Nucleotide sequence of the alpha ribosomal protein operon of *Escherichia coli*

David Bedwell, Geneva Davis*, Mark Gosink, Leonard Post+, Masayasu Nomura, Harry Kestler†, Janice M.Zengel† and Lasse Lindahl†

Institute for Enzyme Research and Departments of Biochemistry and Genetics, University of Wisconsin, Madison, WI 53706, and †Department of Biology, University of Rochester, Rochester, NY 14627, USA

Received 8 April 1985; Accepted 13 May 1985

ABSTRACT

In *Escherichia coli* some 19 transcription units encoding the 52 ribosomal proteins are scattered throughout the genome. One of the units, the alpha operon, encodes genes for the ribosomal proteins S13, S11, S4 and L17 as well as the alpha subunit of RNA polymerase. We report here the complete 3.0 kb nucleotide sequence of the alpha operon. In addition, we have determined by S1 nuclease mapping the site of transcription termination in this operon.

INTRODUCTION

The *Escherichia coli* ribosome consists of 52 ribosomal proteins (r-proteins) and 3 rRNA species. The genes for the r-proteins are organized into about 19 transcription units mapping at different chromosomal locations (1). Twenty-seven of these r-protein genes are located in a region traditionally referred to as the str-spc region at 72 minutes of the *E. coli* genetic map (2,3). This region is organized into 4 transcription units, termed the str, S10, spc, and alpha "operons" (2-4). Recent experiments have shown that during exponential growth the spc and alpha operons are at least partially co-transcribed (5), suggesting that the alpha operon promoter previously identified (6-8) may not function in exponentially growing cells.

In addition to genes for r-proteins, both the spc and the alpha operons are endowed with information for proteins which are not believed to be directly involved in protein synthesis. The spc operon contains in addition to genes for 10 r-proteins a gene (pr1A or secY) implicated in protein secretion (9,10). The alpha operon harbors the gene for the alpha subunit of RNA polymerase (11) as well as the genes for r-proteins S13, S11, S4 and L17. To the extent that the spc and alpha operons are cotranscribed, this region thus represents an 8.6 kb transcription unit encoding 16 proteins (and possibly an additional protein called X; refs. 5,7,12) involved in...
transcription, translation and protein secretion.

The nucleotide sequence of the entire spc operon and parts of the alpha operon have previously been published (5,7,13,14). We have now completed the nucleotide sequence of the alpha operon. We have also located the site of transcription termination just distal to the L17 gene, the last known gene of the alpha operon. Interestingly, the apparent termination signal shows extensive homology with the putative transcription termination signal of the beta-operon which encodes two other subunits of the RNA polymerase.

MATERIALS AND METHODS

Strains

Bacteriophages M13mp8, M13mp9, M13mp10, M13mp11, M13mp18 and M13mp19 (15-17) were grown in JM103 (18) as described in the M13 cloning manual published by Bethesda Research Laboratories, Inc. The strain JM103 (pN02530) was used as a source of RNA for the S1 mapping experiments. The plasmid pN02530 contains the alpha and L17 genes (Fig. 1) expressed from the lac promoter on this plasmid (D. Bedwell and M. Nomura, unpublished experiments).

DNA Sources

The ultimate sources of DNA used for sequencing were λ fus3, λ spc1 and λ spc2 (2,50). The 2.5 kb EcoRI fragment obtained from λ fus3 or λ spc1 (3) as well as the 1.7 kb PstI fragment from λ spc2 (7) were used to sequence the S13, S11 and S4 genes. The sequence of the distal half of the alpha gene and most of the L17 gene was obtained from a 0.9 kb EcoRI/PstI (14,19) fragment from λ spc2A16 (2) or λ spc1. A 3.2 kb PstI fragment from λ spc2 or λ fus3 (14,19) was used as the DNA source for sequencing the end of the L17 gene as well as the termination region. Various segments of the above DNA fragments were subcloned into M13mp8, M13mp9, M13mp10, M13mp11, M13mp18 or M13mp19, using the restriction enzymes indicated in Figure 1. Restriction enzymes and DNA ligase were purchased from New England Biolabs, Bethesda Research Laboratories, or Boehringer Mannheim, or were prepared by Eric Lifson in Rochester.

