JOURNAL OF BACTERIOLOGY, Nov. 1990, p. 6372–6379
0021-9193/90/116372-08$02.00/0

Vol. 172, No. 11

Cloning and Analysis of the *Bacillus subtilis rpsD* Gene, Encoding Ribosomal Protein S4

FRANK J. GRUNDY AND TINA M. HENKIN*

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

Received 11 April 1990/Accepted 27 July 1990

The *rpsD* gene, encoding ribosomal protein S4, was isolated from *Bacillus subtilis* by hybridization with oligonucleotide probes derived from the S4 amino-terminal protein sequence. Sequence analysis of the cloned DNA indicated that *rpsD* is likely to be monocistronic, in contrast to *Escherichia coli rpsD*, which is located in the α operon and is the translational regulator for α operon ribosomal protein gene expression in *E. coli*. The cloned gene was shown to map at position 263° on the *B. subtilis* chromosome, at the position to which mutations conferring alterations in the electrophoretic mobility of protein S4 were localized. A promoter was identified upstream of the *rpsD* coding sequence; initiation of transcription at this promoter would result in a transcript containing a leader region 180 bases in length. Immediately downstream of the *rpsD* coding region were two sequences resembling transcriptional terminators. An open reading frame homologous to tyrosyl-tRNA synthetase (*tyrS*) genes was identified downstream of *rpsD* but in the opposite orientation. The leader region of *rpsD* mRNA is predicted to have extensive secondary structure, resembling a region of *B. subtilis* 16S rRNA where S4 is likely to bind; similar mRNA features have been found to be important in ribosomal gene regulation in *E. coli*. These results provide the first steps toward analysis of the regulation of *rpsD* gene expression in *B. subtilis*.

Ribosome biosynthesis is tightly controlled in the gram-negative bacterium *Escherichia coli*. A number of ribosomal protein operons from this organism have been characterized in detail and were found to be autogenously regulated at the translational level by binding of one ribosomal protein encoded in the operon to a target site on the mRNA (8, 42). In contrast, very little is known about regulation of ribosomal gene expression in procaryotes evolutionarily distant from *E. coli*, such as the gram-positive bacterium *Bacillus subtilis*. Although a large number of genetic determinants for ribosomal components have been localized on the *B. subtilis* chromosomal map (28), the mechanisms for ribosomal gene regulation have not yet been analyzed.

One of the best-characterized *E. coli* ribosomal protein operons is the α operon, which includes the genes for ribosomal proteins S13, S11, S4, and L17 and for the α subunit of RNA polymerase, in the order S13-S11-S4-α-L17 (2). S4 is the translational regulator for this operon and binds to a target site on the α operon mRNA near the start of the S13-coding region (9, 19). Binding of S4 to the mRNA blocks translation of S13 and also inhibits translation of S11 and S4, probably by a translational coupling mechanism (35). Expression of *rpoA*, which encodes the α subunit of RNA polymerase, is not controlled by S4, and there is evidence that a second S4-binding site on the mRNA upstream from the L17-coding region is necessary for S4 regulation of L17 expression (24).

The α operon of *B. subtilis* was recently isolated, and its DNA sequence was determined (3, 33). The most striking feature of the *B. subtilis* α operon is the absence of the S4-coding region. The α operon is located in the major cluster of ribosomal protein genes, at 12° on the 360° *B. subtilis* map (33). Mutations which result in two different alterations in the electrophoretic mobility of S4 protein were mapped to 263°, distant from the site of the α operon (14, 15). Together, these results indicate that the α operon genes in *E. coli* and *B. subtilis* differ in the location of the gene which in *E. coli* encodes the autogenous regulatory protein for the operon. It is therefore of interest to examine regulation of S4 and α operon gene expression in *B subtilis*. We report here a first step toward this goal, which is the cloning and characterization of the *rpsD* gene of *B. subtilis*, encoding ribosomal protein S4. We find that the *rpsD* gene is apparently monocistronic in *B. subtilis*. Also, a gene homologous to tyrosyl-tRNA synthetases from *Bacillus stearothermophilus* (39), *Bacillus caldotenax* (21), and *E. coli* (1) was found to be located immediately downstream of *rpsD* but in the opposite orientation.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. The *B. subtilis* strains used in this study are BR151 (*lys-3 metB10 trpC2*; obtained from G. Chambliss) and 1A92 [*arg(GH)2 aroG932 bioB141 sacA321*; obtained from the Bacillus Genetic Stock Center]. Cultures of *E. coli* and *B. subtilis* were routinely grown in L broth; for propagation of bacteriophage λ, $MgSO_4$ was added to a final concentration of 10 mM. Bacteriophage M13 was propagated in strain JM103 as described by Messing (25). Ampicillin and chloramphenicol (Sigma Chemical Co.) were used at 30 and 5 μg/ml, respectively. Plasmid pMMN13 (26) was obtained from M. Nakano. A 920-bp *Hin*dIII fragment (positions 1 to 920) including the *B. subtilis rpsD* amino-terminal coding region was inserted into the *Hin*dIII site of pMMN13 to generate plasmid pTMH110; this plasmid was integrated into the *rpsD* region of strain 1A92 to generate strain TH110. Plasmid pTMH111 was generated by insertion of a 380-bp *Hin*cII-*Hin*dIII fragment (positions 540 to 920) into plasmid pMMN13 which had been digested with *Sma*I and *Hin*dIII. Plasmid pTMH112 was generated by insertion of a 480-bp

* Corresponding author.

*Sau*3A-*Hin*dIII fragment (positions 440 to 920) into plasmid pMMN13 which had been digested with *Bam*HI and *Hin*dIII.

Cloning procedures. Restriction endonucleases and DNA-modifying enzymes were purchased from New England BioLabs (Beverly, Mass.) and used as instructed by the manufacturer. Chromosomal DNA was isolated from *B. subtilis* BR151 as previously described (14). Phage λEMBL-4 DNA and packaging extracts were purchased from Stratagene, Inc. (San Diego, Calif.) and used as instructed by the manufacturer. Phage λEMBL-4 arms were ligated with *Eco*RI-digested chromosomal DNA (not size fractionated) from strain BR151; since the packaging range of this vector is 9 to 23 kb, many clones contain multiple insert fragments that were not necessarily contiguous on the chromosome. Recombinant phage were selected by using strain P2392 (Spi$^-$ selection) and then propagated on strain LE392. Genetic techniques were as previously described (14).

