Volume 11 Number 9 1983 Nucleic Acids Research

The *spc* ribosomal protein operon of *Escherichia coli:* sequence and cotranscription of the ribosomal protein genes and a protein export gene

Douglas Pat Cerretti, Dennis Dean*, Geneva R.Davis, David M.Bedwell and Masayasu Nomura

Institute for Enzyme Research and the Departments of Biochemistry and Genetics, University of Wisconsin, Madison, WI 53706, USA

Received 18 February 1983; Revised and Accepted 11 April 1983

ABSTRACT

The genes encoding the 52 ribosomal proteins (r-proteins) of Escherichia coli are organized into approximately 19 operons scattered throughout the chromosome. One of these, the spc operon, contains the genes for ten ribosomal proteins: L14, L24, L5, S14, S8, L6, L18, S5, L30 and L15 (rplN, rplX, rplE, rpsN, rpsH, rplF, rplR, rpsE, rpmD, and rplO). We now report the entire 5.9 kb nucleotide sequence of the spc operon. DNA sequence analysis has confirmed the genetic organization and refined the amino acid sequence of the ten r-proteins in this operon. It has also revealed the presence of two open reading frames past the last known gene (L15) of the spc operon. One of these corresponds to a gene (prlA or secY) which recently has been shown by others to be involved in protein export. In addition, S1 mapping experiments indicate that a significant proportion of transcription initiated from the spc operon continues not only into the two putative genes, but also without termination into the downstream α r-protein operon.

INTRODUCTION

The Escherichia coli ribosome contains 52 unique proteins (r-proteins) and three species of RNA (rRNA) (1-3). Genes coding for 27 of the r-proteins as well as several other related genes are clustered in the 72 minute region of the E. coli genome (4,5). The results of a detailed biochemical and genetic analysis (4-6) of this region suggested that these genes could be organized into four transcriptional units. These transcriptional units have been called "operons" and were given the trivial names str, S10, spc, and α based on the genetic or biochemical character of one of their gene products.

Even though these operons have been the subject of intensive study for a number of years, their complete organization and the physiological regulation of their expression is still not completely understood. The application of modern genetic tools that allows a more sophisticated analysis of these operons requires a knowledge of their nucleotide sequence. In the present study the entire nucleotide sequence of the spc operon, including the region between the spc operon and the α operon, was determined. In addition, the nature of the mRNA transcript covering the distal portion of the spc operon, the proximal portion of the α operon, and the region between the two operons was examined by S1 mapping experiments. The spc operon was chosen for such a detailed analysis for the following reasons:

A) Preliminary DNA sequence analysis of the region just preceding the α operon promoter did not reveal a transcription termination structure for the preceding spc operon (7). Furthermore, earlier analysis of the spc and α operons suggested that spc promoter activity might contribute significantly to α operon gene expression (6, 8).

B) Silhavy and his coworkers (9, 10) discovered that mutation of a gene within or very near the spc operon affects protein export. Examination of preliminary DNA sequence information in the region between the α and spc operons revealed possible reading frames that could code for two proteins whose function has not been determined (7). Possibly one or both of these genes codes for proteins that participate in protein export. It is therefore of interest to determine their complete nucleotide sequence and determine if their expression is under the control of the spc promoter.

C) Earlier studies have demonstrated that translation of r-protein mRNA is regulated by an autogenous feedback mechanism (11, for review see 12). The spc operon r-proteins are apparently subdivided into three independently regulated translational units (13 and P. Singer, D. Dean, L. Matthews, and M. Nomura, unpublished experiments). Elucidation of the entire nucleotide sequence of the spc operon allows an examination of possible secondary mRNA structures that could be important in the translational regulation process. In addition, a knowledge of the spc operon sequence will facilitate the construction of hybrid plasmids and/or phages that can be used to test models that attempt to explain how the translational regulation of r-protein mRNA is accomplished.

D) Determination of the complete nucleotide sequence of the spc operon would confirm and refine the amino acid sequence of ten r-proteins determined by others (14-23) using peptide sequencing procedures.

EXPERIMENTAL PROCEDURES

Strains. Bacteriophages M13mp7 (24), M13mp8, M13mp9 (25), plasmid pUC9 (26), and their *E. coli* host JM103 (24) were obtained from Bethesda Research Laboratories (BRL). Methods for their propagation and maintenance were as described by literature supplied by BRL and J. Messing (personal communication). *E. coli* K12 strain GM1, used for S1 mapping experiments, is ara Δ(lac pro) thi/F' lac^+ pro^+. The lac region on the F' carries $lacI^q$ and the lacP mutation, L8 (see 27).

DNA. The ultimate source of DNA for sequencing was λfus3 and λspc1 (4, 28). Cloning of the 5 kb EcoRI fragment (see Fig. 1) has been described (29). This plasmid, pNO1001, was used as a source of DNA to isolate the 5 kb EcoRI fragment. The 1.8 kb EcoRI fragment (see Fig. 1) from λspc1 was cloned into pUC9, designated pNO10558, and was used for the isolation of the 1.8 kb fragment. The 1.9 kb BamH1/Sal1 fragment from λspc1 contains the genes for L14 and L24 and a part of the L5 gene (30). The fragment was cloned into pBR322, designated pNO1507, and used as a source of DNA for sequencing the L14-L24 gene region. Restriction endonucleases and DNA ligase were obtained from BRL or New England Biolabs. Use of enzymes was as recommended by their suppliers.

Cloning into M13. Various portions of the 5 kb EcoRI and 1.8 kb EcoRI fragments were subcloned into M13mp7, M13mp8 and M13mp9 using the restriction enzymes indicated in Fig. 1. In regions where the same insert was isolated in both directions but were too large to be completely sequenced (such as the HincII clones), further subcloning was carried out using a method suggested by C. Squires (personal communication). Equimolar amounts of single stranded DNA from two M13 clones (with an insert in opposite orientations) were incubated in 50 mM Tris HCl, pH 8.0, 50 mM NaCl and 10 mM $MgCl_2$ for 30 min at 68°C. The sample was adjusted to 30 mM sodium acetate, pH 4.5 and 1 mM $ZnSO_4$ and treated with S1 nuclease (2 units/μg DNA for 30 min at 45°C. Digestion was terminated and DNA recovered by ethanol precipitation. The resulting double stranded linear DNA was used directly for further subcloning into M13 using suitable restriction enzymes (such as TaqI and Sau3A).

DNA Sequencing. DNA sequencing was done by the chain termination method of Sanger (31) as described by the BRL sequencing manual. DNA polymerase I, Klenow fragment, was obtained from either BRL, New England Biolabs or P-L Biochemicals with no detectable differences. The single stranded synthetic universal primer, 5'-CCCAGTCACGACGTT-3', was obtained from P-L Biochemicals. Sequencing samples were run on 6% polyacrylamide gels, 0.4 mm thick, containing 7M urea, 100 mM tris borate, pH 8.1 and 2 mM EDTA. Resolution of the sequencing gels was increased by using thinner, 0.25 mm, gels (spacers and combs made from Cardex, Central Plastics Distributors Co., Madison, WI) and by fixing (10% acetic acid for 10 min) and drying (75°C, 1 h) the gels before autoradiography (32). The DNA sequence of the L14 and L24 genes was determined by the method of Maxam and Gilbert (33) as described previously (7) using pNO1507 as a source of DNA. DNA sequence analysis was done with