DNA Sequencing

Most of the DNA sequencing was done by the dideoxynucleotide chain termination method of Sanger et al. (20), in some cases as modified for [35S]nucleotides by Biggin et al. (21). Klenow fragment of DNA polymerase I was obtained from New England Biolabs, Boehringer Mannheim or IBI Biochemicals. The M13 universal primer was obtained from P.L. Biochemicals.
Sequencing gels were prepared as described previously (5,22). The sequences of portions of the S11 and S4 genes and the terminator region were determined both by the method of Maxam and Gilbert (23) and the dideoxy method. The transcription termination region contains an inverted repeat which can form a hairpin structure (see below). This caused problems with premature termination in the primer extension reactions used for the dideoxy sequencing as well as with compression of bands in the sequencing gels. Some of these problems were alleviated by substituting dITP for dGTP in the primer extension reaction mixtures as suggested by Hills and Kramer (24). Furthermore we confirmed the sequence of the DNA containing the possible hairpin structure by dideoxy sequencing of the PstI-HpaII and Rsal-HpaII fragments positioned on either side of HpaII site in the loop at the top of the hairpin (position 8802). DNA sequence analysis was done with programs provided through the University of Wisconsin Genetics Computer Group, or with the microcomputer sequence analysis programs of Pustell (25,26).

**S1 Nuclease Mapping**

In the S1 mapping of the transcription termination site, the DNA fragment (see legend to Fig. 3) was 3' end labeled with Klenow fragment (27), incubated with RNA and then digested with S1 nuclease by the method of Berk and Sharp (28), essentially as described by Barry et al. (29). The products were run on a DNA sequencing gel next to the same radioactively labeled DNA fragment treated with the chemical reactions of Maxam and Gilbert (23). RNA for this experiment was prepared from a culture of JM103 (pNO2530) grown in LB medium supplemented with ampicillin at 50 µg/ml. The same results were obtained with RNA from cells grown in the presence or absence of 1.5 mM IPTG (D. Bedwell and M. Nomura, unpublished). RNA was purified by the hot phenol method (30). S1 nuclease was purchased from Bethesda Research Laboratories.

**RESULTS AND DISCUSSION**

**DNA Sequence of the alpha Operon**

We present here the completed DNA sequence of the alpha operon. Three segments of this operon have previously been reported (Fig. 1). These are: (a) the promoter region through the proximal 109 base pairs of the S13 gene (7), (b) the distal 180 base pairs of the S4 gene through the first 378 base pairs of the alpha gene (13) and (c) the distal part of the alpha gene and the entire L17 gene (but not the terminator region following the L17 gene; ref. 14). We now add to these published sequences new data that complete...
Fig. 1. Map of the spc and alpha operons and sequencing strategy for the alpha operon. The locations of the promoters and r-protein genes of the spc and alpha operons are shown in the upper bar. Sequence determination was done by the dideoxy method of Sanger (20). Portions of the S11 and S4 genes, as well as the terminator region, were also done by the method of Maxam and Gilbert (23). Only relevant restriction enzyme sites are indicated. Horizontal arrows indicate the direction and distance sequenced from a single restriction site. The scale at the bottom represents the numbering of the nucleotide bases used in Fig. 2 and is a continuation of the numbering of the spc operon sequence (5).

The sequence of the entire operon, including the transcription terminator region following the last gene of the operon (the L17 gene). Our sequence deviates at one point from that presented by Meek and Hayward (14). For this reason and for the convenience of the reader, we report the sequence of the entire alpha operon.

DNA derived from λspcl, λspc2 or λfus3 (2) was sequenced employing the strategy shown in Fig. 1 and in the Materials and Methods. The complete nucleotide sequence of the alpha operon is presented in Fig. 2. All of the new sequence data presented here were derived either from both DNA strands or from independent, overlapping clones of the same strand. In addition, most of the new sequences have been determined independently by both the laboratories contributing to this publication.
# Nucleic Acids Research

## Alpha

<table>
<thead>
<tr>
<th>Position</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>7301</td>
<td>Asp</td>
</tr>
<tr>
<td>7501</td>
<td>Ala</td>
</tr>
<tr>
<td>7701</td>
<td>Gly</td>
</tr>
<tr>
<td>7901</td>
<td>Glu</td>
</tr>
<tr>
<td>8101</td>
<td>Thr</td>
</tr>
<tr>
<td>8301</td>
<td>Leu</td>
</tr>
</tbody>
</table>