Hybridization techniques. Chromosomal DNA was separated by agarose gel electrophoresis and transferred to nitrocellulose membranes (BA85; Schleicher & Schuell), using standard techniques. Oligonucleotide probes were 5′ end labeled with [λ-^{32}P]ATP, using T4 polynucleotide kinase, and purified by column chromatography (Stratagene). Radiochemicals were purchased from New England Nuclear or Amersham. Hybridization was at 42° in a solution containing 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M trisodium citrate), 10× Denhardt solution, and 50 μg of sheared, denatured salmon sperm DNA per ml. Filters were washed with 6× SSC at room temperature and at the hybridization temperature; stringency was increased by reducing the SSC concentration or increasing the hybridization and wash temperatures. Plaque lifts of phage λ clones were hybridized by using similar conditions. Plaque lifts of phage M13 clones were performed as described by Wei and Surzycki (38), using nondenaturing conditions for transfer of single-stranded DNA only. Plasmid probes were nick translated by using a kit purchased from Bethesda Research Laboratories. Strand-specific probes were synthesized from M13 template DNA by using a modification of standard DNA sequencing protocols, in the absence of dideoxynucleoside triphosphates.

DNA sequencing. Cloning in phage M13, preparation of single-stranded template DNA, and sequencing of double-stranded plasmid DNA were done by standard techniques. Dideoxynucleotide sequencing was carried out by using Sequenase (United States Biochemicals); dITP was used to resolve ambiguities resulting from compressions in G+C-rich regions. Sequencing primers were M13 primers (New England BioLabs) or synthetic oligonucleotides (The DNA Factory). DNA sequence data were analyzed by using the PC/Gene system and the GenBank On-line Service (Intelligenetics).

Transcriptional analysis. Total RNA was isolated from exponentially growing cultures of strain BR151 as described by Wu et al. (41). RNA probes were synthesized in vitro, using plasmids derived from pGEM7Z(+) and the Riboprobe transcription system (Promega). For primer extension analysis, oligonucleotide S4RC4 (5′-ACCGTCTGCTCCATTCG-3′), corresponding to nucleotides 697 to 713 in Fig. 2, was 5′ end labeled and gel purified and then used in primer extension reactions with 80 μg of cellular RNA. Extension products were resolved on a denaturing 6% polyacrylamide gel. DNA sequencing reactions primed with the same oligonucleotide were used as size standards.

Nucleotide sequence accession number. The sequence reported has been assigned GenBank accession number M34718.

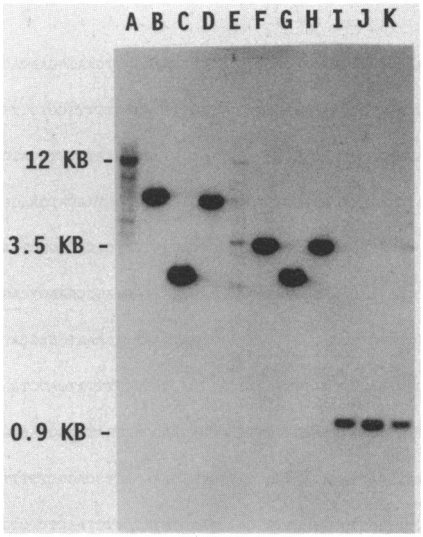

FIG. 1. Hybridization analysis of phage λ clones. Digested DNA was resolved on a 1% agarose gel, transferred to nitrocellulose, and probed with ^{32}P-end-labeled oligonucleotide S4RC2 (5′-ATACCAA GACGAGAAAGTTTCCAAGAAGGACCAGTGTAACGAGC CAT-3°). Lanes: A and E, BR151 chromosomal DNA; B, F, and I, λRC2-1; C, G, and J, λRC2-2; D, H, and K, λRC2-3; A to D, *Eco*RI digests; E to H, *Eco*RI-*Stu*I digests; I to K, *Hin*dIII digests. Positions of the 12-kb *Eco*RI, 3.5-kb *Eco*RI-*Stu*I, and 0.9-kb *Hin*dIII bands are indicated.

RESULTS

Cloning of *B. subtilis rpsD*. Cloning of *rpsD* used synthetic oligonucleotide probes designed on the basis of the *B. subtilis* S4 amino-terminal amino acid sequence (18). Three different oligonucleotide probes were used, all of which hybridized strongly to an *Eco*RI fragment approximately 12 kb in size, and to a 3.5-kb band in DNA digested with *Eco*RI and *Stu*I, in Southern analyses of *B. subtilis* chromosomal DNA.

A library of *Eco*RI-digested *B. subtilis* DNA in λEMBL4 was constructed, and recombinant plaques were screened by hybridization with the oligonucleotides. Three candidates that hybridized with the oligonucleotide probes were purified and screened for the presence of the 12-kb *Eco*RI fragment in Southern hybridization analyses. None of the three candidates contained the expected 12-kb *Eco*RI insert band that hybridized with the oligonucleotide probes (Fig. 1). Clone λRC2-1 contained a 7-kb *Eco*RI band, clone λRC2-2 contained a 3-kb *Eco*RI band, and clone λRC2-3 contained a 6.5-kb *Eco*RI band. All three clones contained λ vector bands of the expected sizes (9 and 20 kb) and additional *Eco*RI insert bands of various sizes that did not hybridize with the probe. In *Eco*RI-*Stu*I double digests, clones λRC2-1 and λRC2-3 showed a hybridizing band 3.5 kb in size, which comigrated with the band detected in the chromosomal digest; the hybridizing band in λRC2-2 was unchanged from that observed in the *Eco*RI digest. All three λ clones showed a 0.9-kb hybridizing band in *Hin*dIII digests.