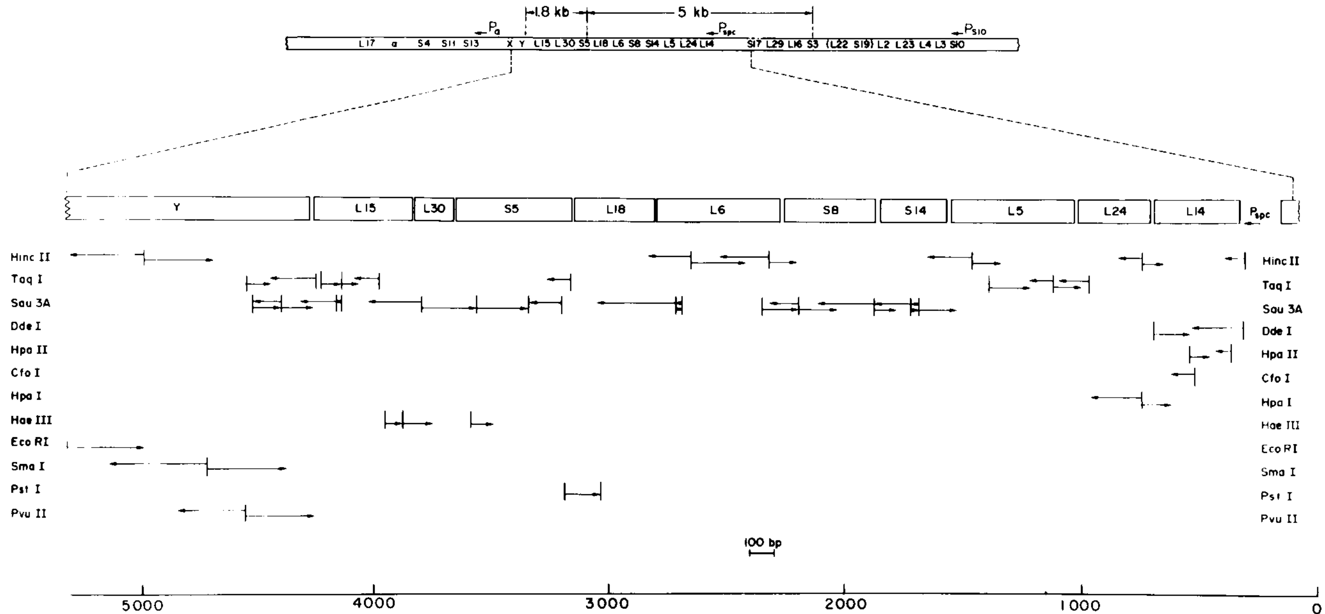

Figure 1
Map of the S10, spc, and α operons and sequencing strategy for the spc operon. The locations of r-protein genes in the S10 operon, the spc operon and the α operon are shown in the upper bar together with the locations of the promoters. The locations of the 5 kb EcoRI fragment and the 1.8 kb EcoRI fragment are also indicated. Sequence determination was done by the method of Sanger (31) except for L14 and L24 which were done by the method of Maxam and Gilbert (33). Only relevant restriction enzyme sites are shown. Each horizontal arrow represents the region covered by a single sequencing experiment. The scale at the bottom represents the numbering of the nucleotide bases used in Figure 2.

computer programs by Staden (34) and with programs designed by John Devereux of the University of Wisconsin Genetics Computer Group.

S1 mapping. S1 mapping was performed by the method of Berk and Sharp (35) as described by Squires et al. (36). Single stranded M13mp8 or M13mp9 hybridization probes containing various segments of the region between the spc and α operons (see Fig. 3A) were cloned in each of the two possible orientations. RNA was extracted from strain GM1 grown in AB glucose (0.4%) supplemented with casamino acids (0.1%) (37) and hybridized to the M13 hybridization probes. After treatment with S1 nuclease, samples were run on 5% polyacrylamide gels and stained with ethidium bromide. In each case, the RNA/DNA hybrids visualized were found only when using a DNA probe whose orientation was complementary to mRNA.

RESULTS AND DISCUSSION

DNA sequence and the organization of r-protein genes in the spc operon. The nucleotide sequence of the promoter region of the spc operon was previously reported (38). The sequence covered a distal part of the S17 gene (the last gene of the S10 r-protein operon) and a proximal part of the L14 gene (the first gene of the spc operon). The promoter region of the α operon was also sequenced previously (7) including approximately 550 nucleotides preceding the transcription start site of the α operon. We have now sequenced the DNA segment (about 5,000 nucleotides long) which is between these two previously determined promoter regions, completing the nucleotide sequence of the entire spc operon.

DNA derived from two sources, λfus3 and λspc1 (4, 28) was used for the sequence work. Fig. 1 shows the experiments used to determine the DNA sequence and Fig. 2 shows the results obtained together with the previously reported sequences of the promoter regions mentioned above. The DNA sequence for the intercistronic region between the L24 gene and the L5 gene was also previously determined (39), and the results are used to complete the spc operon sequences shown in Fig. 2. We have sequenced 70% of the coding regions for r-protein genes from two or more M13 clones, and alignment with the published amino acid sequences of the ten spc operon r-proteins (14-23) gives further support to the reliability of the DNA sequence results. Sequences from other regions were determined from both strands in over 90% of the cases, otherwise they were determined by two independently isolated M13 clones containing the same insert DNA.

