Characterization of the *Bacillus subtilis* *rpsD* Regulatory Target Site

FRANK J. GRUNDY† AND TINA M. HENKIN†*

Department of Biochemistry and Molecular Biology, Louisiana State University Medical Center, Shreveport, Louisiana 71130

Received 8 June 1992/Accepted 17 August 1992

The *Bacillus subtilis* *rpsD* gene, which encodes ribosomal protein S4, is subject to autogenous regulation. Repression of *rpsD* expression by excess S4 protein was previously shown to be affected by mutations in the leader region of the gene. A large number of deletion and point mutations in the leader region were generated, and their effect on repression by S4 in vivo was tested. These studies indicated that the required region was within positions +30 to +190 relative to the transcription start point. Replacement of the *rpsD* promoter with a *lac* promoter derivative which is expressed in *B. subtilis* had no effect, indicating that repression by S4 occurs at a level subsequent to transcription initiation. The *rpsD* leader region was isolated from several *Bacillus* species. Members of the *B. subtilis* group, as defined by analysis of 16S rRNA sequence, contained a leader region target site very closely related in structure to that of *B. subtilis*, despite considerable primary sequence variation; the *B. brevis* *rpsD* leader contained some but not all of the structural features found in the regulatory target sites of the other *Bacillus* species. Very little similarity to the *Escherichia coli* α operon S4 target site was found at either the primary-sequence or the secondary-structure level. Mutagenic and phylogenetic data indicate that the secondary structure of the leader region regulatory target site contains two large stem-loop domains. The first of these helices has a side loop which is essential for autoregulation, is highly conserved among *Bacillus* *rpsD* genes, and is similar to a region of 16S rRNA important in S4 binding.

Little information concerning the mechanisms responsible for control of ribosome biosynthesis in gram-positive bacteria is currently available. The *Bacillus subtilis* *rpsD* gene, which encodes ribosomal protein S4, was chosen for initial analysis because it was found to be located in a region of the chromosome distant from the major cluster of ribosomal protein genes. In *Escherichia coli*, *rpsD* is in the α operon, which contains the genes for ribosomal proteins S13, S11, S4, and L17 and for the α subunit of RNA polymerase (2). S4 acts as a translational repressor to regulate α operon expression, binding to a target site at the start of the S13 coding sequence (6, 18, 25). In *B. subtilis*, the genes for S13, S11, α , and L17 are located within a large transcriptional unit in the major ribosomal gene cluster at 12° on the 360° map (3), while *rpsD* is located at 263° (14) and is monocistronic (10). It was therefore of interest to determine whether this difference in gene organization was associated with a difference in gene regulation.

Transcription of the *B. subtilis* *rpsD* gene initiates at a near-consensus σ^A -type promoter sequence approximately 180 bp upstream of the start of the S4 coding sequence (10). Previous studies demonstrated that expression of an *rpsD-lacZ* translational fusion is repressed in the presence of excess S4 (11). Deletions and point mutations in the leader region which result in loss of repression by S4 were identified, indicating that the leader sequence contains the target site for autogenous regulation. This target region exhibits a number of self-complementary sequences, leading to a number of possible structural models, including models in which the leader forms a pseudoknot-like structure similar in basic form to the model for the *E. coli* α operon S4 target site (25,

26). Elements conserved with sequences within the region of 16S rRNA with which S4 is proposed to interact (24, 28) were also identified.

In an effort to elucidate the basic structure of the *B. subtilis* *rpsD* regulatory target site and to identify sequence and structural elements critical for autogenous regulation, we have employed phylogenetic analysis of the *rpsD* gene isolated from several *Bacillus* species. We have also expanded the mutagenic analysis of the *B. subtilis* *rpsD* leader region. We find that the sequences required for repression by S4 are located within positions +30 to +190 of the mRNA and that this region is unlikely to form a pseudoknot-like structure in *Bacillus* spp. that is analogous to the *E. coli* α operon pseudoknot.

MATERIALS AND METHODS

Bacterial strains and growth conditions. The bacterial strains used in this study are listed in Table 1. *Bacillus* isolates were normally grown at 37°C in Luria-Bertani medium (21) containing the appropriate antibiotics when necessary; *Bacillus stearothermophilus* was grown at 65°C in Luria-Bertani medium supplemented with glucose (0.25%, wt/vol), FeCl₃ (20 ng/ml), and CaCl₂ (200 μ M). *B. subtilis* BR151C1 (normal levels of S4) and BR151P1 (S4 overproducing) were previously described (11). *E. coli* JM103 and DH5 α were used for propagation of bacteriophage M13 and plasmids, respectively; strain RZ1032 was used for preparation of uracil-containing M13 template DNA for oligonucleotide mutagenesis. Antibiotics were used at the following concentrations: ampicillin, 50 μ g/ml; chloramphenicol, 5 μ g/ml; neomycin, 5 μ g/ml; erythromycin, 1 μ g/ml; and lincomycin, 25 μ g/ml. Chloramphenicol and erythromycin were used at 0.1 μ g/ml for induction of the *cat* and *erm* genes, respectively. All antibiotics were purchased from Sigma Chemical Co. β -Galactosidase production was de-

* Corresponding author.

† Present address: Department of Biochemistry and Molecular Biology, Albany Medical College, New Scotland Ave., Albany, NY 12208.