The simplest explanation for these results is that the 12-kb *Eco*RI band was unstable during propagation in the λ vector, and the three different-size *Eco*RI bands in the λ clones represent three different deletion products. Two of the three deletion events apparently occurred outside of the 3.5-kb

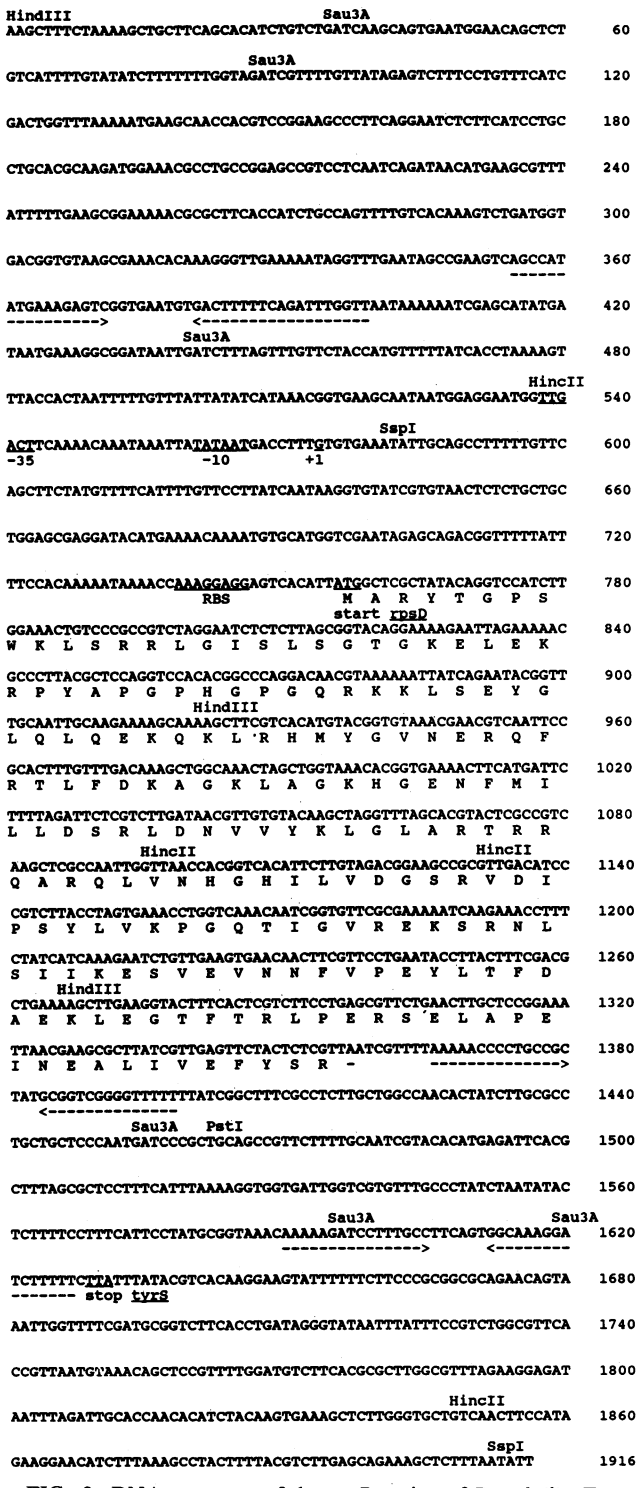

FIG. 2. DNA sequence of the *rpsD* region of *B. subtilis*. Translation of the *rpsD* coding region is shown below the DNA sequence. Certain restriction sites are indicated above the line. The putative promoter region (−35, −10, and +1) and ribosome-binding site (RBS) are underlined. Inverted repeat sequences that resemble transcriptional terminators are shown as dashed arrows below the line. The stop codon for the *tyrS* open reading frame, which originates downstream of the end of the sequence shown and is directed toward the end of *rpsD*, is underlined.

*Eco*RI-*Stu*I band, whereas all three events occurred outside of the 0.9-kb *Hin*dIII band. Since the 0.9-kb *Hin*dIII band was the smallest band that hybridized to the oligonucleotide probes, it was predicted that the amino-terminal coding region of S4 was within this fragment. The 0.9-kb *Hin*dIII fragment from λRC2-1 was subcloned into phage M13mp18, and a probe derived from this subclone was used to probe chromosomal digests of *B. subtilis*. As predicted, this probe hybridized to 12-kb *Eco*RI and 3.5-kb *Eco*RI-*Stu*I bands in chromosomal digests which were identical in size to those detected by the oligonucleotide probes (data not shown). This result indicated that the 0.9-kb *Hin*dIII band was in fact derived from the 12-kb *Eco*RI fragment. The DNA sequence of the 0.9-kb *Hin*dIII fragment was determined and examined for a coding region matching the amino-terminal sequence of *B. subtilis* S4. The sequence was found to correspond to the S4 amino terminus, indicating that the 0.9-kb fragment contains the amino-terminal coding sequence of S4. Overlapping fragments from the λRC2-1 clone were subcloned and sequenced.

DNA sequence analysis of the *B. subtilis rpsD* gene. The DNA sequence for a 1.9-kb region encompassing the S4-coding sequence is shown in Fig. 2. The S4-coding region is located at positions 755 to 1355 and is preceded by a sequence showing an excellent match to ribosome-binding sites for *B. subtilis* genes (23). The predicted amino-terminal S4 sequence is identical to the amino-terminal protein sequence for *B. subtilis* obtained by Higo et al. (18; Fig. 3); the amino-terminal methionine is apparently removed in both *B. subtilis* and *E. coli* (18, 30). In addition, the carboxyl-terminal amino acids (Phe-Tyr-Ser-Arg) matched data obtained by carboxypeptidase digestion and amino acid analysis of *B. subtilis* S4 (T. Henkin, Ph.D. thesis, University of Wisconsin, Madison, 1984).

The predicted *B. subtilis* S4 amino acid sequence was compared with the *E. coli* S4 sequence (Fig. 3). Amino acid identity was 52%; with the addition of conservative amino acid substitutions, homology was 61%. The similarity between the two proteins was highest at the amino-terminal and carboxyl-terminal regions of the protein and was 72% (including conservative amino acid substitutions) in the region of the protein shown for *E. coli* S4 to be critical for binding to 16S rRNA (6, 7; underlined in Fig. 3). In *E. coli*, mutations that change the glutamine at position 53 to leucine confer the *ram* (ribosome ambiguity) phenotype (36); this glutamine is conserved in *B. subtilis* S4. No *ram* mutants have yet been isolated in *B. subtilis* (T. Henkin, unpublished results).