## Beta

<table>
<thead>
<tr>
<th>Position</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>7401</td>
<td>Ser</td>
</tr>
<tr>
<td>7601</td>
<td>Val</td>
</tr>
<tr>
<td>7801</td>
<td>Glu</td>
</tr>
<tr>
<td>8001</td>
<td>Pro</td>
</tr>
<tr>
<td>8201</td>
<td>Ser</td>
</tr>
<tr>
<td>8401</td>
<td>Thr</td>
</tr>
<tr>
<td>8601</td>
<td>Val</td>
</tr>
</tbody>
</table>

---

**Sequences:**

**Alpha**

```
AGCTTTACTCCAACTAACCTTAGTACCAAAGAGACACACACAATOCAGOCRCTGTAGACAGAGTTTCTAAAACCCreGCCTG
```

**Beta**

```
AACMACACCTCATCCTCATC&AGCTTTACTCCAACTAACCTTAGTACCAAAGAGACACACACAATOCAGOCRCTGTAGACAGAGTTTCTAAAACCCreGCCTG
```

---

**Annotations:**

- **Alpha** sequence is annotated with amino acid positions from 7301 to 7400.
- **Beta** sequence is annotated with amino acid positions from 7401 to 7500.

---

**References:**

- [Nucleic Acids Research](https://nucleic-acids-research.nature.com)
The nucleotide sequence confirms the gene organization of the alpha operon previously determined from biochemical and genetic experiments (2,7,11-13). In addition, it confirms and refines the published amino acid sequences for the five proteins encoded by the operon. The DNA sequence of the S13, S11, alpha and L17 genes show complete agreement with the published amino acid sequences of the protein products of these genes (31-34). However, we have found several discrepancies between the published amino acid sequence of the S4 protein (35) and the amino acid sequence deduced from the nucleotide sequence of the gene for this protein. Thus our DNA sequence predicts (a) an additional leu residue between amino acids 89 and 90; (b) glu at amino acid 93 instead of gln; (c) ser at amino acid 136 instead of asp; (d) asp at amino acid 139 instead of ser; (e) an additional ser residue between amino acids 141 and 142; (f) gln at amino acid 149 instead of glu; and (g) glu at amino acid 163 instead of gln (Fig. 2). The latter two discrepancies were noted previously (13).

We also found one discrepancy with the DNA sequence of the alpha/L17 intergenic region published by Meek and Hayward (14). They reported a C between the two G's at positions 8368 and 8369 of our sequence. Our results indicate that this C is not present. Each of the corrections mentioned above have been confirmed by sequencing both DNA strands as well as by independent experiments in Madison and Rochester.

The frequency of codon usage in the five genes of the alpha operon is shown in Table 1. Both the r-protein genes and the alpha gene show the highly non-random codon usage that has previously been observed for r-protein genes and other highly expressed genes (5,22,36). The preferentially used codons correspond to the most abundant tRNA species. The possible significance of this observation has previously been discussed (36-40).

Identification of the 3' End of the alpha Operon Transcript

The precise location of transcription termination at the end of the Fig. 2. DNA sequence of the alpha operon. The numbering system is a continuation of that used for the DNA sequence of the spc operon (5). Parts of this sequence have been published previously (7,13,14). Transcripts initiated at the alpha promoter start at position 5806 (7). The differences between the amino acid sequence inferred from our DNA sequence and the amino acid sequence published for r-protein S4 (35) have been indicated with asterisks (***) above the corrected sequence. The differences indicated by asterisks in parentheses were previously noted (13). We also note one discrepancy between our DNA sequence and the one previously published (14) in the intergenic region between the alpha and L17 genes [indicated by a +] at position 8369. See text for further details.
Table 1. Codon Usage for the Alpha Subunit of RNA Polymerase and the Ribosomal Protein Genes of the Alpha Operon