TABLE 1. Bacterial strains

Strain	Genotype	Source or reference
<i>B. subtilis</i>		
BR151C1	<i>lys-3 trpC2 amyE::M13C1neo</i>	11
BR151P1	<i>lys-3 trpC2 amyE::M13P1neo</i>	11
ZB307A	<i>SPβc2del2::Tn917::pSK10Δ6</i>	P. Zuber (30)
ZB449	<i>trpC2 pheA1 abrB703</i> , SPβ cured	P. Zuber (30)
<i>B. megaterium</i> 7A1	Prototroph	BGSC ^a
<i>B. licheniformis</i> 5A1	Prototroph	BGSC
<i>B. stearothermophilus</i> 9A2	Prototroph	BGSC
<i>B. circulans</i> 16A1	Prototroph	BGSC
<i>B. brevis</i> ATCC 8185	Prototroph	N. Pace
<i>E. coli</i>		
JM103	<i>endA1 supE44 sbcBC thi-1 rpsL Δ(lac-pro)/F' traD36 lacI^qZΔM15 proAB</i>	20
DH5α	<i>φ80dlacZΔM15 endA1 recA1 supE44 hsdR17 (r_K⁻ m_K⁺) thi-1 gyrA96 relA1 Δ(lacZYA-argF)U169</i>	Bethesda Research Laboratories
RZ1032	<i>thi-1 relA1 spoT1 dut-1 ung-1 zbd-279::Tn10 supE44 HfrKL16 Po/45</i>	T. Kunkel (19)

^a *Bacillus* Genetic Stock Center, The Ohio State University.

ected by the addition of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside; Gold Biotechnologies) to solid medium at a concentration of 40 μg/ml.

Plasmids and bacteriophage. Construction of plasmid pTMH123, which contains a 480-bp *Hind*III fragment of *rpsD* including the promoter and amino-terminal S4 coding sequence fused in frame to the β-galactosidase gene in plasmid pZΔ329, has been described elsewhere (11). Sequence alterations, most of which were designed to create or destroy restriction endonuclease cleavage sites, were introduced by oligonucleotide-directed mutagenesis of this region carried in phage M13 and confirmed by DNA sequencing. The altered DNA was then introduced into plasmid pZΔ329 to generate variants of plasmid pTMH123. The leader region sequence and altered positions are shown in Fig. 1; the specific alterations present in each fusion construct are listed in Table 2. Deletion derivatives were constructed from plasmids containing new restriction sites.

Plasmids pTMH154 and pTMH155 were constructed by

FIG. 1. *B. subtilis* *rpsD* mRNA leader region sequence. Arrows indicate positions altered by oligonucleotide-directed mutagenesis; alterations are shown below the arrows. The sequence alterations present in each mutant are listed in Table 2. Regions exhibiting sequence similarity to 16S rRNA are underlined; relevant 16S rRNA sequences are shown above the line. The ribosome-binding site (AAGGAGG and AUG) is underlined.

replacing the *rpsD* promoter regions of plasmids pTMH133 and pTMH125, respectively, with a fragment derived from plasmid pTMH35 which contains the *P_{lac-1015}* variant of the *E. coli* *lacUV5* promoter (Fig. 2B); this variant, which contains changes in the -16 region of the promoter, is expressed 29-fold more efficiently in *B. subtilis* than is *P_{lacUV5}* (15). In these plasmids, the *P_{lac-1015}* fragment includes sequences to an *Xba*I site 60 bp downstream from the transcription start site fused to position +30 (plasmid pTMH154) or position +90 (plasmid pTMH155) of the *rpsD* leader. For construction of plasmid pTMH152, *Xba*I sites were introduced near the +1 positions of *P_{lac-1015}* and *P_{rpsD}* so that the resulting plasmid contains a near-perfect fusion of the *lac* promoter to the *rpsD* leader with only one less nucleotide at the 5' end of the resulting transcript.

The *rpsD-lacZ* fusions were introduced into *B. subtilis* specialized transducing phage SPβ by homologous recombination of the plasmid with SPβc2del2::Tn917::pSK10Δ6 by using strain ZB307A and selection for the plasmid-carried *cat* (chloramphenicol resistance) determinant. Phage carrying the fusion were purified by passage through strain ZB449 (SPβ cured) and were introduced into strains BR151C1 and BR151P1 for measurement of β-galactosidase activity under normal or S4-overproducing conditions, as previously described (11). These strains contain an extra copy of the *rpsD* gene inserted at the *amyE* locus; strain BR151C1 has an extra copy of the wild-type allele and contains normal levels of S4 protein, while strain BR151P1 has a derepressed version of *rpsD* inserted at *amyE* and accumulates excess S4 protein (11).

DNA manipulations. Restriction endonucleases and DNA-modifying enzymes were purchased from New England BioLabs and used as recommended by the manufacturer. Dideoxynucleotide sequencing was carried out with Sequenase (United States Biochemical) and single-stranded M13 or double-stranded plasmid templates. DNA sequence data were analyzed with the PC/Gene system (Intelligenetics, Inc.). *Bacillus* chromosomal DNA isolation and transformation were carried out as previously described (14).