Analysis of the sequences flanking the S4-coding sequence revealed a putative vegetative RNA polymerase promoter sequence approximately 200 bp upstream of the S4 start codon (Fig. 2). This sequence (TTGACT-17 bp-TATAAT), starting at position 538, shows five-of-six and six-of-six adherence to the consensus sequences for the −35 and −10 regions of *B. subtilis* σ^A promoters (TTGACA and TATAAT) and perfect spacer length (12); in addition, it shows three-of-five adherence to a conserved element (RTRTG) at the −16 region of the promoter (16). This sequence would therefore be predicted to act as a promoter in vivo in both *B. subtilis* and *E. coli*. Fusions of fragments containing this region to a promoterless *lacZ* gene did in fact result in high-level synthesis of β-galactosidase in both organisms, and transcriptional mapping experiments indicated that a transcript originating at this position was present in exponential-growth-phase *B. subtilis* cells (see below).

Immediately downstream of the carboxyl-terminal coding

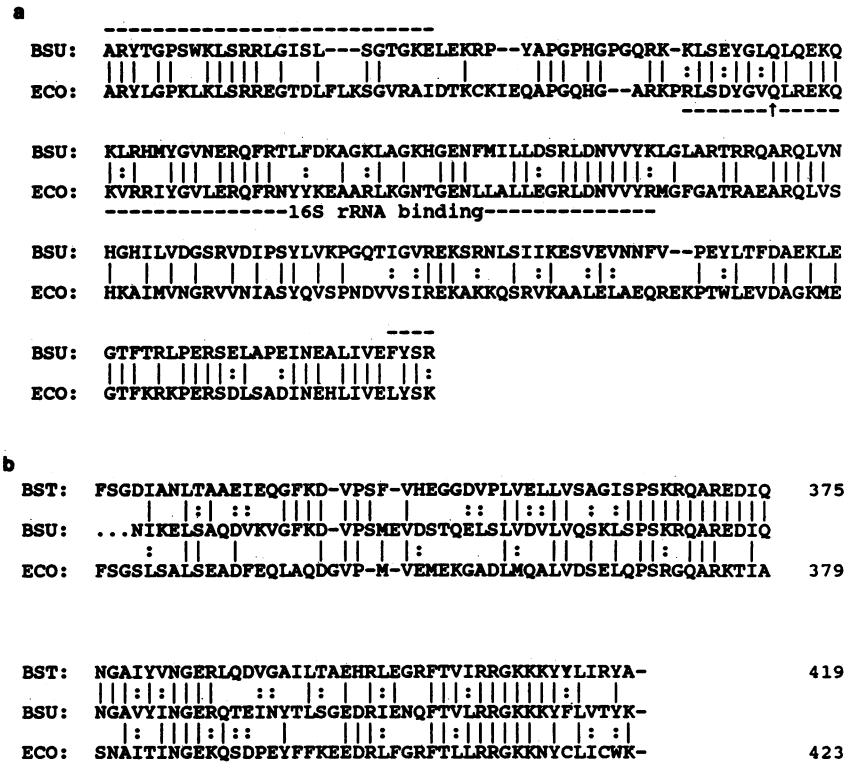

FIG. 3. Amino acid sequence comparison of S4 and TyrS proteins from *B. subtilis* and *E. coli*. (a) Amino acid sequences for S4 protein from *B subtilis* (BSU) and *E. coli* (ECO), derived from Bedwell et al. (2), are shown. Symbols: |, identical amino acids; :, conservative amino acid substitutions (A, S, and T; D and E; N and Q; R and K; I, L, M, and V; F, Y, and W). The regions of S4 sequence previously known by amino acid sequence analysis (22; Henkin, Ph.D. thesis, 1984) are overlined; the region of *E. coli* S4 shown to be important for binding to 16S rRNA (6, 7) is underlined; the position of glutamine-to-leucine mutations which result in the *ram* phenotype (36) is indicated with an arrow. (b) Amino acid sequence of ORF-3 and *tyrS* coding regions. The amino acid sequence of *B. subtilis* ORF-3 (BSU) is aligned with the carboxyl-terminal regions of *tyrS* genes from *E. coli* (ECO; 1) and *B. stearothermophilus* (BST; 39); the sequence of *B. caldotenax tyrS* (21) is identical to that of *B. stearothermophilus* in this region. The positions in the *B. stearothermophilus* and *E. coli* sequences are indicated on the right.

sequence of S4 were two inverted repeat sequences that would be predicted to form a structure resembling Rho-independent transcriptional terminators (29). The predicted stabilities of these two structures are −23 and −29 kcal (ca. −96 and -121 kJ), respectively, according to the algorithm of Zuker and Stiegler (43). A third structure of this type, with a predicted stability of −13 kcal (−53 kJ), is located upstream of the position of the putative promoter (Fig. 4).

Computer analysis of sequences in the region surrounding *rpsD* revealed three additional open reading frames extending beyond the sequenced region. Open reading frame 1 (ORF-1) is 102 amino acids long and extends from upstream of the *Hin*dIII site toward *rpsD* (Fig. 4); the inverted repeat structure described above (positions 355 to 399) is located downstream of the end of ORF-1 (position 307) and could represent the terminator for that gene. ORF-2, 153 amino acids in length, extends in the reverse orientation from position 496 past the *Hin*dIII site at the beginning of the sequenced region, with a possible ribosome-binding site and methionine codon at position 457; the inverted repeat structure is located near the beginning of ORF-2. No sequences resembling *B. subtilis* vegetative promoters were found in the appropriate orientation to direct transcription of this region, but it is possible that this region is transcribed by a

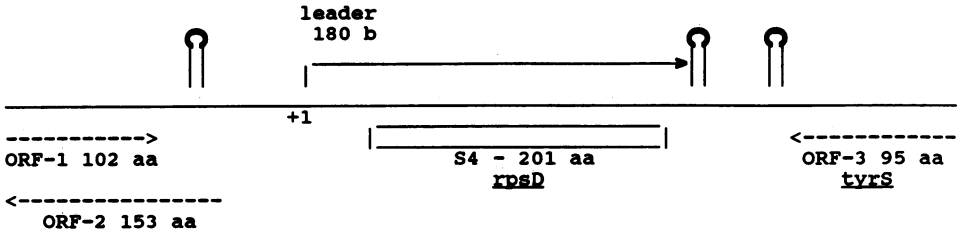

FIG. 4. Map of the 1.9-kb *rpsD* region. The S4-coding region is shown as an open box below the line. ORF-1, -2, and -3 are shown as dashed arrows, indicating the orientation of the coding sequences; ORF-3 has been identified as the carboxy-terminal coding region of the *tyrS* gene. The putative start point and direction of transcription are shown above the line, with the expected 180-bp leader region indicated. The stem-loop structures above the line designate the positions of putative transcriptional terminator structures.