<table>
<thead>
<tr>
<th>Codons</th>
<th>r-Proteins</th>
<th>Alpha</th>
<th>Codons</th>
<th>r-Proteins</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly GGG</td>
<td>1</td>
<td>1</td>
<td>Trp TGG</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Gly GGA</td>
<td>1</td>
<td>1</td>
<td>End TAG</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Gly GGT</td>
<td>30</td>
<td>9</td>
<td>Cys TGT</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Gly GCC</td>
<td>15</td>
<td>9</td>
<td>Cys TGC</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Glu GAG</td>
<td>15</td>
<td>17</td>
<td>End TAG</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Glu GAA</td>
<td>21</td>
<td>19</td>
<td>End TAA</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Asp GAT</td>
<td>12</td>
<td>13</td>
<td>Tyr TAT</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Asp GAC</td>
<td>10</td>
<td>8</td>
<td>Tyr TAC</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Val GTG</td>
<td>8</td>
<td>11</td>
<td>Leu TTG</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Val GTA</td>
<td>7</td>
<td>4</td>
<td>Leu TTA</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Val GTT</td>
<td>20</td>
<td>10</td>
<td>Phe TTT</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Val GTC</td>
<td>4</td>
<td>5</td>
<td>Phe TTC</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Ala GCG</td>
<td>9</td>
<td>6</td>
<td>Ser TCG</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Ala GCA</td>
<td>24</td>
<td>7</td>
<td>Ser TCA</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Ala GCT</td>
<td>18</td>
<td>7</td>
<td>Ser TCT</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Ala GCC</td>
<td>12</td>
<td>3</td>
<td>Ser TCC</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Arg AGG</td>
<td>0</td>
<td>0</td>
<td>Arg CGG</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Arg AGA</td>
<td>2</td>
<td>1</td>
<td>Arg CGA</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Ser AGT</td>
<td>3</td>
<td>1</td>
<td>Arg CGT</td>
<td>49</td>
<td>14</td>
</tr>
<tr>
<td>Ser AGC</td>
<td>9</td>
<td>4</td>
<td>Arg CGC</td>
<td>18</td>
<td>7</td>
</tr>
<tr>
<td>Lys AAG</td>
<td>25</td>
<td>3</td>
<td>Gln CAG</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>Lys AAA</td>
<td>23</td>
<td>13</td>
<td>Gln CAA</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Asn AAT</td>
<td>4</td>
<td>2</td>
<td>His CAT</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Asn AAC</td>
<td>20</td>
<td>6</td>
<td>His CAC</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Met ATG</td>
<td>9</td>
<td>4</td>
<td>Leu CTG</td>
<td>31</td>
<td>23</td>
</tr>
<tr>
<td>Ile ATA</td>
<td>1</td>
<td>0</td>
<td>Leu CTA</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Ile ATT</td>
<td>1</td>
<td>10</td>
<td>Leu CTT</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Ile ATC</td>
<td>22</td>
<td>14</td>
<td>Leu CTC</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Thr AGG</td>
<td>3</td>
<td>3</td>
<td>Pro CCG</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Thr ACA</td>
<td>1</td>
<td>2</td>
<td>Pro CCA</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Thr ACT</td>
<td>14</td>
<td>4</td>
<td>Pro CCT</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Thr ACC</td>
<td>9</td>
<td>11</td>
<td>Pro CCC</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Codons listed do not include initiation codons.

alpha operon was determined by S1 nuclease mapping (29,41). A 139 base pair HinPI/RsaI DNA fragment labeled at the 3' end of the HinPI site (Fig. 3B) was hybridized with total cellular RNA made from cells carrying the Plac-alpha-L17 plasmid pN02530 (see Materials and Methods). We used RNA from both uninduced and isopropylthiogalactoside induced cells. In either case the 3'-ends of the alpha operon transcripts map in the run of U's of a
structure typical of a Rho-independent terminator (42). The results from the experiment with RNA from the uninduced cells is shown in Fig. 3A. Since no full-length hybridization probe was protected by RNA against S1 nuclease degradation, it is likely that essentially all transcription terminates at the indicated site. Note that the same termination sites were observed whether or not the plasmid borne Plac-alpha-L17 operon was induced. Thus transcripts initiated both at the normal chromosomal promoter and at the lac promoter terminate at the same positions.

Interestingly, the terminator structure at the end of the alpha operon is preceded by another possible stem-loop structure (Fig. 3C) in a configuration very similar to that proposed for the end of the transcripts of the operon coding for the beta and beta' subunits of RNA polymerase (43; Fig. 3D). The similarity includes the identical sequence GAGUAAUC, with the UAA stop codon (for the last genes of the operons) located in the loop of the stem-loop structure preceding the terminator structures. In addition, the terminator stem-loop structures themselves are very similar with 10 identical bases pairs in the stems and loops of 4 bases. Transcripts from the operon encoding the sigma subunit of RNA polymerase also terminate at a Rho-independent terminator which is preceded by another stem-loop structure (41). However, the tandem stem-loop structures of the sigma operon differ both in structure and spacing from the corresponding structures of the alpha and beta operons. We do not know the significance of the very similar tandem stem-loop structures found in the alpha and the beta operons. In the case of the beta operon, it was previously suggested that the stem-loop structure may help to prevent degradation of mRNA by 3' exonucleases (43; see also 44). This suggestion may also apply to the alpha operon transcript. However, as mentioned above, practically all of the alpha operon mRNA terminates at the consecutive U's at the distal side of the second stem-loop structure. No transcripts of shorter length can be detected, making it unclear why the first stem-loop should be involved in protection against exonucleases.