M13 template DNA preparation and oligonucleotide-directed mutagenesis were done as described by Kunkel (19). Synthetic oligonucleotides were purchased from Genosys Biotechnologies (Woodlands, Tex.) or The DNA Factory

TABLE 2. β -Galactosidase activity of *rpsD-lacZ* fusions

Plasmid carrying <i>rpsD-lacZ</i> fusion	Leader region mutation	β -Galactosidase activity ^a in host strain		Repression ratio (CI/P1) ^b	Regulation ^c
		BR151C1	BR151P1		
pTMH123	Wild type	730	330	2.2	+
pTMH125	Xba-1 (G86U, G89A)	1,900	2,100	0.90	—
pTMH127	Xba-2 (A129C, G130U, C133A)	900	1,100	0.82	—
pTMH129	Δ Xba1-Xba2 (+90 to +130)	290	280	1.0	—
pTMH131	Pst-1 (G71C, U74C, A76G)	2,200	2,200	1.0	—
pTMH133	Xba-3 (G29U, Δ U32, C33A, U34G)	1,300	530	2.5	+
pTMH136	Δ Xba4-Xba3 (+1 to +30)	1,100	490	2.2	+
pTMH137	Δ Xba3-Xba2 (+30 to +130)	1,900	2,000	0.95	—
pTMH138	Pst-RC (G71C, U74C)	1,500	1,200	1.3	±
pTMH139	Pst-RG (U74C, A76G)	1,400	940	1.5	±
pTMH140	Pst-RD (U74C)	1,200	900	1.3	±
pTMH141	Sac-1 (G94U, A95C)	2,100	1,900	1.1	—
pTMH142	Sau-1 (U78G, C79A)	2,000	1,900	1.1	—
pTMH143	Taq-1 (U80G, C81A)	1,700	1,200	1.4	±
pTMH144	R3S (U188G, C189A)	780	380	2.1	+
pTMH145	R3S-Sac-1	1,100	990	1.1	—
pTMH146	Sau-1-Sac-1	1,900	1,800	1.1	—
pTMH147	Taq-1-Sac-1	1,900	2,000	0.95	—
pTMH148	R3C (U188A, C189U)	570	260	2.2	+
pTMH149	Sac1R (G94U)	1,100	710	1.5	±
pTMH150	Xba-4 (U2A) ^d	380	190	2.0	+
pTMH151	Δ Sph-Hind (+190 to +350)	220	96	2.3	+
pTMH152	<i>P</i> _{lacX} -Xba4	170	89	1.9	+
pTMH154	<i>P</i> _{lac} -Xba3	170	170	1.0	—
pTMH155	<i>P</i> _{lac} -Xba1	190	190	1.0	—
pTMH156	Nru-1 (G90U, A91C)	1,400	1,400	1.0	—
pTMH157	Sau-1-Nru-1	340	260	1.3	±
pTMH158	Sal-1 (A127C)	820	310	2.6	+
pTMH159	Xba-5 (A126U, A127C)	700	340	2.1	+
pTMH160	A126U	980	430	2.3	+

^a β -Galactosidase specific activity was determined according to the method of Miller (21). All samples were assayed in duplicate in at least two separate experiments; values varied no more than 10%.

^b Ratio of β -galactosidase activity obtained in a normal strain (BR151C1) to that obtained in a strain overproducing S4 protein (BR151P1). A ratio of 1.0 indicates loss of sensitivity of the *rpsD-lacZ* fusion to repression by excess S4.

^c +, repression ratio similar to that of the wild-type fusion (pTMH123); —, complete loss of repression by excess S4; ±, partial derepression.

^d Xba-4 mutation includes three additional changes upstream of the transcription start point; see Fig. 2B.

(San Diego, Calif.). Mutants were screened for acquisition or loss of restriction sites where appropriate, and alterations were confirmed by DNA sequencing.

Isolation of the *rpsD* leader region from other *Bacillus* species. Oligonucleotide RC2 (5'-ATACCAAGACGACGAGAAAGTTTCCAAGAAGGACCAGTGTAAACGAGCCAT-

3'; 41 of 50 match to *B. subtilis rpsD*), based on the *B. subtilis* amino-terminal S4 coding sequence, was used previously for isolation of the *B. subtilis rpsD* gene (10). Oligonucleotide BST1 (5'-ATTTTCCACATCGGGCCCGTGTAGCGCGCCAT-3') was based on the amino-terminal *B. stearothermophilus* S4 protein sequence (17). These oligo-

FIG. 2. Deletion analysis. (A) Sequences present in each *rpsD-lacZ* fusion. Dark box, *P*_{*rpsD*}; line, leader mRNA; open box, S4 amino-terminal coding sequence; hatched box, β -galactosidase coding sequence; dotted box, *P*_{*lac-1015*}; +, wild-type repression by S4; —, loss of repression by S4. Leader region sequence coordinates are shown at the bottom. (B) Construction of *P*_{*lac-rpsD*} fusion construct (plasmid pTMH152). Arrows indicate positions altered to introduce *Xba*I sites at the transcription start points of *P*_{*rpsD*} and *P*_{*lac-1015*}.

nucleotides were 5' end labeled with [32 P]ATP and T4 polynucleotide kinase and used in Southern hybridization analyses of digests of chromosomal DNA from various *Bacillus* species for identification of convenient restriction fragments containing the *rpsD* region. Appropriate DNA preparations were size fractionated, and M13 libraries were screened by plaque hybridization (29).

β -Galactosidase assays. Cells were grown to mid-logarithmic growth phase in Luria-Bertani medium containing the appropriate antibiotics. Samples (1 ml) were harvested, permeabilized with toluene, and assayed as described by Miller (21).

