRYAKKSGEILDK
          || || ||||| ||||| ||||| ||||| ||||| ||||| 83%
BSU L24:  PQ-GGISNQEAPIHVSVMPLDPKTGEVTRVGKVEDGKKVRVAKKSGQVLDK
          || || ||||| ||||| ||||| ||||| ||||| ||||| 47%
ECO L24:  NQPGGEVEKEAAIQVSNVAIFNAATGKADRVGFRFEDGKKVRFFK-SNSETIK

BSU L5:   MNRLKEKYNKEIAPALMTKFNYDSVMQVPKIEKIVINMGVDVAQNAKAI
          || || || ||||| ||||| ||||| ||||| |||||
ECO L5:   MAKLHDYYKDEVVKLMTFNYNSVMQVPRVEKITLNMVGGEAIADKKLL

BSU L5:   DSAVEELTFIAGQKPVVTRAKSIAQFRLREGMPIGAKVTLRGERMYDFL
          || || ||||| ||||| ||||| ||||| ||||| |||||
ECO L5:   DNAAADLAAISGQKPLITKARKSVAGFKIRQGYPIGCKVTLRGERMWEFF

BSU L5:   DKLISVSLPRVRDFRGVSKKSFDRGNITLGIKEQLIFPEIDYDKVTKVR
          || || ||||| ||||| ||||| ||||| ||||| |||||
ECO L5:   ERLITIAVPRIRDFRGLSAKSFDRGNYSMVGREQIIFPEIDYDKVDRVR

BSU L5:   GMDIVIVTTANTDEEARELLTQVGMPPFQK
          || || ||||| ||||| ||||| ||||| 60%
ECO L5:   GLDITITTTAKSDEEGRALLAAFDFFPRK

BST S14:  -AKKSMA-----[KQKRTPKFKVRAYT]..
          ||||| ||||| ||||| |||||
BSU S14:  MAKKSMA-----[KQRTPKFKVQEYT]-----
          ||||| ||||| ||||| |||||
ECO S14:  MAKQSMKAAREVKRVALADKYFAKRAELKAIADVNASDEDRWNAVLKIQ

BSU S14:  -----RCERCGRPHSVIRKFKLCRICFRELAYKGQIPGVKKASW
          || ||||| ||||| ||||| ||||| |||||
ECO S14:  TLPRDSSPSRQRNRCRQTGRPHGFLRKFGLSRIKVREAAMRGEIPGLKKG 44%