Biochemical (5) and genetic experiments (4, 7) performed previously in

```
                                                                         END
                                                                         S17
  I  R  E  C  R  P  L  S  K  T  K  S  W  T  L  V  R  V  V  E  K  A  V  L  *
AATCCGCGAATGCCGTCCGCTGTCCAAGACTAAATCCTGGACGCTGGTTCGCGTTGTAGAGAAAGCGGTTCTGTAATACAGTACACTCTCTCAATACGAATAAACGGCTCAGAAATGAGC
         10        20        30        40        50        60        70        80        90       100       110       120

                                             Pspc
CGTTTATTTTTTCTACCCATATCCTTGAAGCGGTGTTATAATGCCGCGCCCTCGATATGGGGATTTTTAACGACCTGATTTCGGGTCTCAGTAGTAGTTGACATTAGCGGAGCCTAAAA
        130       140       150       160       170       180       190       200       210       220       230       240
START
L14
M  I  Q  E  Q  T  M  L  N  V  A  D  N  S  G  A  R  R  V  M  C  I  K  V  L  G  G  S  H  R  R  Y  A  G  V  G  D  I  I  K  I
TGATCCAAGAACAGACTATGCTGAACGTCGCCGACAACTCCGGTGCACGTCGCGTAATGTGTATCAAGGTTCTAGGGGGCTCGCACCGACGATACGCAGGCGTAGGCGACATCATCAAGA
        250       260       270       280       290       300       310       320       330       340       350       360

  T  I  K  E  A  I  P  R  G  K  V  K  K  G  D  V  L  K  A  V  V  V  R  T  K  K  G  V  R  R  P  D  G  S  V  I  R  F  D  G
TCACCATCAAAGAAGCAATTCCGCGTGGTAAGGTCAAAAAAGGTGATGTGCTGAAGGCGGTAGTGGTGCGCACCAAGAAGGGTGTTCGTCGCCCGGACGGTTCTGTCATTCGCTTCGATG
        370       380       390       400       410       420       430       440       450       460       470       480

  N  A  C  V  L  L  N  N  N  S  E  Q  P  I  G  T  R  I  F  G  P  V  T  R  E  L  R  S  E  K  F  M  K  I  I  S  L  A  P  E
GTAATGCTTGTGTTCTTCTGAACAACAACAGCGAGCAGCCTATCGGTACGCGTATTTTTGGGCCGGTAACTCGTGAGCTTCGTAGTGAGAAGTTCATGAAAATTATCTCTCTGGCACCAG
        490       500       510       520       530       540       550       560       570       580       590       600
           END         START
           L14         L24
  V  L  *              M  A  A  K  I  R  R  D  D  E  V  I  V  L  T  G  K  D  K  G  K  R  G  K  V  K  N  V  L  S  S  G  K
AAGTACTCTAAGGAGCGAATCATGGCAGCGAAAATCCGTCGTGATGACGAAGTTATCGTGTTAACCGGTAAAGATAAAGGTAAACGCGGTAAAGTTAAGAATGTCCTGTCTTCCGGCAAG
        610       620       630       640       650       660       670       680       690       700       710       720

V  I  V  E  G  I  N  L  V  K  K  H  Q  K  P  V  P  A  L  N  Q  P  G  G  I  V  E  K  E  A  A  I  Q  V  S  N  V  A  I  F
GTCATTGTTGAAGGTATCAACTTGGTTAAGAAACATCAGAAGCCGGTTCCGGCCTTGAACCAACCGGGTGGCATCGTTGAAAAAGAAGCCGCTATTCAGGTTTCCAACGTAGCAATCTTC
        730       740       750       760       770       780       790       800       810       820       830       840
                                                                                                 END             START
                                                                                                 L24              L5
N  A  A  T  G  K  A  D  R  V  G  F  R  F  E  D  G  K  K  V  R  F  F  K  S  N  S  E  T  I  K  *                 M  A  K  L
AATGCGGCAACCGGCAAGGCTGACCGTGTAGGCTTTAGATTCGAAGACGGTAAAAAAGTCCGTTTCTTCAAGTCTAACAGCGAAACTATCAAGTAATTTGGAGTAGTACGATGGCGAAAC
        850       860       870       880       890       900       910       920       930       940       950       960

  H  D  Y  Y  K  D  E  V  V  K  K  L  M  T  E  F  N  Y  N  S  V  M  Q  V  P  R  V  E  K  I  T  L  N  M  G  V  G  E  A  I
TGCATGATTACTACAAAGACGAAGTAGTTAAAAAACTCATGACTGAGTTTAACTACAATTCTGTCATGCAAGTCCCTCGGGTCGAGAAGATCACCCTGAACATGGGTGTTGGTGAAGCGA
        970       980       990      1000      1010      1020      1030      1040      1050      1060      1070      1080

  A  D  K  K  L  L  D  N  A  A  A  D  L  A  A  I  S  G  Q  K  P  L  I  T  K  A  R  K  S  V  A  G  F  K  I  R  Q  G  Y  P
TCGCTGACAAAAAACTGCTGGATAACGCAGCAGCAGACCTGGCAGCAATCTCCGGTCAAAAACCGTTGATCACCAAAGCACGCAAATCTGTTGCAGGCTTCAAAATCCGTCAGGGCTATC
       1090      1100      1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
```

```
  I  G  C  K  V  T  L  R  G  E  R  M  W  E  F  F  E  R  L  I  T  I  A  V  P  R  I  R  D  F  R  G  L  S  A  K  S  F  D  G
CGATCGGCTGTAAAGTAACTCTGCGTGGCGAACGCATGTGGGAGTTCTTTGAGCGCCTGATCACTATTGCTGTACCTCGTATCCGTGACTTCCGTGGCCTGTCCGCTAAGTCTTTCGACG
      1210      1220      1230      1240      1250      1260      1270      1280      1290      1300      1310      1320

  R  G  N  Y  S  M  G  V  R  E  Q  I  I  F  P  E  I  D  Y  D  K  V  D  R  V  R  G  L  D  I  T  I  T  T  T  A  K  S  D  E
GTCGTGGTAACTACAGCATGGGTGTCCGTGAGCAGATCATCTTCCCAGAAATCGACTACGATAAAGTCGACCGCGTTCGTGGTTTGGACATTACCATTACCACTACTGCGAAATCTGACG
      1330      1340      1350      1360      1370      1380      1390      1400      1410      1420      1430      1440
                                            END              START
                                            L5                S14
  E  G  R  A  L  L  A  A  F  D  F  P  F  R  K  *                M  A  K  Q  S  M  K  A  R  E  V  K  R  V  A  L  A  D  K
AAGAAGGCCGCGCTCTGCTGGCTGCCTTTGACTTCCCGTTCCGCAAGTAAGGTAGGGTTACTAAATGGCTAAGCAATCAATGAAAGCACGCGAAGTAAAGCGCGTAGCTTTAGCTGATAA
      1450      1460      1470      1480      1490      1500      1510      1520      1530      1540      1550      1560

  Y  F  A  K  R  A  E  L  K  A  I  I  S  D  V  N  A  S  D  E  D  R  W  N  A  V  L  K  L  Q  T  L  P  R  D  S  S  P  S  R
ATACTTCGCGAAACGCGCTGAACTGAAAGCGATCATCTCTGATGTGAACGCTTCCGACGAAGATCGTTGGAACGCTGTTCTCAAGCTGCAGACTCTCCCCCGTGATTCCAGCCCGTCTCG
      1570      1580      1590      1600      1610      1620      1630      1640      1650      1660      1670      1680

  Q  R  N  R  C  R  Q  T  G  R  P  H  G  F  L  R  K  F  G  L  S  R  I  K  V  R  E  A  A  M  R  G  E  I  P  G  L  K  K  G
TCAGCGTAACCGTTGCCGTCAAACAGGTCGTCCGCATGGTTTCCTGCGAAAGTTTGGGTTGAGCCGTATTAAGGTCCGTGAAGCCGCTATGCGCGGTGAAATCCCGGGTCTGAAAAAAGG
      1690      1700      1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
END                                     START
S14                                      S8
 *                                       M  S  M  Q  D  P  I  A  D  M  L  T  R  I  R  N  G  Q  A  A  N  K  A  A  V
CTAGCTGGTAATTGTCACCAATTGAATCACGGGAGGTAAAGACAGATGAGCATGCAAGATCCGATCGCGGATATGCTGACCCGTATCCGTAACGGTCAGGCCGCGAACAAAGCTGCGGTC
      1810      1820      1830      1840      1850      1860      1870      1880      1890      1900      1910      1920

 T  M  P  S  S  K  L  K  V  A  I  A  N  V  L  K  E  E  G  F  I  E  D  F  K  V  E  G  D  T  K  P  E  L  E  L  T  L  K  Y
ACCATGCCTTCCTCCAAGCTGAAAGTGGCAATCGCCAACGTGCTGAAGGAAGAAGGTTTTATTGAAGATTTTAAAGTTGAAGGCGACACCAAGCCTGAACTGGAACTTACTCTGAAGTAT
      1930      1940      1950      1960      1970      1980      1990      2000      2010      2020      2030      2040

 F  Q  G  K  A  V  V  E  S  I  Q  R  V  S  R  P  G  L  R  I  Y  K  R  K  D  Q  L  P  K  V  M  A  G  L  G  I  A  V  V  S
TTCCAGGGCAAAGCTGTTGTTGAAAGCATTCAGCGTGTCAGCCGCCCAGGCTTGCGCATCTATAAACGTAAAGATCAGCTGCCCAAAGTTATGGCGGGTCTGGGTATCGCAGTTGTTTCT
      2050      2060      2070      2080      2090      2100      2110      2120      2130      2140      2150      2160
                                                                         END          START
                                                                         S8            L6
 T  S  K  G  V  M  T  D  R  A  A  R  Q  A  G  L  G  G  E  I  I  C  Y  V  A  *              M  S  R  V  A  K  A  P  V  V
ACCTCTAAAGGTGTTATGACTGATCGTGCAGCGCGCCAAGCTGGTCTTGGTGGCGAAATTATCTGCTACGTAGCCTAATCGGACGAAAAAATGTCTCGTGTTGCTAAAGCACCGGTCGTT
      2170      2180      2190      2200      2210      2220      2230      2240      2250      2260      2270      2280

 V  P  A  G  V  D  V  K  I  N  G  Q  V  I  T  I  K  G  K  N  G  E  L  T  R  T  L  N  D  A  V  E  V  K  H  A  D  N  T  L
GTTCCTGCCGGCGTTGACGTAAAAATCAACGGTCAGGTTATTACGATCAAAGGTAAAAACGGCGAGCTGACTCGTACTCTCAACGATGCTGTTGAGGTTAAACATGCAGATAATACCCTG
      2290      2300      2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
```

```
 T  F  G  P  R  D  G  Y  A  D  G  W  A  Q  A  G  T  A  R  A  L  L  N  S  M  V  I  G  V  T  E  D  F  T  K  K  L  Q  L  V
ACCTTTGGTCCGCGTGATGGTTACGCAGACGGTTGGGCACAGGCTGGTACCGCGCGTGCCCTGCTGAACTCAATGGTTATCGGTGTTACCGAGGACTTCACTAAGAAGCTGCAGCTGGTT
      2410      2420      2430      2440      2450      2460      2470      2480      2490      2500      2510      2520

 G  V  G  Y  R  A  A  V  K  G  N  V  I  N  L  S  L  G  F  S  H  P  V  D  H  Q  L  P  A  G  I  T  A  E  C  P  T  Q  T  E
GGTGTAGGTTACCGTGCAGCGGTTAAAGGGAATGTGATTAACCTGTCTCTGGGTTTCTCTCATCCTGTTGACCATCAGCTGCCTGCGGGTATCACTGCTGAATGTCCGACTCAGACTGAA
      2530      2540      2550      2560      2570      2580      2590      2600      2610      2620      2630      2640

 I  V  L  K  G  A  D  K  Q  V  I  G  Q  V  A  A  D  L  R  A  Y  R  R  P  E  P  Y  K  G  K  G  V  R  Y  A  D  E  V  V  R
ATCGTGCTGAAAGGCGCTGATAAGCAGGTGATCGGCCAGGTTGCAGCGGATCTGCGCGCCTACCGTCGTCCTGAGCCTTATAAAGGCAAGGGTGTTCGTTACGCCGACGAAGTCGTCCGC
      2650      2660      2670      2680      2690      2700      2710      2720      2730      2740      2750      2760
                     END         START
                     L6           L18
 T  K  E  A  K  K  K  *           M  D  K  K  S  A  R  I  R  R  A  T  R  A  R  R  K  L  Q  E  L  G  A  T  R  L  V  V  H
ACCAAAGAGGCTAAGAAGAAGTAAGGTAACACTATGGATAAGAAATCTGCTCGTATCCGTCGTGCGACCCGCGCACGCCGCAAGCTCCAGGAGCTGGGCGCAACTCGCCTGGTGGTACAT
      2770      2780      2790      2800      2810      2820      2830      2840      2850      2860      2870      2880

 R  T  P  R  H  I  Y  A  Q  V  I  A  P  N  G  S  E  V  L  V  A  A  S  T  V  E  K  A  I  A  E  Q  L  K  Y  T  G  N  K  D
CGTACCCCGCGTCACATTTACGCACAGGTAATTGCACCGAACGGTTCTGAAGTTCTGGTAGCTGCTTCTACTGTAGAAAAAGCTATCGCTGAACAACTGAAGTACACCGGTAATAAAGAC