Nucleotide sequence accession numbers. The sequences reported in this study have been deposited in the GenBank and EMBL data bases under accession numbers M34718 (*B. subtilis*), X58866 (*B. stearothermophilus*), M99040 (*Bacillus brevis*), M99041 (*Bacillus circulans*), M99042 (*Bacillus megaterium*), and M99043 (*Bacillus licheniformis*).

RESULTS

Deletion analysis of the *B. subtilis rpsD* leader. Constructs containing various portions of the *B. subtilis rpsD* leader fused to *lacZ* were inserted into specialized transducing phage SP β and transferred into strains BR151C1 (containing normal levels of protein S4) and BR151P1 (containing high levels of protein S4). It has previously been shown that under these conditions, expression of an intact *rpsD-lacZ* fusion is repressed, while expression of a fusion with alterations in the *rpsD* regulatory target site is unaffected by the presence of excess S4 (11). The results of β -galactosidase measurements are shown in Table 2 and summarized in Fig. 2A. The wild-type fusion (pTMH123) exhibited 2.2-fold repression in the S4-overproducing host strain. Deletions in the central portion of the leader (pTMH137 and pTMH129, with leader regions +30 to +130 and +90 to +130, respectively, deleted; Fig. 2A) resulted in loss of repression by high levels of S4 protein, as previously shown (11). Deletion of the 5' portion of the leader sequence (pTMH136, positions +1 to +30) had no effect on repression by S4, indicating that the sequence in this region is not required.

With an intact leader region, deletion of the amino-terminal S4 coding region (as in pTMH151) had no effect on repression by S4 (Fig. 2A, Table 2), although the level of β -galactosidase observed was lower than in the wild-type fusion; the amino-terminal S4 sequences in the wild-type *rpsD-lacZ* may serve to stabilize the fusion protein, resulting in higher activity levels. These studies indicate that leader sequences from +30 to +190 relative to the start point of transcription are sufficient for the response to S4. This is in contrast to the S4 target site in the *E. coli* α operon, where sequences in the S13 coding region are required (6, 25).

Replacement of the *rpsD* promoter. A possible role for the *rpsD* promoter sequence in autogenous regulation was tested by using constructs in which the *rpsD* promoter was replaced by a derivative of the *E. coli lacUV5* promoter, *P_{lac-1015}*, which is capable of expression in *B. subtilis*. In the first constructs, pTMH154 and pTMH155, the *P_{lac}* fragment included approximately 60 bp of downstream *lac* sequences. These constructs, in which the fusion point in *rpsD* was at +30 or +90, resulted in loss of repression by S4 (Table 2, Fig. 2A). Since the leader sequences from +1 to +30 were shown above to be unnecessary for regulation, the possibility that the loss of repression in pTMH154 was caused by interference by the *P_{lac}* downstream sequences was explored. Short regions within the *P_{lac}* downstream sequences

BSU:	GUGUGAAAUUUGCAGCCUUUUUGUUCAGCUUUAUUGUUUUAUUCUUUAUC---	57
BMG:	A.....U.....	57
BLC:	A.....U.....	56
BST:	A.....G.....CU.....G.....CG.....	56
BCR:	A.....G.....A.....G.....CU.....G.....C.....U.....C.....AAUUAU	57
BBV:	AA..UG.....A.....U.....AGC.....AGAA.....CG.....CAA.....C.....GU.....	48
BSU:	-AAU---AAGGUGUAUCGUGUAA-CUCUCU-----GCUGCUGGAGCGAGG	97
BMG:	-C-----	97
BLC:	-U---A.....U.....	97
BST:	---CGCG.....C.....U.....GA.....	97
BCR:	U.....UUG.U.....U.G.....C.....UG.....A.....A.....A.....	101
BBV:	---AUGUC-----GA.C.....G.....A.....UAUACUCCUUGUGUCA..UAA.....A.....U---	98
BSU:	AUACAUGAAAAACAAAUUGUGCAUGGUCGAAUAG--AGCAGA---CGGUUUUUUUUCCA	152
BMG:	A.....	152
BLC:	GA.....CU.....U.....A.....UG.....	151
BST:	G.....GC.....U.....U.....UUG.....GC.....CAU.....U.....C.....	151
BCR:	G.....G.....G.....AAU.....AAAA.....AU.....UU.....CAU.....U.....A.....C	155
BBV:	AGC.....GGU.....AA.....UCGAG.....U.....CAA.....C.....U---	144
BSU:	CAAAAAUAAACCC--AAAGGAGGAGUCACAUUUGGCUUGCUAUACA	197
BMG:		99
BLC:	G.....GU.....ACGA.....U.....U	91
BST:	-GGG.....G.....GU.....ACGA.....U.....U	73
BCR:	A.....AUG.....A.....U.....U	198
BBV:	UA.....UUU.....CAU.....A.....U.....A.....U.....C.....U	186

FIG. 3. Alignment of *rpsD* leader regions from *Bacillus* species. BSU, *B. subtilis*; BMG, *B. megaterium*; BLC, *B. licheniformis*; BST, *B. stearothermophilus*; BCR, *B. circulans*; BBV, *B. brevis*. Dots indicate bases identical to those in the *B. subtilis* sequence; dashes represent spaces inserted to maximize alignment. The Shine-Dalgarno sequence and AUG start codon are underlined.

that are complementary to sequences within the *rpsD* target site were found, but no obvious alternate structure potentially responsible for loss of repression in pTMH154 was identified. Plasmid pTMH152 was constructed by creation of *Xba*I restriction sites close to the +1 positions of both *P_{lac-1015}* and *P_{rpsD}* (Fig. 2B) so that the resulting *P_{lac-rpsD}* construct contained only one less nucleotide at the 5' end of the leader region and no *lac* mRNA sequences. This fusion exhibited normal regulation, indicating that the *rpsD* promoter is not required for autogenous regulation and that repression by S4 occurs post-transcription initiation.