BST S8:   -VMTDPIADMLTRIRN..
          ||||| ||||| ||||| |||||
BSU S8:   MVMTDPIADMLTRIRNANMVRHEKL..
          || ||||| ||||| ||||| |||||
ECO S8:   MSMQDPIADMLTRIRNGQAANKAAV.. 56%

```

Figure 4. Sequence comparisons of S10-*spc* ribosomal proteins. Amino acid sequences for ribosomal proteins from *B. subtilis* (BSU), *B. stearothermophilus* (BST) and *E. coli* (ECO) are shown. Spaces (-) were introduced to maximize the alignment. Vertical lines indicate identical amino acids. Incomplete sequences are indicated by "...". "X" indicates positions in the *B. stearothermophilus* amino acid sequence which were ambiguous (24,25).

cal genetic analysis of ribosomal protein mutations mapping in this gene cluster, located at 12° on a 360° map of the *B. subtilis* genome (32), agree with this gene order, and indicate that *str* operon genes are located upstream of S10 (5), as is also the case in *E. coli* (2). In addition, the localization of mutations conferring resistance to the antibiotics rifampicin and streptolydigin upstream of *str* operon genes (5,7) indicates that the *rif* operon equivalent is also located in the major ribosomal gene cluster in *B. subtilis*, while the *rif* operon is located elsewhere in the *E. coli* genome (1,2).

The ribosomal proteins characterized in this study show a high degree of structural relatedness to their *E. coli* homologs, ranging from 38-63% amino acid identity. One of the proteins exhibiting the lowest homology is protein S14, which is conserved at its amino and carboxy termini, but diverges in the central region. The *B. subtilis* S14 protein closely resembles the *B. stearothermophilus* protein (24). It should be noted that *E. coli* S14 is essential for *in vitro* reconstitution of functionally active 30S subunits and that the *B. stearothermophilus* S14 homolog can replace *E. coli* S14 to form completely active subunits as judged by poly(U)-dependent poly(phenylalanine) synthesis (33). Thus, the large reduction in size and the sequence alteration in the central portion of S14 does not appear to affect the ability of this protein to interact with 16S rRNA (and/or other ribosomal proteins). The sequence of the protein S14 homolog from rice chloroplasts was recently reported (34), and shows a structure resembling the *E. coli* form of the protein, suggesting a closer evolutionary relationship of ribosomal protein genes between chloroplasts and Gram-negative bacteria than Gram-positive bacteria. Further studies on ribosomal gene structure and organization will provide added information for such comparisons.

In *E. coli*, expression of the *spc* operon is regulated by binding of protein S8 to a target site on the mRNA located between nucleotide positions 949 and 1039 (see Figure 3a), at the beginning of the L5 coding sequence (4,35). The secondary

structure of the mRNA in this region was previously deduced from in vitro studies using structure-specific nucleases, sequence analysis of the genes from three other Enterobacteria species (which revealed several compensatory nucleotide alterations), and site-directed mutagenesis experiments (4). The mutagenesis experiments led to the conclusion that specific nucleotide residues located between two helical regions (helices III and IV in Figure 3a) are necessary for interaction with S8, and that the two helical regions are important presumably to maintain the proper orientation of these nucleotide residues (4). The structure of the S8 target site on spc operon mRNA was highly conserved among the four Enterobacteria species analyzed (E. coli, Salmonella typhimurium, Proteus vulgaris and Serratia marcescens), and is similar in structure to the S8 binding site on 16S rRNA (4; see Figure 3). We therefore examined whether B. subtilis spc mRNA sequences would also form a structure similar to the S8 target site on E. coli spc mRNA, as shown in Figure 3a. While the sequence forming the right half of the proposed E. coli mRNA helices (III and IV in Figure 3a) is conserved, the sequence forming the left half of the helices is not well-conserved, and no structure similar to that proposed for the E. coli S8 target site is predicted to be formed. In addition, two of the three A residues proposed to be crucial for S8 recognition, A957, A959 and A1031 (4,36; see the legend to Figure 3), are altered in the B. subtilis sequence. We searched for sequences which would form a structure similar to the E. coli S8 target site in other regions of the B. subtilis spc operon sequence (determined in this work) but failed to find them.