      2890      2900      2910      2920      2930      2940      2950      2960      2970      2980      2990      3000

 A  A  A  A  V  G  K  A  V  A  E  R  A  L  E  K  G  I  K  D  V  S  F  D  R  S  G  F  Q  Y  H  G  R  V  Q  A  L  A  D  A
GCGGCTGCAGCTGTGGGTAAAGCTGTCGCTGAACGCGCTCTGGAAAAAGGCATCAAAGATGTATCCTTTGACCGTTCCGGGTTCCAATATCATGGTCGTGTCCAGGCACTGGCAGATGCT
      3010      3020      3030      3040      3050      3060      3070      3080      3090      3100      3110      3120
                        END                START
                        L18                 S5
 A  R  E  A  G  L  Q  F  *                  M  A  H  I  E  K  Q  A  G  E  L  Q  E  K  L  I  A  V  N  R  V  S  K  T  V  K  G
GCCCGTGAAGCTGGCCTTCAGTTCTAAGGTAGAGGTGTAAGATGGCTCACATCGAAAAACAAGCTGGCGAACTGCAGGAAAAGCTGATCGCGGTAAACCGCGTATCTAAAACCGTTAAAG
      3130      3140      3150      3160      3170      3180      3190      3200      3210      3220      3230      3240

   G  R  I  F  S  F  T  A  L  T  V  V  G  D  G  N  G  R  V  G  F  G  Y  G  K  A  R  E  V  P  A  A  I  Q  K  A  M  E  K  A
GTGGTCGTATTTTCTCCTTCACAGCTCTGACTGTAGTTGGCGATGGTAACGGTCGCGTTGGTTTTGGTTACGGTAAAGCGCGTGAAGTTCCAGCAGCGATCCAGAAAGCGATGGAAAAAG
      3250      3260      3270      3280      3290      3300      3310      3320      3330      3340      3350      3360

   R  R  N  M  I  N  V  A  L  N  N  G  T  L  Q  H  P  V  K  G  V  H  T  G  S  R  V  F  M  Q  P  A  S  E  G  T  G  I  I  A
CCCGTCGCAATATGATTAACGTCGCGCTGAATAACGGCACTCTGCAACACCCTGTTAAAGGTGTTCACACGGGTTCTCGCGTATTCATGCAGCCGGCTTCCGAAGGTACCGGTATCATCG
      3370      3380      3390      3400      3410      3420      3430      3440      3450      3460      3470      3480

   G  G  A  M  R  A  V  L  E  V  A  G  V  H  N  V  L  A  K  A  Y  G  S  T  N  P  I  N  V  V  R  A  T  I  D  G  L  E  N  M
CCGGTGGTGCAATGCGCGCCGTTCTGGAAGTCGCTGGGGTTCATAACGTTCTCGCTAAAGCCTATGGTTCCACCAACCCGATCAACGTGGTTCGTGCAACTATTGATGGCCTGGAAAATA
      3490      3500      3510      3520      3530      3540      3550      3560      3570      3580      3590      3600
```

```
                                                                  END  START
                                                                  S5    L30
 N  S  P  E  M  V  A  A  K  R  G  K  S  V  E  E  I  L  G  K  *     M  A  K  T  I  K  I  T  Q  T  R  S  A  I  G  R  L  P
TGAATTCTCCAGAAATGGTCGCTGCCAAGCGTGGTAAATCCGTTGAAGAAATTCTGGGGAAATAAACCATGGCAAAGACTATTAAAATTACTCAAACCCGCAGTGCAATCGGTCGTCTGC
      3610      3620      3630      3640      3650      3660      3670      3680      3690      3700      3710      3720

 K  H  K  A  T  L  L  G  L  G  L  R  R  I  G  H  T  V  E  R  E  D  T  P  A  I  R  G  M  I  N  A  V  S  F  M  V  K  V  E
CGAAACACAAGGCAACGCTGCTTGGCCTGGGTCTGCGTCGTATTGGTCACACCGTAGAGCGCGAGGATACTCCTGCTATTCGCGGTATGATCAACGCGGTTTCCTTCATGGTTAAAGTTG
      3730      3740      3750      3760      3770      3780      3790      3800      3810      3820      3830      3840
   END  START
   L30   L15
 E  *     M  R  L  N  T  L  S  P  A  E  G  S  K  K  A  G  K  R  L  G  R  G  I  G  S  G  L  G  K  T  G  G  R  G  H  K  G
AGGAGTAAGAGATGCGTTTAAATACTCTGTCTCCGGCCGAAGGCTCCAAAAAGGCGGGTAAACGCCTGGGTCGTGGTATCGGTTCTGGCCTCGGTAAAACCGGTGGTCGTGGTCACAAAG
      3850      3860      3870      3880      3890      3900      3910      3920      3930      3940      3950      3960

 Q  K  S  R  S  G  G  G  V  R  R  G  F  E  G  G  Q  M  P  L  Y  R  R  L  P  K  F  G  F  T  S  R  K  A  A  I  T  A  E  I
GTCAGAAGTCTCGTTCTGGCGGTGGCGTACGTCGCGGTTTCGAGGGTGGTCAGATGCCTCTGTACCGTCGTCTGCCGAAATTCGGCTTCACTTCTCGTAAAGCTGCGATTACAGCCGAAA
      3970      3980      3990      4000      4010      4020      4030      4040      4050      4060      4070      4080

 R  L  S  D  L  A  K  V  E  G  G  V  V  D  L  N  T  L  K  A  A  N  I  I  G  I  Q  I  E  F  A  K  V  I  L  A  G  E  V  T
TTCGTTTGTCTGACCTGGCTAAAGTAGAAGGCGGTGTAGTAGACCTGAATACGCTGAAAGCGGCTAACATTATCGGTATCCAGATCGAGTTCGCGAAAGTGATCCTGGCTGGCGAAGTAA
      4090      4100      4110      4120      4130      4140      4150      4160      4170      4180      4190      4200
                                                                                                  END      START
                                                                                                  L15        Y
 T  P  V  T  V  R  G  L  R  V  T  K  G  A  R  A  A  I  E  A  A  G  G  K  I  E  E  *              M  A  K  Q  P  G  L  D  F
CGACTCCGGTAACTGTTCGTGGCCTGCGTGTTACTAAAGGCGCTCGTGCTGCTATCGAAGCTGCTGGCGGTAAAATCGAGGAATAAGTAGCAGATGGCTAAACAACCGGGATTAGATTTT
      4210      4220      4230      4240      4250      4260      4270      4280      4290      4300      4310      4320

 Q  S  A  K  G  G  L  G  E  L  K  R  R  L  L  F  V  I  G  A  L  I  V  F  R  I  G  S  F  I  P  I  P  G  I  D  A  A  V  L
CAAAGTGCCAAAGGTGGCTTAGGCGAGCTGAAACGCAGACTGCTGTTTGTTATCGGTGCGCTGATTGTGTTCCGTATTGGCTCTTTTATTCCGATCCCTGGTATTGATGCCGCTGTACTT
      4330      4340      4350      4360      4370      4380      4390      4400      4410      4420      4430      4440

 A  K  L  L  E  Q  Q  R  G  T  I  I  E  M  F  N  M  F  S  G  G  A  L  S  R  A  S  I  F  A  L  G  I  M  P  Y  I  S  A  S
GCCAAACTGCTTGAGCAACAGCGAGGCACCATCATTGAGATGTTTAACATGTTCTCTGGTGGTGCTCTCAGCCGTGCTTCTATCTTTGCTCTGGGGATCATGCCGTATATTTCGGCGTCG
      4450      4460      4470      4480      4490      4500      4510      4520      4530      4540      4550      4560

 I  I  I  Q  L  L  T  V  V  H  P  T  L  A  E  I  K  K  E  G  E  S  G  R  R  K  I  S  Q  Y  T  R  Y  G  T  L  V  L  A  I
ATCATTATCCAGCTGCTGACGGTGGTTCACCCAACGTTGGCAGAAATTAAGAAAGAAGGGGAGTCTGGTCGTCGTAAGATCAGCCAGTACACCCGCTACGGTACTCTGGTGCTGGCAATA
      4570      4580      4590      4600      4610      4620      4630      4640      4650      4660      4670      4680

 F  Q  S  I  G  I  A  T  G  L  P  N  M  P  G  M  Q  G  L  V  I  N  P  G  F  A  F  Y  F  T  A  V  V  S  L  V  T  G  T  M
TTCCAGTCGATCGGTATTGCTACCGGTCTGCCGAATATGCCTGGTATGCAAGGCCTGGTGATTAACCCGGGCTTTGCATTCTACTTCACCGCTGTTGTAAGTCTGGTCACAGGAACCATG
      4690      4700      4710      4720      4730      4740      4750      4760      4770      4780      4790      4800
```

```
 F  L  M  W  L  G  E  Q  I  T  E  R  G  I  G  N  G  I  S  I  I  I  F  A  G  I  V  A  G  L  P  P  A  I  A  H  T  I  E  Q
TTCCTGATGTGGTTGGGCGAACAGATTACTGAACGAGGTATCGGCAACGGTATTTCAATCATTATCTTCGCCGGTATTGTCGCGGGACTCCCGCCAGCCATTGCCCATACTATCGAGCAA
      4810      4820      4830      4840      4850      4860      4870      4880      4890      4900      4910      4920

 A  R  Q  G  D  L  H  F  L  V  L  L  L  V  A  V  L  V  F  A  V  T  F  F  V  V  F  V  E  R  G  Q  R  R  I  V  V  N  Y  A
GCGCGTCAAGGCGACCTGCACTTCCTCGTGTTGCTGTTGGTTGCAGTATTAGTATTTGCAGTGACGTTCTTTGTTGTATTTGTTGAGCGTGGTCAACGCCGCATTGTGGTAAACTACGCG
      4930      4940      4950      4960      4970      4980      4990      5000      5010      5020      5030      5040

 K  R  Q  Q  G  R  R  V  Y  A  A  Q  S  T  H  L  P  L  K  V  N  M  A  G  V  I  P  A  I  F  A  S  S  I  I  L  F  P  A  T
AAACGTCAGCAAGGTCGTCGTGTCTATGCTGCACAGAGCACACATTTACCGCTGAAAGTGAATATGGCGGGGGTAATCCCGGCAATCTTCGCTTCCAGTATTATTCTGTTCCCGGCGACC
      5050      5060      5070      5080      5090      5100      5110      5120      5130      5140      5150      5160

 I  A  S  W  F  G  G  G  T  G  W  N  W  L  T  T  I  S  L  Y  L  Q  P  G  Q  P  L  Y  V  L  L  Y  A  S  A  I  I  F  F  C
ATCGCGTCATGGTTCGGGGGCGGTACTGGTTGGAACTGGCTGACAACAATTTCGCTGTATTTGCAGCCTGGGCAACCGCTTTATGTGTTACTCTATGCGTCTGCAATCATCTTCTTCTGT
      5170      5180      5190      5200      5210      5220      5230      5240      5250      5260      5270      5280

 F  F  Y  T  A  L  V  F  N  P  R  E  T  A  D  N  L  K  K  S  G  A  F  V  P  G  I  R  P  G  E  Q  T  A  K  Y  I  D  K  V
TTCTTCTACACGGCGTTGGTTTTCAACCCGCGTGAAACAGCAGATAACCTGAAGAAGTCCGGTGCATTTGTACCAGGAATTCGTCCGGGAGAGCAAACGGCGAAGTATATCGATAAAGTA
      5290      5300      5310      5320      5330      5340      5350      5360      5370      5380      5390      5400

 M  T  R  L  T  L  V  G  A  L  Y  I  T  F  I  C  L  I  P  E  F  M  R  D  A  M  K  V  P  F  Y  F  G  G  T  S  L  L  I  V
ATGACCCGCCTGACCCTGGTTGGTGCGCTGTATATTACCTTTATCTGCCTGATCCCGGAGTTCATGCGTGATGCAATGAAAGTACCGTTCTACTTCGGTGGGACCTCACTGCTTATCGTT
      5410      5420      5430      5440      5450      5460      5470      5480      5490      5500      5510      5520
                                                                                                           END
                                                                                                            Y
 V  V  V  I  M  D  F  M  A  Q  V  Q  T  L  M  M  S  S  Q  Y  E  S  A  L  K  K  A  N  L  K  G  Y  G  R  *
GTTGTCGTGATTATGGACTTTATGGCTCAAGTGCAAACTCTGATGATGTCCAGTCAGTATGAGTCTGCATTGAAGAAGGCGAACCTGAAAGGCTACGGCCGATAATTGGTCGCCCGAGAA
      5530      5540      5550      5560      5570      5580      5590      5600      5610      5620      5630      5640
               START
                 X
                 M  K  V  R  A  S  V  K  K  L  C  R  N  C  K  I  V  K  R  D  G  V  I  R  V  I  C  S  A  E  P  K  H  K  Q
GTTACGGAGAGTAAAAATGAAAGTTCGTGCTTCCGTCAAGAAATTATGCCGTAACTGCAAAATCGTTAAGCGTGATGGTGTCATCCGTGTGATTTGCAGTGCCGAGCCGAAGCATAAACA
      5650      5660      5670      5680      5690      5700      5710      5720      5730      5740      5750      5760
         END
          X
 R  Q  G  *                                                   Palpha
GCGCCAAGGCTGATTTTTTCGCATATTTTTCTTGCAAAGTTGGGTTGAGCTGGCTAGATTAGCCAGCCAATCTTTTGTATGTCTGTGCGTTTCCATTTGAGTATCCTGAAAACGGGCTTT
      5770      5780      5790      5800      5810      5820      5830      5840      5850      5860      5870      5880
                                 START
                                  S13
                                   M
TCAGCATGGAACGTACATATTAAATAGTAGGAGTGCATAGTG
      5890      5900      5910      5920
```