Phylogenetic analysis of *rpsD*. Analysis of the *rpsD* regions from several different *Bacillus* species was used to provide information about evolutionary conservation of features of the leader target sequence. *B. megaterium* and *B. licheniformis* were chosen because they were predicted to be fairly closely related to *B. subtilis*; *B. brevis* was chosen because it was predicted to be distant from *B. subtilis*, on the basis of 16S rRNA sequence data; *B. stearothermophilus* was chosen because it is a thermophile, with an optimum growth temperature of 60 to 70°C; and *B. circulans* was selected without prior knowledge of phylogenetic relatedness to *B. subtilis*, although it has recently been shown to fall within the *B. subtilis* group (1, 23).

The sequences of the entire *rpsD* genes from *B. megaterium* and *B. stearothermophilus* were determined (12). The *B. megaterium* sequence was nearly identical to the *B. subtilis* sequence throughout the entire *rpsD* gene; however, sequences upstream and downstream of the gene exhibited considerable variation. The *B. brevis rpsD* clone was nearly complete, extending to a position only three amino acids from the predicted end of the coding sequence (12). The *B. licheniformis* and *B. circulans* clones contained only the promoter, leader, and amino-terminal S4 coding regions.

Alignment of the *Bacillus rpsD* leader regions is shown in Fig. 3. The *B. megaterium* sequence exhibited only two nucleotide substitutions in the entire region compared with *B. subtilis rpsD*. The *B. licheniformis rpsD* leader was 91% conserved compared with the *B. subtilis* leader. All but four variable sequence positions in this sequence were among those altered in the *B. stearothermophilus* and *B. circulans*

FIG. 4. Sequence variations in *Bacillus rpsD* leader regions. (A) *B. subtilis* sequence from positions +33 to +173, with arrows indicating changes in *B. circulans* (uppercase letters) and *B. stearothermophilus* (lowercase letters). Double bars indicate positions exhibiting covariance. (B) Structural model of the *B. brevis rpsD* leader. BCR, *B. circulans*; BST, *B. stearothermophilus*.

leaders, which were both approximately 70% conserved compared with the *B. subtilis* sequence. Certain regions of the leader sequences were very highly conserved, while others were more variable.

Each of the *rpsD* leader sequences from members of the *B. subtilis* group exhibited all of the structural features proposed for the *B. subtilis rpsD* leader despite considerable primary sequence variation. The sequence alterations found in the *B. stearothermophilus* and *B. circulans* leaders in comparison to the *B. subtilis* structure are shown in Fig. 4A. The side stem-loop of the first large stem (+51 to +64) was especially variable, but the basic structure was always present. The side bulge at positions +71 to +76 is a region previously noted as being conserved with positions 505 to 510 of 16S rRNA (10, 11) within a region proposed to interact with protein S4 (24, 28); this region exhibits very little variability in bacterial 16S rRNA sequences (8, 13). This bulge is also very well conserved in the *rpsD* leaders, with a single U-to-G alteration at position +72 in the *B. circulans* sequence; this position is a G in the *B. subtilis* 16S sequence (9). The proposed stem sequences were well conserved and exhibited covariance of residues on opposite sides of the helix at several positions (C42-G115, U67-A100, U78-A91, G138-C165, and G139-C164), providing strong evidence of a base-paired structure in these regions. In the second large stem (positions +135 to +165), the *B. subtilis* sequence contains an unpaired U residue at position +144; this residue is absent in the *B. stearothermophilus* sequence, while the *B. circulans* sequence contains an additional A residue at a position directly opposite this position. Thus, the stem is predicted to be more stable in the last two cases.

The regions separating the two large stem-loop structures were highly variable in sequence but similar in length (20 to 22 bases). All of the sequences contained an element related to the sequence originally noted as being conserved between the *B. subtilis* leader and the 16S rRNA (+125 to +130 in *B. subtilis rpsD*); however, this element was shown to be

unnecessary for regulation by the mutagenetic analysis described below.

The *B. circulans* sequence was especially useful in distinguishing among alternate base-pairing possibilities at the top of the first helix. The U78C and A91G variations support the model shown, where residues 77 to 80 are base paired with residues 89 to 92, and fail to support alternate models. In addition, the G96A variation is inconsistent with base pairing of residues 94 to 97 with residues 70 and 77 to 79, which was possible in the *B. subtilis* and *B. stearothermophilus* sequences.

The region downstream of the AUG start codon was conserved in all of the sequences except that of *B. brevis*; however, mutational analysis indicated that the sequence in this region is not important for regulation and is unlikely to be base paired in a pseudoknot-like structure (see below). A second possible base-paired region, involving positions 84 to 88 and 128 to 132, was not conserved, indicating that this region is unlikely to be base paired, although mutations in this area disrupted regulation.