Although the binding site for protein S8 on B. subtilis 16S rRNA has not yet been characterized, inspection of the published rRNA sequence in this region (37) indicates that the structure identified as the S8 binding site on E. coli 16S rRNA is conserved in B. subtilis (Figure 3b); there is 80% nucleotide identity in this region, and none of the substitutions occurring in helix regions alters the base pairing scheme proposed for E. coli 16S rRNA. Nucleotide sequences shared by E. coli 16S rRNA and the S8 target site on E. coli spc mRNA (4; boxed in Figure 3), which include the three A residues important for S8

recognition (36; see above), are conserved in B. subtilis 16S rRNA. The conservation of the S8 binding site on 16S rRNA is consistent with previous observations that B. stearothermophilus S8, which is almost certainly very similar to B. subtilis S8 (see Figure 4), can replace E. coli S8 in the reconstitution of functionally active 30S subunits (33), indicating a specific interaction between B. stearothermophilus S8 and E. coli 16S rRNA. Functional interaction between E. coli S8 and B. stearothermophilus 16S rRNA must also take place in view of the results obtained in earlier heterologous reconstitution studies (38). In summary, while essential elements in the S8 binding site on 16S rRNA are well-conserved among diverse species of bacteria including E. coli and B. subtilis, the structure of the S8 binding site on E. coli spc mRNA is not found on B. subtilis spc mRNA. These results suggest the possibility that regulation of the spc operon in B. subtilis occurs by a mechanism different from that found in E. coli.

The alpha operon in E. coli contains the genes for ribosomal proteins S13, S11, S4 and L17 and RNA polymerase subunit α (in the order S13, S11, S4, α , L17); S4 acts as a translational repressor regulating the synthesis of S13, S11, S4 and L17 (14,39,40). In B. subtilis, the homologous genes are located in the same order distal to the spc operon (31,41), except that the gene for S4 is located away from this gene cluster in a separate region of the chromosome (8). Thus, as discussed by Price and co-workers (31), the regulation of the alpha operon genes in B. subtilis may also be different from that found in E. coli.

It appears that evolution of the translational feedback regulation system for control of ribosomal protein gene expression found in E. coli is fairly recent, probably occurring after the separation of Bacillus ancestors from Enterobacteria ancestors. It is quite remarkable to note that in the mRNA structure shown in Figure 3a, there are only a few nucleotide alterations in the region covering the right half of the large stem-loop structure (positions 999-1037), while there are extensive nucleotide changes in the upstream left half, which is also within the L5 protein coding sequence of the mRNA. It seems likely that the S8 target site and the extensive secondary structure seen in the

present day Enterobacteria species arose by repeated mutational tinkering of this upstream L5 coding region, starting from an ancestral gene common for both E. coli and B. subtilis.

The translational feedback mechanism found in E. coli appears to be an efficient means for preventing wasteful over-production of ribosomal proteins. Since ribosomal protein genes of B. subtilis corresponding to two E. coli ribosomal protein operons are apparently not regulated by the same mechanism as in E. coli, it will be of great interest to determine the mechanism for production of ribosomal proteins stoichiometrically with rRNA in B. subtilis.

ACKNOWLEDGEMENTS

We would like to thank Brian Glass and Lynn Pittman-Cooley for technical assistance, Dr. Frank Grundy for assistance with DNA sequence analysis, and Dr. John Keaner for critical reading of the manuscript. This work was supported by National Institutes of Health grants R29 GM40650 (T.H.) and R01 GM35959 (M.N.), National Science Foundation grant DMB 8543776 (M.N.), and American Cancer Society Junior Faculty Research Award JFRA-226 (T.H.).