this laboratory identified ten r-protein genes in the spc operon and indicated their order except for L30 and L15. The order of L30 and L15 was subsequently established by DNA sequencing across the site between the 5 kb EcoRI and the 1.8 kb EcoRI fragments (L. Post, Ph.D. thesis, University of Wisconsin, 1979). The present DNA sequence results have confirmed these previous conclusions regarding the presence of r-protein genes and their order. In addition, the results have revealed the presence of two open reading frames (Y and X) downstream of the last r-protein gene, L15, in the operon. These will be discussed below.

The DNA sequence has also confirmed and refined the amino acid sequences of the ten r-proteins of the spc operon. The amino acid sequences as deduced from the DNA sequence for L24, S8, L18, S5, L30 and L15 show complete agreement with published amino acid sequences (15, 18, 20, 21, 22, 23). However, discrepancies have been found for L14, L5, S14 and L6 (14, 16, 17, 19). In the case of L14 and L6, the amino acid sequences as deduced from the DNA sequences have been shown to agree with revised amino acid sequence data (H. G. Wittmann, personal communication). The discrepancies not resolved are: ile thr thr thr from the DNA sequence instead of ile thr thr at position 155 in L5 from the protein sequence, and ser from the DNA sequence instead of gly at position 98 in S14. Table 1 shows the codon usage of the spc operon r-proteins. The highly nonrandom codon usage is the same as that observed for other r-proteins (40). The codons preferentially used are those recognized by the most abundant tRNA species. This usage probably reflects the need to efficiently translate r-protein mRNA as has been discussed in more detail previously (40-44).