form of RNA polymerase with a different promoter specificity. ORF-3, 95 amino acids in length, is located downstream of *rpsD*, in the orientation opposite to that of the S4-coding sequence. It is possible that the second terminatorlike structure downstream of the S4-coding sequence represents the transcriptional terminator for this gene. Searches of the GenBank and EMBL data bases for sequences homologous to these three open reading frames revealed no significant homology for the region upstream of *rpsD*; confirmation of the significance of the upstream open reading frames will require determination of additional DNA sequence information for this region. The open reading frame downstream of *rpsD*, in the opposite orientation, exhibited very high homology with the carboxyl-terminal portion of tyrosyl-tRNA synthetase (*tyrS*) genes from *B. stearothermophilus* and *B. caldotenax* (54% amino acid identity, 72% with conservative substitutions; 21, 39) and *E. coli* (42% amino acid identity, 55% including conservative substitutions; 1). The alignment of these coding regions is shown in Fig. 3b. Homology between the *B. subtilis* and *E. coli* sequences is similar to that between *B. stearothermophilus* and *E. coli*. The *tyrS* gene locus has not previously been reported in *B. subtilis* (28); our results indicate that this gene is located adjacent to *rpsD*, at 263°. Determination of the DNA sequence of the amino-terminal portion of the *tyrS* gene is in progress.

Transcriptional mapping. The start point of *rpsD* transcription was analyzed by RNase protection and primer extension experiments. For RNase protection, an RNA probe corresponding to positions 540 to 699 of the DNA sequence was hybridized to cellular RNA and digested with RNases A and T_1. A protected fragment approximately 125 bases in length was detected, and no full-length protection of the probe was observed (data not shown). To more precisely define the transcription initiation point within this region, primer extension analysis was performed. A single primer extension product was detected, corresponding to a transcript initiating at the G at position 574 of the DNA sequence (Fig. 5). These results demonstrate that the promoter element identified in the DNA sequence was utilized in vivo and that no transcription originated upstream of this site. Initiation of transcription at this position would result in a transcript with a leader region 180 bases in length preceding the S4-coding region. Northern analysis of RNA isolated from exponential-growth-phase *B. subtilis* cells resulted in the detection of a single transcript, approximately 0.8 kb in length (data not shown), which is the size predicted for a transcript beginning at the start point at position 574 and ending at the terminatorlike structure immediately downstream of the S4-coding sequence.

Identification of the position of *rpsD* in the *B. subtilis* chromosome. Mutations that result in alterations in the electrophoretic mobility of S4 were previously isolated and mapped to a position on the *B. subtilis* chromosomal map between *argGH* (260°) and *aroG* (264°; 28), at 263° (14, 15). To determine whether these mutations do in fact define the *rpsD* structural gene, the map location of the cloned *rpsD* gene was analyzed by integrational mapping. Plasmid pTMH110, which contains the amino-terminal coding region of *rpsD*, was introduced into strain 1A92 (*argGH aroG*) by transformation, and Cm^r transformants were selected. Since *rpsD* is likely to be an essential gene, transformants could arise only if the insert DNA contained one end of the *rpsD* transcriptional unit, since integration of the plasmid via internal fragments would result in insertional disruption of the chromosomal *rpsD* gene. Recovery of Cm^r transformants therefore provided further confirmation that the 0.9-kb *Hin*dIII fragment contained in pTMH110 includes the *rpsD* promoter. Plasmid pTMH111, a similar construct containing a 380-bp *Hin*cII-*Hin*dIII fragment, did not result in the generation of Cm^r transformants, indicating that this fragment no longer contained the intact promoter; the *Hin*cII site is in the center of the putative −35 sequence. Plasmid pTMH112, which contains the 480-bp *Sau*3A-*Hin*dIII fragment, including the putative promoter region, did result in the generation of Cm^r transformants, indicating that the 5′ end of the transcriptional unit is likely to be between the *Sau*3A and *Hin*cII sites. These results are in agreement with those obtained in transcript mapping experiments cited above.

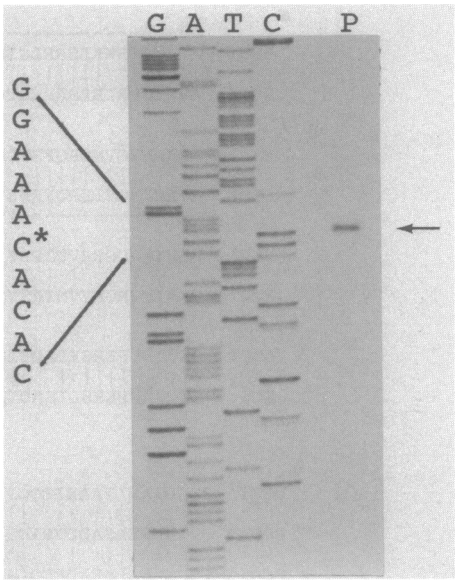

FIG. 5. Primer extension analysis of *rpsD* transcription. RNA from exponential-growth-phase *B. subtilis* cells was hybridized with an oligonucleotide corresponding to positions 697 to 713 of the DNA sequence, and the reverse transcription product was resolved on a 6% denaturing polyacrylamide gel and subjected to autoradiography. Sequencing reactions using the same oligonucleotide as primer were used as size standards. The labeled region corresponds to positions 569 and 578 of the DNA sequence; the primer extension product aligns with position 574 of the DNA sequence.