*To whom correspondence should be addressed.

+Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115, USA

REFERENCES

1. Lindahl, L., and Zengel, J. (1986) Ann. Rev. Genet. **20**,297-326.
2. Jinks-Robertson, S., and Nomura, M. (1987) In Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology. Ed. Neidhardt, F.C. American Society for Microbiology, Washington, D.C. pp. 1358-1385.
3. Sor, F., and Nomura, M. (1987) Mol. Gen. Genet. **210**,52-59.
4. Cerretti, D.P., Mattheakis, L., Kearney, K.R., Vu, L., and Nomura, M. (1988) J. Mol. Biol. **204**,309-329.
5. Dabbs, E.R. (1983) Mol. Gen. Genet. **192**,124-130.
6. Dabbs, E.R. (1983) Mol. Gen. Genet. **192**,386-390.
7. Dabbs, E.R. (1984) J. Bacteriol. **159**,770-772.
8. Henkin, T.M., and Chambliss, G.H. (1984) Mol. Gen. Genet. **193**,364-369.
9. Smith, I. (1982) In The Molecular Biology of the Bacilli, Vol. 1. Ed. Dubnau, D. Academic Press, N.Y. pp. 111-145.
10. Cerretti, D.P., Dean, D., Davis, G.R., Bedwell, D.M., and

-
- Nomura, M. (1983) *Nucleic Acids Res.* **11**,2599-2616.
11. Wada, A., and Sako, T. (1987) *J. Biochem.* **101**,817-820.
 12. Ueiguichi, C., Wittekind, M., Nomura, M., Akiyama, Y., and Ito, K. (1989) *Mol. Gen. Genet.* **217**,1-5.
 13. Dean, D., Yates, J.L., and Nomura, M. (1981) *Nature* **289**,89-91.
 14. Yates, J.L., Arfsten, A.E., and Nomura, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**,1837-1841.
 15. Olins, P.O., and Nomura, M. (1981) *Nucleic Acids Res.* **9**,1757-1764.
 16. Mattheakis, L.C., and Nomura, M. (1988) *J. Bacteriol.* **170**,4484-4492.
 17. Mattheakis, L., Vu, L., Sor, F., and Nomura, M. (1989) *Proc. Natl. Acad. Sci. USA* **86**,448-452.
 18. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 19. Feinberg, A.P., and Vogelstein, B. (1983) *Anal. Biochem.* **132**,6-13.
 20. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**,103-119.
 21. Messing, J. (1983) *Methods Enzymol.* **101**,20-78.
 22. Band, L., Yansura, D.G., and Henner, D.J. (1983) *Gene* **26**,313-315.
 23. Higo, K., Otaka, E., and Osawa, S. (1982) *Mol. Gen. Genet.* **185**,239-244.
 24. Hori, H., Higo, K., and Osawa, S. (1977) *J. Mol. Evol.* **9**,191-201.
 25. Kimura, M., Kimura, J., and Ashman, K. (1985) *Eur. J. Biochem.* **150**,491-497.
 26. Zurawski, G., and Zuraski, S.M. (1985) *Nucleic Acids Res.* **13**,4521-4526.
 27. Post, L.E., Arfsten, A.E., Reusser, F., and Nomura, M. (1978) *Cell* **15**,215-229.
 28. Murray, C.L., and Rabinowitz, J.C. (1982) *In Molecular Cloning and Gene Regulation in Bacilli*. ED. Ganesan, A., Hoch, J., and Chang, S. Academic Press, NY. pp. 271-285.
 29. Graves, M.C., and Rabinowitz, J.C. (1986) *J. Biol. Chem.* **261**,11409-11415.
 30. Henkin, T.M., and Sonenshein, A.L. (1987) *Mol. Gen. Genet.* **209**,467-474.
 31. Boylan, S.A., Suh, J.-W., Thomas, S.M., and Price, C.W. (1989) *J. Bacteriol.* **171**,2553-2562.
 32. Piggot, P.J., and Hoch, J.A. (1985) *Microbiol. Rev.* **49**,158-179.
 33. Higo, K., Held, W., Kahan, L., and Nomura, M. (1973) *Proc. Natl. Acad. Sci. USA* **70**,944-948.
 34. Cote, J.-C., and Wu, R. (1989) *Nucleic Acids Res.* **17**,1780.
 35. Gregory, R.J., Cahill, P.B.F., Thurlow, D.L., and Zimmerman, R.A. (1988) *J. Mol. Biol.* **204**,295-307.
 36. Mougél, M., Eyermann, F., Westhof, E., Romby, P., Expert-Bezançon, A., Ebel, J.-P., Ehresmann, B., and Ehresmann, C. (1987) *J. Mol. Biol.* **198**,91-107.
 37. Green, C.J., Stewart, G.C., Hollis, M.A., Vold, B.S., and Bott, K.F. (1985) *Gene* **37**,261-266.
 38. Nomura, M., Traub, P., and Bechmann, H. (1968) *Nature* **219**,793-799.
-

39. Dean, D., and Nomura, M. (1980) *Proc. Natl. Acad. Sci. USA* **77**,3590-3594.
40. Thomas, M.S., Bedwell, D.M., and Nomura, M. (1987) *J. Mol. Biol.* **196**,333-345.
41. Suh, J.-W., Boylan, S.A., and Price, C.W. (1986). *J. Bacteriol.* **168**,65-71.
42. Zimmermann, R.A., Mackie, G.A., Muto, A., Garret, R.A., Ungeweckell, E., Ehresmann, E., Stiegler, P., Ebel, J.-P., and Fellner, P. (1975) *Nucleic Acids Res.* **2**,279-302.
43. Woese, C.R., Gutel, R.R., Gupta, R., and Noller, H.F. (1983) *Microbiol. Rev.* **47**,621-669.
44. Ferrari, F.A., Trach, T., and Hoch, J.A. (1985) *J. Bacteriol.* **161**,556-562.
45. Moriya, S., Ogasawara, N., and Yoshikawa, H. (1985) *Nucleic Acids Res.* **13**,2251-2265.