Cotranscription of the spc operon and α operon. The α operon, located immediately distal to the spc operon, was originally defined by analyzing expression of r-protein genes in UV irradiated cells infected with transducing phages that carry various portions of the 72 minute region of the *E. coli* chromosome (see Fig. 1 for structures of the spc and α operons). A DNA fragment containing the S13, S11 and S4 genes, but not the upstream spc operon genes, was subcloned into a charon vector phage (45). This phage could direct synthesis of S13, S11, and S4 in a λ lysogen and therefore

Figure 2
DNA sequence of the spc operon. According to the numbering system used previously, the first nucleotide is arbitrary set at the third base of the codon for amino acid 59 of r-protein S17, which is the last r-protein in the S10 operon (38). Some of these sequences have been published previously (7, 38, 39).

Table 1

Codon Usage for the Ribosomal Protein Genes of the Spc Operon and Reading Frame Y

Codons		r-Proteins	Y	Codons		r-Proteins	Y
phe	TTT	11	12	ser	TCT	27	6
phe	TTC	28	20	ser	TCC	17	3
leu	TTA	3	5	ser	TCA	2	3
leu	TTG	6	7	ser	TCG	1	4
leu	CTT	6	4	pro	CCT	13	3
leu	CTC	8	4	pro	CCC	2	0
leu	CTA	1	0	pro	CCA	5	3
leu	CTG	65	28	pro	CCG	24	14
ile	ATT	28	22	thr	ACT	30	5
ile	ATC	58	22	thr	ACC	26	10
ile	ATA	0	1	thr	ACA	3	5
met	ATG	35	15	thr	ACG	6	5
val	GTT	53	10	ala	GCT	53	10
val	GTC	22	4	ala	GCC	20	6
val	GTA	31	10	ala	GCA	37	12
val	GTG	14	11	ala	GCG	29	13
cys	TGT	4	1	tyr	TAT	6	8
cys	TGC	2	1	tyr	TAC	16	7
***	TGA	0	0	***	TAA	9	1
trp	TGG	3	4	***	TAG	1	0
arg	CGT	68	12	his	CAT	9	2
arg	CGC	35	5	his	CAC	8	2
arg	CGA	3	3	gln	CAA	13	12
arg	CGG	1	0	gln	CAG	29	9
ser	AGT	2	4	asn	AAT	13	2
ser	AGC	8	3	asn	AAC	34	8
arg	AGA	1	1	lys	AAA	79	10
arg	AGG	0	0	lys	AAG	42	7
gly	GGT	82	20	asp	GAT	27	5
gly	GGC	41	12	asp	GAC	27	2
gly	GGA	0	5	glu	GAA	54	5
gly	GGG	7	6	glu	GAG	21	9

Codons listed do not include the initiation codon
*based on initiation at nucleotide 4294 (see text)

provided evidence that these genes carry their own promoter. Furthermore, DNA sequencing and *in vitro* transcription experiments were used to identify a transcription start site and a promoter structure immediately preceding the S13 gene (7). Finally, deletion analysis of recombinant phages gave *in vivo* evidence for the presence of an α promoter activity in the region identified by the above *in vitro* experiments (30). Although there is compelling evidence that the α promoter does have the ability to be physiologically active, there is also evidence that the expression of the α operon genes might also be dependent (at least in part) on the activity of the *spc* promoter. Insertion mutations which nearly abolish expression of the *spc* operon genes also substantially reduce expression of α operon genes (up to 50%) (6, 8). Thus, the possibility was considered that the α promoter is a "secondary" promoter and is used in *spc* insertion mutants but not under normal physiological conditions.