Strain TH110, a transformant created by the insertion of plasmid pTMH110 into strain 1A92, was used as a recipient in a transformation cross using donor DNA from strain BR151 (Arg^+ Aro^+ Cm^s). Transformants were selected independently for Arg^+ and Aro^+ and screened for Cm^s; linkage was very similar to that obtained for the *rpsD1* and *rpsD2* mutations (14, 15). These results showed that the cloned DNA mapped to the same site of the chromosome as the *rpsD1* and *rpsD2* mutations which conferred alterations in the electrophoretic mobility of protein S4, confirming that this locus is in fact the site of the structural gene for protein S4.

DISCUSSION

We report here the isolation and characterization of the *B. subtilis rpsD* gene, which encodes ribosomal protein S4. The sequence of this protein is 61% homologous (including conservative amino acid substitutions) to *E. coli* S4, but the location of this gene in the genome is very different in the

two organisms. In *E. coli*, the *rpsD* gene is part of the α operon, located in the *str-spc* cluster, which contains four ribosomal protein gene operons, in the order *str*-S10-*spc*-α (20). In *B. subtilis*, the majority of ribosomal protein genes, including genes homologous to those in the *str*, S10, *spc*, and α operons of *E. coli*, as well as *rif* operon genes and the S9 and S20 ribosomal protein genes, are located in a large cluster at 12° on a 360° map of the genome (28). However, there is a marked difference in the arrangement of the α operon genes in the two organisms. In *E. coli*, the α operon contains four ribosomal protein genes and the gene for the α subunit of RNA polymerase, in the order S13-S11-S4-α-L17. Although the majority of α operon transcripts arise as readthrough transcripts originating at the beginning of the upstream *spc* operon (4), α operon genes are regulated as a unit at the translational level. This regulation is mediated by the action of the S4 protein, which acts as a translational repressor (8, 19, 42). In *B. subtilis*, the gene for S4 is not located between the genes for S11 and α; a 180-bp intercistronic region is found at this position (33). In addition, it appears that the *infA* and *rpmJ* genes, located upstream of the S13 gene, are also a part of the α operon transcriptional unit in *B. subtilis* (3).

Mutations conferring alterations in the electrophoretic mobility of protein S4 were localized to a position at 263° on the *B. subtilis* map (14, 15), distant from the major ribosomal gene cluster. The locus defined by these mutations was presumed to be the structural gene for S4 but could have been a gene involved in posttranslational modification of S4. In this study, the structural gene for S4 was cloned and its DNA sequence was determined; the cloned DNA was shown to be derived from the same region of the chromosome as the site to which the S4 mutations were mapped. In the accompanying paper (15), we describe the sequence alterations in the *rpsD* mutants, unequivocably demonstrating that this is the chromosomal locus of the *rpsD* gene. In *B. subtilis*, then, the *rpsD* gene is located at a position distant from the α operon. Since S4 is the translational repressor for α operon