The *B. brevis rpsD* leader sequence exhibited lower similarity to the other *Bacillus rpsD* leaders (51% identity to *B. subtilis*), and only some of the structural features were found. The basic double-stem-loop structure was conserved (Fig. 4B). However, the side stem-loop in the first helix (corresponding to positions +51 to +64 in *B. subtilis*) was absent, and the upper region of this part of the structure was no longer base paired. The region corresponding to the side bulge at +71 to +76 of the *B. subtilis* sequence (GCGUAA) was located within a larger loop sequence at the top of the first helical domain; conservation with 16S rRNA in this region was maintained, with a single U-to-C alteration at the position corresponding to +72; this residue also varied in the *B. circulans* sequence and does not match the 16S sequence for any of the leaders. The 4-base side bulge at positions +93 to +96 in the *B. subtilis* sequence is 17 bases in length in the *B. brevis* sequence, while the 4-base side bulge at positions

FIG. 5. Secondary-structure models of the *B. subtilis rpsD* regulatory target site and the S4-binding region of 16S rRNA. (A) *rpsD* target site, positions +33 to +173. The effects of nucleotide substitution mutations in various regions of the leader on regulation are indicated by + (wild-type repression by S4) or - (loss of repression by S4). (B) Putative binding site for S4 on *B. subtilis* 16S rRNA. The region from positions 437 to 547 is shown. Dashed boxes show regions conserved between 16S rRNA and the *rpsD* regulatory target site.

+103 to +106 is one nucleotide larger in *B. brevis* and maintains a GAA sequence conserved in all of the *Bacillus rpsD* leaders. The second stem-loop region contains a shorter helical domain than is found in the other *Bacillus* leaders; part of the Shine-Dalgarno sequence (GGAGG) is included within the base-paired region in the *B. brevis* leader, while in the other *Bacillus* sequences, the Shine-Dalgarno sequence is located downstream of this helix.

Mutagenic analysis of the *rpsD* leader region. The roles of various regions within the *rpsD* leader region in repression by S4 were tested by oligonucleotide-directed mutagenesis. The sequence alterations tested are shown in Fig. 1. Their effects on regulation are listed in Table 2 and summarized in Fig. 5A. Sequence alterations outside the +30 to +190 region had no effect on regulation, while the majority of mutations tested within this region resulted in total loss of sensitivity to repression by excess S4.

The region at positions +70 to +78 exhibits structural and primary sequence similarity to a portion of the region of 16S rRNA with which S4 is believed to interact (Fig. 5B), and the Pst-1 triple mutant (affecting residues 71, 74, and 76) was previously shown to abolish repression by S4 (11). Mutants containing one or two of the alterations present in this region were most important. Any of the single or double mutants tested exhibited partial derepression; full derepression apparently requires all three alterations present in the original Pst-1 triple mutant, although not all combinations were tested.

The region at positions +125 to +130 was also found to resemble a portion of the probable S4 target site on 16S rRNA, and the Xba-2 mutant, which contains alterations in two of the conserved residues (A129 and G130), was previ-

ously shown to be derepressed (11). The Xba-5 (A126U, A127C), Sal-1 (A127C), and A126U mutants, which alter other residues conserved with the 16S rRNA sequence, were therefore constructed. These mutants exhibited normal regulation, indicating that this sequence is unlikely to play a role in regulation by S4. The residues altered in the Xba-2 mutation must therefore be important for reasons different from conservation with 16S rRNA.

The upper portion of the upstream stem-loop structure in the regulatory target site contains several self-complementary regions, permitting the design of alternate structural models differing in the specific base-paired residues. In an attempt to distinguish among these models, several mutants (Sac-1, Nru-1, Sau-1, and Taq-1) in which possible base-paired residues were altered were constructed. Double mutations were generally used to ensure disruption of helical regions. In addition, mutations on opposite sides of the putative helical regions were combined in order to test whether compensatory mutations which were predicted to restore base-pairing (Sac-1-Taq-1, Sau-1-Nru-1, and Sau-1-Sac-1) would restore function. All sequence alterations in this region, whether in base-paired or loop regions, resulted in loss of regulation, although the single mutation in Sac1R (G94U) resulted in only partial derepression. No combination of possible compensatory mutations was able to restore normal regulation. The Sau-1-Nru-1 mutant exhibited partial repression; this is consistent with the phylogenetic data, which indicated that the pairing predicted to be restored in this mutant is most conserved in other *Bacillus* species. The failure to restore normal regulation with compensatory base changes suggests that this region is very sensitive to sequence alteration, even in regions predicted to be helical.

The possibility that the region immediately downstream

from the translational start codon could be base paired with the region near position +90 (11) to form a pseudoknot-like structure similar to the S4 target site at the start of the S13 gene in the *E. coli* α operon (25) was tested by construction of the R3C and R3S mutants, in which the putative base pairing was disrupted. These mutations did not affect regulation, although the level of β -galactosidase was reduced, possibly because of reduced stability of the *rpsD-lacZ* fusion product. Furthermore, expression was still derepressed in the R3C-Sac-1 mutant, in which base pairing was predicted to be restored. These results provide no support for a model similar to the structure found in the *E. coli* S4 target site.

DISCUSSION

The *B. subtilis* *rpsD* gene, which encodes ribosomal protein S4, is subject to autogenous control; sequences in the leader region of the gene are required for repression by excess S4 protein. In *E. coli*, S4 regulates expression of the α operon, in which it is encoded, by binding to a target site at the start of the S13 gene, the first gene in the operon (6, 25). In *B. subtilis*, *rpsD* is monocistronic and is not in the major cluster of ribosomal genes, where the remaining α operon genes are located. The goal of this study was to delimit the sequences required for repression of *rpsD* expression by S4 and to improve our understanding of the target site structure by using a combination of mutagenic and phylogenetic analyses.