After completion of the DNA sequence for the *spc* operon, the sequence was examined for possible transcription termination signals in the region between the last r-protein gene in the *spc* operon and the α promoter. No rho-independent transcription termination sequence of the type described previously was found (for example, see ref. 46). A similar situation was previously noted (40) in the sequence between the L11 operon (*rplKA* operon and the adjacent β operon (*rplJL*-*rpoBC* operon), and it was subsequently found that these two operons are in fact co-transcribed under "normal" growth conditions (47). We therefore searched for transcription termination sites between the *spc* operon and the α operon using S1 mapping techniques (35, 36).

M13 single stranded DNA probes were constructed that together span the region between the *spc* and α operons (see Fig. 3A). These probes were then used to hybridize total RNA isolated from exponentially growing cells. Treatment of such hybrid molecules with S1 nuclease removes single stranded RNA or DNA regions, leaving only RNA-DNA hybrid molecules intact. These hybrids can be sized on polyacrylamide or agarose gels and consequently a site(s) for transcription initiation, processing, or termination can be deduced based on the size of the S1 treated hybrid molecules. The results of such an experiment are shown in Fig. 3B. This experiment used two DNA probes (labelled A and B in Fig. 3A) that together span the region between the *spc* and α operons. The sizes of the most abundant RNA-DNA hybrids (bands I and III in Fig. 3B) correspond to the full length of each of the DNA probes. Similar results were obtained with probe C (data not shown). It is clear from these experiments that the majority of transcription originating in the *spc* operon continues into the α operon under normal growth conditions. However, there appears to be some transcription termination and/or processing in this region since several small minor bands could be detected (for example, bands II and IV in Fig. 3B). Although the nature of the minor

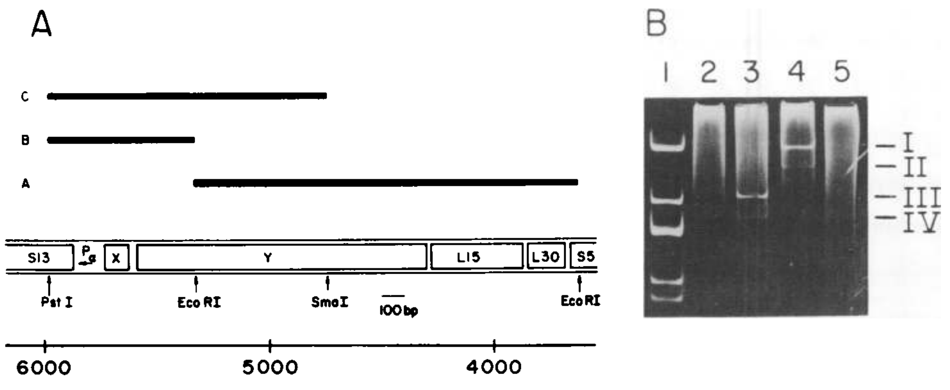

Figure 3
A) The interoperonic region between the spc and α operons. Included are the proposed reading frames X and Y. Bars A, B and C represent the DNA fragments cloned into M13mp8 and M13mp9 for S1 mapping experiments.
B) RNA/DNA hybrids. Lane 1: HaeII digest of pBR322, fragments sizes are 1875 bp, 622 bp, 439 bp, 430 bp, 370 bp and 227 bp; Lane 2: M13 probe B-, which contains the insert B with the same sense as mRNA; Lane 3: M13 probe B+, which contains the insert B complementary to mRNA; Lane 4: M13 probe A+, which contains the insert A complementary to mRNA; Lane 5: M13 probe A-, which contains the insert A with the same sense as mRNA. The scale at the bottom represents the numbering of the nucleotide bases used in Figure 2.

bands has not been examined, it appears that transcription initiated from the α promoter is small compared to readthrough from the spc operon. Thus it appears that α promoter activity, like β promoter activity (47), is suppressed in exponentially growing cells when transcription from the upstream genes is strong. A similar situation was reported with respect to the transcription from the P_L promoter of phage λ, and this phenomenon has been called "promoter occlusion" (48). It is not clear why transcription of two r-protein operons is carried out in this way rather than as two completely independent transcription units.

Recently this laboratory has shown that the r-protein genes in the spc operon are organized into three translational control units, while the α operon contains one (or possibly two) such units, (for review see 12). The S1 mapping experiments reported here indicate that transcription initiated at the spc promoter results in primarily one spc-α transcript, containing four or more translational control units. It would be interesting to know whether the S10 and spc operons are also co-transcribed (see Fig. 1). It has previously been noted (38) that a rho-independent transcription termination sequence exists distal to the S10 operon, suggesting that these operons are not co-transcribed. This possibility is currently under investigation.

Reading frames Y and X in the spc operon and a protein export gene. The previous sequence studies on the α operon promoter revealed two open reading frames upstream from the α operon promoter, one corresponding to a small basic protein (called "X") and the other (called "Y") spanning the EcoRI site (7; and Fig. 3A). However, it was not clear whether there are any additional open reading frames between L15 of the spc operon and the α operon, nor was it clear whether these presumptive genes, X and Y, are a part of the spc operon.

Recently, genetic experiments have been used to identify a gene, prlA (9) or secY, (49) that is located in this region and is involved in protein export. Complementation studies indicated that the functional prlA (or secY) gene covers both the SmaI site at nucleotide 4746 and the EcoRI site at nucleotide 5357 (see Fig. 2 and 3). Analysis of the DNA sequence in this region shows that there is only one reading frame that is not interrupted between these two sites, and this is reading frame Y. The pattern of codon usage in the Y reading frame (see Table 1) is characteristic of weakly or moderately expressed genes in E. coli such as the lacI (50) and trp operon (51) genes.

The actual initiation codon used to initiate the protein encoded in the Y reading frame has not been determined. However the results of genetic studies (9, 49) indicate the the initiation codon is upstream from the SmaI restriction enzyme site at position 4746. It seems likely that the initiation codon located at position 4294 is used because only this codon is preceded by a "good" Shine-Dalgarno sequence (52). Furthermore, Silhavy and his coworkers (10) have fused a portion of the lacZ gene and a portion of the Y coding region in a recombinant plasmid in the "correct" reading frame. Size analysis of the resultant fusion protein is consistent with the initiation codon being located at position 4294 but does not rule out other possibilities.

Examination of the DNA sequence does not reveal any promoter-like structure in or following the L15 gene that could be used for the initiation of the transcription of the prlA (or secY) gene. In addition, the S1 mapping experiments described above show that the region corresponding to Y is, at least in part, co-transcribed with the preceding r-protein genes. These results are consistent with the conclusion obtained from the genetic experiments (9, 49) that the prlA (secY) gene is cotranscribed with the r-protein genes in the spc operon.

As discussed in the previous paper on the α operon promoter sequence (7), it is likely that the reading frame X also codes for a protein _in vivo_. Although nothing is known about the (hypothetical) protein encoded by X, we note that the proteins encoded by both X and Y are basic and hydrophobic (their relative hydrophobicity is similar to membrane-bound proteins, see ref. 53 for calculations) as deduced from the DNA sequences. It has been suggested from the genetic studies (9, 49) that protein Y may play a function in protein export that is analogous to a protein in the signal recognition particle discovered in eukaryotic systems (54, 55). The eukaryotic signal recognition particle consists of a small RNA and several protein subunits (55). Therefore, it is possible that the hypothetical protein encoded by X also functions as a part of the protein export machinary analogous to the signal recognition particle. This hypothesis is under current investigation.