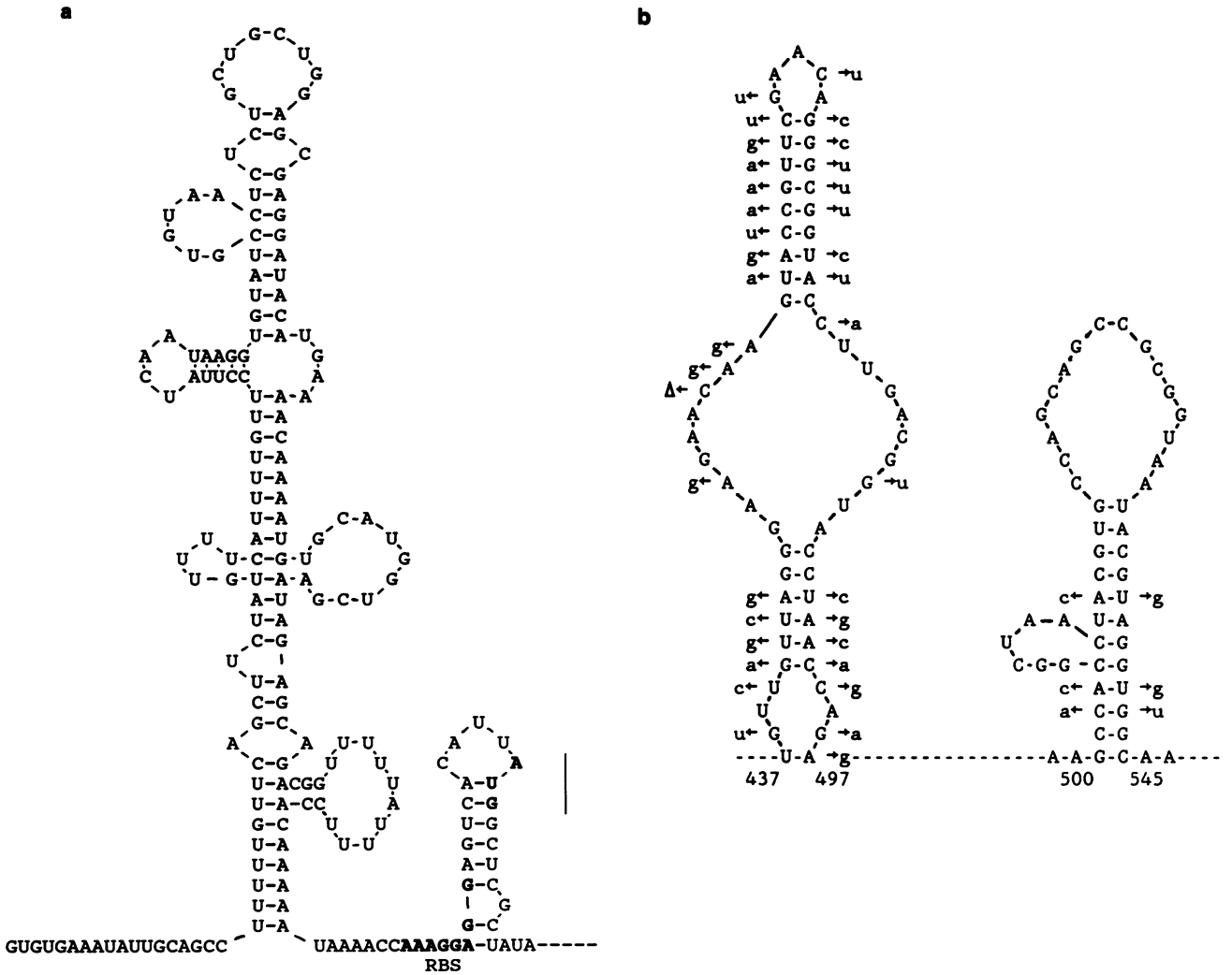

FIG. 6. Secondary-structure model for the leader region of *rpsD* mRNA and the S4-binding region of 16S rRNA. (a) *rpsD* leader region. This structure, derived from nucleotide positions 574 (the +1 position for transcription) to 768 (just past the *rpsD* translation initiation codon) is predicted to have a stability of −58 kcal (−243 kJ) (43). The AUG initiator codon for S4 (vertical line) and ribosome-binding site (RBS) are indicated. Alternate structures for this region can be drawn with similar predicted free energy. (b) 16S rRNA. Sequence of the region of *B. subtilis* 16S rRNA (*rrnB* operon sequence; 13) corresponding to helices 437–497 and 500–545 (*E. coli* numbering; 46) is shown, with arrows and lowercase letters indicating nucleotide differences in *E. coli* 16S rRNA.

gene expression in *E. coli*, this result suggests that regulation of these genes in *B. subtilis* may operate by a different mechanism.

The results presented here indicate that the *rpsD* gene is monocistronic in *B. subtilis*. Transcription originated approximately 180 bp upstream of the S4 translational start, and two sequences resembling Rho-independent transcriptional terminators were located immediately downstream of the translational stop codon for S4. Northern analysis revealed a single *rpsD* transcript, 0.8 kb in length, indicating that the first terminatorlike structure is used in vivo. There was an additional transcriptional terminatorlike structure upstream of the *rpsD* promoter. These additional structures flanking *rpsD* may represent signals for termination of transcription of upstream and downstream genes. The identity of the upstream open reading frames remains to be established, whereas the open reading frame immediately downstream of *rpsD* has been identified as the *tyrS* gene, encoding tyrosyl-tRNA synthetase. Since the *tyrS* gene is in the orientation opposite to that of *rpsD*, it is very unlikely that there are additional genes in the *rpsD* transcriptional unit.

In *E. coli*, S4 regulates α operon ribosomal protein gene expression by binding to a target site on the α operon mRNA overlapping the translation initiation region of the S13 gene, the first gene in the operon (9–11). Binding of S4 to this site is believed to inhibit translation of the mRNA. The S4 target site on α operon mRNA has been studied in detail and has been proposed to form a complicated pseudoknot structure (34). We therefore examined the putative *rpsD* mRNA leader region for sequences with the capacity to form similar structures. This sequence has the potential to form extensive secondary structure, according to the algorithm of Zuker and Stiegler (43) (Fig. 6a). Although there is little sequence similarity to the S4-binding site on *E. coli* α operon mRNA, there may be conservation of structural elements; the capacity to form structures which involve the ribosome-binding site and translational start codon is particularly intriguing. It is not possible at this time to predict whether the structural features of the leader region are involved in regulation of *B. subtilis rpsD* at the translational level; studies on the regulation of *rpsD* expression are in progress.

In several of the ribosomal protein regulatory systems in *E. coli*, the regulatory ribosomal protein has been found to be a protein that binds to ribosomal RNA early in ribosome assembly (20). The target site on the mRNA has been shown in some cases to resemble the binding site for the regulatory ribosomal protein on rRNA (5, 22). In the case of protein S4, the target sites on the mRNA and on 16S rRNA are apparently structurally complex (31, 32, 34, 37), and it has been difficult to identify structural similarities with the S4-binding site on 16S rRNA. Protein S4 is one of the first proteins to bind to 16S rRNA in 30S subunit assembly, initiating the binding of a large set of secondary proteins (27). The binding site for S4 on 16S rRNA in *E. coli* has been analyzed by a number of techniques and has been shown to involve positions 27 to 556, with interactions clustered in helices 437–497 and 500–545 (32). Although binding studies have not been done with *Bacillus* ribosomes, the homology of *E. coli* and *B. subtilis* S4 proteins, especially in the region of S4 known to be involved in binding to 16S rRNA (Fig. 3a), as well as the ability of *Bacillus* and *E. coli* ribosomal proteins and rRNA to be interchanged in heterologous reconstitution studies (17) and the high level of conservation of 16S rRNA sequence (13), suggest that S4 binding to *B. subtilis* 16S rRNA is similar to that of *E. coli*. Comparison of the predicted S4-binding sites on *E. coli* and *B. subtilis* 16S rRNAs (Fig. 6b) shows that although there is low sequence conservation between *B. subtilis* and *E. coli* in helix 437–497 (*E. coli* numbering; 40), the overall structure is maintained in helices 437–497 and 500–545 while the unpaired regions show high sequence conservation. These results suggest that the structure, rather than the sequence, is most important in helix 436–497 of the 16S rRNA, whereas helix 500–540 is very highly conserved.

There appears to be significant conservation of structure between the *rpsD* leader region and helix 500–545 of 16S rRNA (Fig. 6). We also note that the sequence at positions 656 to 663 (GCUGCUGG) of the *B. subtilis rpsD* region, within the leader mRNA, is complementary to the 16S rRNA sequence at positions 518 to 525 (CCAGCAGC), which is within the predicted S4-binding site. It will be of interest to examine the possible role of these sequence elements in future studies on *rpsD* gene expression in *B. subtilis*.

ACKNOWLEDGMENTS

We thank Brian Glass and Lynn Pittman-Cooley for technical assistance, John Wills for assistance with DNA sequence analysis, and Glenn Chambliss and Peter Zuber for critical reading of the manuscript.

This work was supported by Public Health Service grant GM-40650 from the National Institutes of Health and American Cancer Society Junior Faculty research award JFRA-226 to T.H.