Deletion analysis indicated that the *rpsD* regulatory target site was located within positions +30 to +190 of the transcript (the AUG start codon is at position +183). This is in contrast to the S4 target site in the *E. coli* α operon, which spans the beginning of the S13 coding sequence, including sequences 32 bases downstream from the translational start codon (position +126 of the mRNA; 25). Replacement of the *rpsD* promoter with a derivative of the *E. coli lacUV5* promoter had no effect on autogenous regulation, indicating that repression occurs post-transcription initiation. Repression by S4 could occur at the level of translation or by an effect on mRNA stability; biochemical analyses will be necessary to distinguish between these possibilities.

In *E. coli*, the S4 target site on the α operon mRNA has been shown by biochemical and genetic studies to form a complex pseudoknot-like structure (25, 26). There is very little primary sequence homology between the *E. coli* α operon S4 target site and the *B. subtilis* *rpsD* target site. Although it was also possible to model the *B. subtilis* target site as a pseudoknot-like structure (11), this model is not supported by the data presented in this study. Sequences downstream from the translation initiation site which were postulated to base pair with upstream sequences to form a pseudoknot could in fact be altered to disrupt base pairing with no effect on repression by S4. Furthermore, while the major structural features of the target site were conserved in other *Bacillus* species, the potential to form a pseudoknot structure was not conserved. We therefore conclude that the simpler two-helix model is more likely to be an accurate representation of the *rpsD* regulatory target site in *Bacillus* species.

The *rpsD* leader region was highly conserved in members of the *B. subtilis* group. Conservation was especially high in *B. megaterium* and *B. licheniformis* and was lower in *B. stearothermophilus* and *B. circulans*. Leader region conservation was correlated with the conservation of the S4 amino acid sequence for those genes for which the coding-region sequence was determined. On the basis of 16S rRNA se-

quence data, *B. brevis* has been proposed to belong to a *Bacillus* subgroup different from the *B. subtilis* group (23); the *rpsD* gene of *B. brevis* exhibited much greater divergence, both in the leader region sequence and structure and in the S4 coding region (65% amino acid identity to *B. subtilis* S4 compared with 82% for *B. stearothermophilus* versus *B. subtilis* [12] and 52% for *E. coli* versus *B. subtilis* [10]). It therefore appears that the pattern of phylogenetic relatedness derived from 16S rRNA sequence data can also be observed in an analysis of highly conserved ribosomal protein genes. *B. stearothermophilus* was placed farther from *B. subtilis* on the basis of a simple analysis of 16S rRNA homology scores (1), presumably because of differences in G+C content, while analysis of *rpsD* sequences indicates that these species are in fact closely related. Also, on the basis of analysis of *rpsD* versus 16S rRNA, *B. megaterium* appears to be more closely related to *B. subtilis*.

S4 acts as a negative regulator for its own expression in both *Bacillus* species and *E. coli*; in the latter, S4 also regulates the expression of other genes in the α operon with *rpsD*, while *rpsD* is monocistronic in *Bacillus* species. We have not yet tested whether *B. subtilis* S4 is involved in regulation of other ribosomal protein genes; however, overproduction of S4 does not adversely affect growth in *B. subtilis* (12), suggesting that S4 does not repress the expression of other genes.

The *Bacillus* and *E. coli* S4 proteins are approximately 50% identical, and conservation is especially high in the region of the protein shown for *E. coli* S4 to interact with 16S rRNA (4, 5). The proteins are likely to play very similar roles in ribosomal structure and function; the region of 16S rRNA with which S4 interacts is highly conserved in eubacteria, and active ribosomal subunits could be assembled in heterologous reconstitution experiments mixing *E. coli* and *B. stearothermophilus* components (16, 22). *E. coli* S4 was shown to bind in vitro to 16S rRNA from *B. stearothermophilus* but not to that from *B. subtilis* or *B. brevis* (27), indicating that there are significant differences in the RNA recognition properties of the different S4 species. Analysis of the ability of the S4 variants now available to interact with rRNA and *rpsD* leader region mRNA of different origins will provide valuable information about the specificity of these protein-RNA interactions.

The molecular mechanism of α operon repression by S4 in *E. coli* has been studied in detail and has been proposed to involve an allosteric alteration in the conformation of the mRNA target site such that the 30S ribosomal subunit can still bind but cannot initiate translation (7, 26). No information is yet available concerning the molecular mechanism for S4 repression of *rpsD* expression in *B. subtilis*. It will be of interest to determine whether the structural differences between the S4 target sites in the two systems are associated with fundamental differences in the mechanism of repression by S4.

ACKNOWLEDGMENTS

We thank Tsetska Takova and Brian Glass for technical assistance and David Draper, George Fox, and Robin Gutell for providing data prior to publication.

This work was supported by National Institutes of Health grant GM-40650 and American Cancer Society Junior Faculty Research Award JFRA-226 to T.H.