In conclusion, we have completed the nucleotide sequence of the region that historically has been called the _spc_ operon. There are two interesting, although not completely unexpected, observations that have emerged. Firstly, in addition to the ten r-proteins contained within the _spc_ operon, there is at least one and possibly two other genes whose products are involved in protein export. Whether these proteins are as yet undiscovered r-proteins is not known. Secondly, we have demonstrated that the majority of the α operon transcripts originates within the _spc_ operon. The α operon contains genes that code for at least five (and perhaps more) proteins, including four r-proteins and the α subunit of RNA polymerase. Thus the transcriptional unit that includes the _spc_ and α operons contains at least 17 genes with a variety of functions, including roles in translation, protein export, and transcription. Clearly this unit of gene expression is remarkable not only for its size but also for the heterogeneity of its gene products. The DNA sequence now available allows a direct examination of the many questions posed by this unique organization.

ACKNOWLEDGMENTS

This work was supported in part by the College of Agricultural and Life Sciences, University of Wisconsin-Madison, by Grant GM-20427 from the National Institutes of Health, and by Grant PCM79-10616 from the National Science Foundation. D.D. was supported by National Institutes of Health Postdoctoral Fellowship GMO 7553. This is paper number 2627 from the Laboratory of Genetics, University of Wisconsin-Madison.

*Present address: Charles F. Kettering Research Laboratory, Yellow Springs, OH 45387

REFERENCES

1. Nomura, M. and Post, L. E. (1980) in Ribosomes: Structure, Function and Genetics. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L.,and Nomura, M., eds. pp. 641-669, University Park Press, Baltimore.
2. Isono, K. (1980) in Ribosomes: Structure, Function and Genetics. Chambliss, G., Craven, G. R., Davies, J., Davis, K., Kahan, L. and Nomura, M., eds. pp. 267-294, University Park Press, Baltimore.
3. Lindahl, L. and Zengel, J. M. (1982) Advances in Genetics 21, 53-121.
4. Jaskunas, S. R., Fallon, A. M., and Nomura, M. (1977) J. Biol. Chem. 252, 7323-7336.
5. Lindahl, L., Post, L., Zengel, J., Gilbert, S. F., Strycharz, W. A. and Nomura, M. (1977) J. Biol. Chem. 252, 7365-7383.
6. Jaskunas, S. R. and Nomura, M. (1977) J. Biol. Chem. 252, 7337-7343.
7. Post, L. E., Arfsten, A. E., Davis, G. R. and Nomura, M. (1980). J. Biol. Chem. 255, 4653-4659.
8. Jaskunas, S. R., Lindahl, L. and Nomura, M. (1975) Nature 256 183-187.
9. Emr, S. D., Hamley-Way, S., and Silhavy, T. J. (1981) Cell 23 79-88.
10. Shultz, J., Silhavy, T. J., Berman, M. L., Fiil, N. and Emr, S. D. (1982) Cell 31, 227-235
11. Nomura, M., Yates, J. L., Dean, D. and Post, L. E. (1980) Proc. Natl. Acad. Sci. USA 77, 7084-7088.
12. Nomura, M., Dean, D., and Yates, J. L. (1982) Trends in Biochemical Sciences 7, 92-95.
13. Dean, D., Yates, J. L. and Nomura, M. (1981) Nature 289, 89-91.
14. Morinaga, T., Funatsu, G., Funatsu, M., Wittmann-Liebold, B., and Wittmann, H. G. (1978) FEBS Lett. 91, 74-77.
15. Wittmann-Liebold, B. (1979) FEBS Lett. 108, 75-80.
16. Chen, R. and Ehrke, G. (1976) FEBS Lett. 69, 240-245.
17. Wittmann, H. G., personal communication for the amino acid sequence of S14.
18. Allen, G. and Wittman-Liebold, B. (1978) Hoppe-Seyler's Z. Physiol. Chem. 359, 1509-1525.
19. Chen, R., Afrsten, U., and Chen-Schmeisser (1977) Hoppe-Seyler's Z. Physiol. Chem. 358, 531-535.
20. Brosius, J., Schiltz, E., and Chen, R. (1975). FEBS Lett. 56, 359-361.
21. Wittmann-Liebold, B. and Greuer, B. (1978) FEBS Lett. 95, 91-98.
22. Ritter, E. and Wittmann-Liebold, B. (1975) FEBS Lett. 60, 153-155.
23. Giorginis, S. and Chen, R. (1977) FEBS Lett. 84, 347-350.
24. Messing, J., Crea, R., and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309-321.
25. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
26. Vieira, J. and Messing, J. (1982) Gene 19,259-268.
27. Coulondre, C. and Miller, J. H. (1977) J. Mol. Biol. 117, 525-575.
28. Jaskunas, S. R., Lindahl, L. and Nomura, M. (1975) Proc. Natl. Acad. Sci. USA 72, 6-10.
29. Fallon, A. M., Jinks, C. S., Yamamoto, M. and Nomura, M. (1979) J. Bacteriol. 138, 383-396.
30. Miura, A., Krueger, J. H., Itoh, S., de Boer, H. A. and Nomura, M. (1981) Cell 25, 773-782.
31. Sanger, F., Wicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
32. Garoffe, H. and Ansorge, W. (1981) Anal. Biochem. 115, 450-457.

33. Maxam, A. M. and Gilbert, W. (1977) Natl. Acad. Sci. USA 74, 560-564.
34. Staden, R. (1980) Nucleic Acids Res. 8, 3673-3694.
35. Berk. A. J. and Sharp, P. A. (1978) Proc. Natl. Acad. Sci. USA 75, 1274- 1278.
36. Squires, C., Krainer, A., Barry, G., Shen, W.-F., and Squires, C. L. (1981) Nucleic Acids Res. 9, 6827-6840.
37. Clark, D. J. and Maaløe, O. (1967) J. Mol. Biol. 23, 99-112.
38. Post, L. E., Arfsten, A. E., Reusser, F. and Nomura, M. (1978) Cell 15, 215-229.
39. Olins, P. O. and Nomura, M. (1981) Nucleic Acids Res. 9, 1757-1764.
40. Post, L. E., Strycharz, G. D., Nomura, M., Lewis, H. and Dennis, P. P. (1979) Proc. Natl. Acad. Sci. USA 76, 1697-1701.
41. Post, L. E. and Nomura, M. (1980) J. Biol. Chem. 255, 4660-4666.
42. Ikemura, T. (1981) Cell 151, 389-409.
43. Grosjean, H. and Fiers, W. (1982) Gene 18, 199-209.
44. Goury, M. and Gautier, C. (1982) Nucleic Acids Res. 10, 7055-7074.
45. Jaskunas, S. R., Fallon, A. M., Nomura, M., Williams, B. G. and Blattner, F. R. (1977) J. Biol. Chem. 252, 7355-7364.
46. Platt, T. (1981) Cell 24, 10-23.
47. Bruckner, R. and Matzura, H. (1981) Mol. Gen. Genet. 183, 277-282.
48. Adhya, S. and Gottesman, M. (1982) Cell 29, 939-944.
49. Ito, K., Wittekind, M., Nomura, M., Miura, A., Shiba, K., Yura, T. and Nashimoto, H. (1983) Cell, in press.
50. Farabaugh, P. J. (1978) Nature 274, 765-769.
51. Yanofsky, C., Platt, T., Crawford, I. P., Nichols, B. P., Christie, G. E., Horowitz, H., Van Cleemput, M. and Wu, A. M. (1981) Nucleic Acids Res. 9, 6647-6668.
52. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA 71, 1342- 1346.
53. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132.
54. Walter, P. and Blobel, G. (1980) Proc. Natl. Acad. Sci. USA 77, 7112- 7116.
55. Walter, P. and Blobel, G. (1982) Nature 299, 691-698.