LITERATURE CITED

1. **Barker, D. G., C. J. Bruton, and G. Winter.** 1982. The tyrosyl-tRNA synthetase from *Escherichia coli*. FEBS Lett. **150**:419–423.
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**:3891–3902.
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. **Cerretti, D. P., D. Dean, G. R. Davis, D. M. Bedwell, and M. Nomura.** 1983. The *spc* ribosomal protein operon of *Escherichia coli*: sequence and cotranscription of the ribosomal protein genes and a protein export gene. Nucleic Acids Res. **11**:2599–2616.
5. **Cerretti, D. P., L. Mattheakis, K. R. Kearney, L. Vu, and M. Nomura.** 1988. Translational regulation of the *spc* operon in *Escherichia coli*. Identification and structural analysis of the target site for S8 repressor protein. J. Mol. Biol. **204**:309–329.
6. **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.
7. **Conrad, R. C., and G. R. Craven.** 1987. A cyanogen bromide fragment of S4 that specifically rebinds 16S rRNA. Nucleic Acids Res. **15**:10331–10343.
8. **Dean, D., and M. Nomura.** 1980. Feedback regulation of ribosomal protein gene expression in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **77**:3590–3594.
9. **Deckman, I. C., and D. E. Draper.** 1985. Specific interaction between ribosomal protein S4 and the α operon messenger RNA. Biochemistry **24**:7860–7865.
10. **Deckman, I. C., and D. E. Draper.** 1987. S4-α mRNA translation regulation complex. II. Secondary structure of the RNA regulatory site in the presence and absence of S4. J. Mol. Biol. **196**:323–332.
11. **Deckman, I. C., D. E. Draper, and M. S. Thomas.** 1987. S4-α mRNA translation repression complex. I. Thermodynamics of formation. J. Mol. Biol. **196**:313–322.
12. **Graves, M. C., and J. C. Rabinowitz.** 1986. *In vivo* and *in vitro* transcription of the *Clostridium pasteurianum* ferredoxin gene.

Evidence for "extended" promoter elements in Gram-positive organisms. J. Biol. Chem. **261**:11409–11415.
13. **Green, C. J., G. C. Stewart, M. A. Hollis, B. S. Vold, and K. F. Bott.** 1985. Nucleotide sequence of the *Bacillus subtilis* RNA operon, *rrnB*. Gene **37**:261–266.
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., G. H. Chambliss, and F. J. Grundy.** 1990. *Bacillus subtilis* mutants with alterations in ribosomal protein S4. J. Bacteriol. **172**:6380–6385.
16. **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.
17. **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.
18. **Higo, K., E. Otaka, and S. Osawa.** 1982. Purification and characterization of 30S ribosomal proteins from *Bacillus subtilis*: correlation to *Escherichia coli* 30S proteins. Mol. Gen. Genet. **185**:239–244.
19. **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.
20. **Jinks-Robertson, S., and M. Nomura.** 1987. Ribosomes and tRNA, p. 1358–1385. *In* F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: molecular and cellular biology. American Society for Microbiology, Washington, D.C.
21. **Jones, M. D., D. M. Lowe, T. Borgford, and A. R. Fersht.** 1986. Natural variation of tyrosyl-tRNA synthetase and comparison with engineered mutants. Biochemistry **25**:1887–1891.
22. **Kearney, K. R., and M. Nomura.** 1987. Secondary structure of the autoregulatory mRNA binding site of ribosomal protein L1. Mol. Gen. Genet. **210**:60–68.
23. **McLaughlin, J. R., C. L. Murray, and J. C. Rabinowitz.** 1981. Unique features in the ribosome binding site sequence of the Gram-positive *Staphyloccus aureus* β-lactamase gene. J. Biol. Chem. **256**:11283–11291.
24. **Meek, D. W., and R. S. Hayward.** 1984. Nucleotide sequence of the *rpoA-rplQ* DNA of *Escherichia coli*: a second regulatory binding site for protein S4? Nucleic Acids Res. **12**:5813–5821.
25. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
26. **Nakano, M. M., and P. Zuber.** 1989. Cloning and characterization of *srfB*, a regulatory gene involved in surfactin production and competence in *Bacillus subtilis*. J. Bacteriol. **171**:5347–5353.
27. **Noller, H. F., and M. Nomura.** 1987. Ribosomes, p. 104–125. *In* F. C. Neidhardt, J. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: molecular and cellular biology. American Society for Microbiology, Washington, D.C.
28. **Piggot, P. J.** 1989. Revised genetic map of *Bacillus subtilis* 168, p. 1–42. *In* I. Smith, R. A. Slepecky, and P. Setlow (ed.), Regulation of procaryotic development. American Society for Microbiology, Washington, D.C.
29. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of transcription. Annu. Rev. Genet. **13**:319–353.
30. **Schiltz, E., and J. Reinbolt.** 1975. Determination of the complete amino-acid sequence of protein S4 from *Escherichia coli* ribosomes. Eur. J. Biochem. **56**:467–481.
31. **Stern, S., B. Weiser, and H. F. Noller.** 1988. Model for the three-dimensional folding of 16S ribosomal RNA. J. Mol. Biol. **204**:447–481.
32. **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.
33. **Suh, J.-W., S. A. Boylan, and C. W. Price.** 1986. Gene for the alpha subunit of *Bacillus subtilis* RNA polymerase maps in the ribosomal protein gene cluster. J. Bacteriol. **168**:65–71.
34. **Tang, C. K., and D. E. Draper.** 1989. Unusual mRNA pseudoknot structure is recognized by a protein translational repressor. Cell **57**:531–536.
35. **Thomas, M. S., D. M. Bedwell, and M. Nomura.** 1987. Regulation of α operon expression in *Escherichia coli*. A novel form of translational coupling. J. Mol. Biol. **196**:333–345.
36. **van Acken, U.** 1975. Protein chemical studies on ribosomal proteins S4 and S12 from *ram* (ribosomal ambiguity) mutants of *E. coli*. Mol. Gen. Genet. **140**:61–68.
37. **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.
38. **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.
39. **Winter, G., G. L. E. Koch, B. S. Hartley, and D. G. Barker.** 1983. The amino acid sequence of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus*. Eur. J. Biochem. **132**:383–387.
40. **Woese, C. R., R. R. Gutel, R. Gupta, and H. F. Noller.** 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol. Rev. **47**:621–669.
41. **Wu, J.-J., M. G. Howard, and P. J. Piggot.** 1989. Regulation of transcription of the *Bacillus subtilis spoIIA* locus. J. Bacteriol. **171**:692–698.
42. **Yates, J. L., A. E. Arfsten, and M. Nomura.** 1980. In vitro expression of *Escherichia coli* ribosomal protein genes—autogenous inhibition of translation. Proc. Natl. Acad. Sci. USA **77**:1837–1841.
43. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133–148.