REFERENCES

1. Ash, C., J. A. E. Farrow, S. Wallbanks, and M. D. Collins. 1991. Phylogenetic heterogeneity of the genus *Bacillus* revealed by

- comparative analysis of small-subunit-ribosomal RNA sequences. *Lett. Appl. Microbiol.* **13**:202–206.
2. Bedwell, D., G. Davis, M. Gosink, L. Post, M. Nomura, H. Kestler, J. M. Zengler, and L. Lindahl. 1985. Nucleotide sequence of the alpha ribosomal protein operon of *Escherichia coli*. *Nucleic Acids Res.* **13**:3891–3903.
 3. Boylan, S. A., J.-W. Suh, S. M. Thomas, and C. W. Price. 1989. Gene encoding the alpha core subunit of *Bacillus subtilis* RNA polymerase is cotranscribed with the genes for initiation factor 1 and ribosomal proteins B, S13, S11 and L17. *J. Bacteriol.* **171**:2553–2562.
 4. Changchien, L.-M., and G. R. Craven. 1976. The function of the N-terminal region of ribosomal protein S4. *J. Mol. Biol.* **108**:381–401.
 5. Conrad, R. C., and G. R. Craven. 1987. A cyanogen bromide fragment of S4 that specifically rebinds 16S RNA. *Nucleic Acids Res.* **15**:10331–10343.
 6. Deckman, I. C., and D. E. Draper. 1985. Specific interaction between ribosomal protein S4 and the α operon messenger RNA. *Biochemistry* **24**:7860–7865.
 7. Draper, D. E. 1992. Personal communication.
 8. Fox, G. E. 1992. Personal communication.
 9. Green, C. J., G. C. Stewart, M. A. Hollis, B. S. Vold, and K. F. Bott. 1985. Nucleotide sequence of the *Bacillus subtilis* ribosomal RNA operon, *rnbB*. *Gene* **37**:261–266.
 10. Grundy, F. J., and T. M. Henkin. 1990. Cloning and analysis of the *Bacillus subtilis* *rpsD* gene, encoding ribosomal protein S4. *J. Bacteriol.* **172**:6372–6379.
 11. Grundy, F. J., and T. M. Henkin. 1991. The *rpsD* gene, encoding ribosomal protein S4, is autogenously regulated in *Bacillus subtilis*. *J. Bacteriol.* **173**:4595–4602.
 12. Grundy, F. J., and T. M. Henkin. 1992. Unpublished results.
 13. Gutell, R. 1991. Personal communication.
 14. Henkin, T. M., and G. H. Chambliss. 1984. Genetic mapping of a mutation causing an alteration in *Bacillus subtilis* ribosomal protein S4. *Mol. Gen. Genet.* **193**:364–369.
 15. Henkin, T. M., and A. L. Sonenshein. 1987. Mutations of the *Escherichia coli* *lacUV5* promoter resulting in increased expression in *Bacillus subtilis*. *Mol. Gen. Genet.* **209**:467–474.
 16. Higo, K., W. Held, L. Kahan, and M. Nomura. 1973. Functional correspondence between 30S ribosomal proteins of *Escherichia coli* and *Bacillus stearothermophilus*. *Proc. Natl. Acad. Sci. USA* **70**:944–948.
 17. Higo, K., and K. Loertscher. 1974. Amino-terminal sequences of some *Escherichia coli* 30S ribosomal proteins and functionally corresponding *Bacillus stearothermophilus* ribosomal proteins. *J. Bacteriol.* **118**:180–186.
 18. Jinks-Robertson, S., and M. Nomura. 1982. Ribosomal protein S4 acts in *trans* as a translational repressor to regulate expression of the α operon in *Escherichia coli*. *J. Bacteriol.* **151**:193–202.
 19. Kunkel, T. A. 1985. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**:488–492.
 20. Messing, J. 1983. New M13 vectors for cloning. *Methods Enzymol.* **101**:20–78.
 21. Miller, J. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 22. Nomura, M., P. Traub, and H. Bechmann. 1968. Hybrid 30S ribosomal particles reconstituted from components of different bacterial origins. *Nature (London)* **219**:793–799.
 23. Rossler, D., W. Ludwig, K. H. Schleifer, C. Lin, T. J. McGill, J. D. Wisotzkey, P. Jurtshuk, Jr., and G. E. Fox. 1991. Phylogenetic diversity in the genus *Bacillus* as seen by 16S rRNA sequencing studies. *Syst. Appl. Microbiol.* **14**:266–269.
 24. Stern, S., R. C. Wilson, and H. F. Noller. 1986. Localization of the binding site for protein S4 on 16S ribosomal RNA by chemical and enzymatic probing and primer extension. *J. Mol. Biol.* **192**:101–110.
 25. Tang, C. K., and D. E. Draper. 1989. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. *Cell* **57**:531–536.
 26. Tang, C. K., and D. E. Draper. 1990. Evidence for allosteric coupling between the ribosome and repressor binding sites of a translationally regulated mRNA. *Biochemistry* **29**:4434–4439.
 27. Thurlow, D. L., and R. A. Zimmermann. 1978. Conservation of ribosomal protein binding sites in prokaryotic 16S RNAs. *Proc. Natl. Acad. Sci. USA* **75**:2859–2863.
 28. Vartikar, J. V., and D. E. Draper. 1989. S4-16S ribosomal RNA complex. Binding constant measurements and specific recognition of a 460-nucleotide region. *J. Mol. Biol.* **209**:221–234.
 29. Wei, Y.-G., and S. J. Surzycki. 1986. Screening recombinant clones containing sequences homologous to *Escherichia coli* genes using single-stranded bacteriophage vector. *Gene* **48**:251–256.
 30. Zuber, P., and R. Losick. 1987. Role of AbrB in Spo0A- and Spo0B-dependent utilization of a sporulation promoter in *Bacillus subtilis*. *J. Bacteriol.* **169**:2223–2230.