Regulation of Protein Synthesis

Although the gene for the alpha subunit of RNA polymerase is cotranscribed with r-protein genes, the regulation of alpha synthesis is different from that of r-protein synthesis under some conditions. For example, r-protein synthesis is subject to stringent control during amino acid limitation, whereas the synthesis of alpha, like the synthesis of the other RNA polymerase subunits beta, beta' and sigma, is not subject to this
Fig. 3. S1 nuclease mapping of the transcription termination region of the alpha operon. The probe used is shown in (B). A 192 base pair HinPI fragment was 3' end labeled using [32P]dGTP and Klenow fragment of DNA polymerase I. It was then cut with Rsal, and the purified 139 base pair HinPI/Rsal fragment was used as a hybridization probe. The results from such an experiment are shown in (A): Lane 1 shows the probe alone; lane 2 shows the probe hybridized to yeast RNA; lane 3 shows the probe hybridized to total RNA from E. coli JM103 (pNO2530) grown in LB without IPTG (essentially the same result was obtained with RNA from the same strain grown in LB with IPTG). The next two lanes show the same probe treated by the G and G+A sequence reactions of Maxam and Gilbert (23). Our
interpretation of these results are shown in (C), with arrows designating
the termination points of the alpha operon mRNA. (C) also shows a probable
secondary structure of the terminator region. The ΔG values calculated
according to the revised Tinoco rules (49) for the proximal and distal stem-
loop structures are -9.3 kcal/mol and -23.5 kcal/mol, respectively.
(D) shows a similar secondary structure previously proposed for the end of
the operon encoding the beta and beta' subunits of RNA polymerase (43). The
calculated ΔG values for the proximal and distal stem-loop structures are
-14.2 kcal/mol and -17.4 kcal/mol, respectively.

control (45,46; D. Bedwell and M. Nomura, unpublished). In the case of the
beta and beta' genes, which are also cotranscribed with r-protein genes,
differential regulation could be explained by attenuation in the region
between the beta gene and the preceding r-protein genes (29,47). In the
case of the gene for the sigma subunit, the presence of both an attenuator
and a secondary promoter have been demonstrated and could account, at least
in part, for the differences in regulation of the sigma gene and the
upstream S21 gene (41).

The beta, beta' and sigma genes are located distal to the r-protein
genes in their respective transcription units. The alpha gene is unique in
the sense that it is located between r-protein genes. Therefore it is not
clear how this gene could be regulated differently from the flanking r-
protein genes. We have examined the nucleotide sequence for possible clues
to this differential regulation but failed to find any sequence which might
act as an attenuator or a secondary promoter. We have also performed S1
nuclease mapping with probes covering the region between 6028 (in the S13
gene) and the termination site to see if we could detect any transcripts
which could reflect initiation or RNA processing within the alpha operon.
No such transcripts were identified using RNA from either exponentially
growing cells or from cells harvested during amino acid starvation
conditions where differential regulation of alpha and r-protein synthesis
was observed (D. Bedwell and M. Nomura, unpublished observations). These
results therefore suggest that the differential regulation of alpha and r-
protein synthesis during amino acid limitation may take place at the level
of translation.

The autogenous regulation of the alpha operon also results in a
differential effect on the synthesis of the r-proteins and alpha. Ribosomal
protein S4, the product of the third gene of the alpha operon, inhibits the
translation of the three r-protein cistrons on the proximal side as well as
the L17 cistron on the distal side of the alpha gene (48). However, alpha
synthesis is only partially affected by S4 (1,51). Inspection of the DNA
sequence of the alpha operon offers no obvious clues to the differential regulation of alpha and r-protein synthesis. Hopefully, the DNA sequence of the entire alpha operon presented in this paper will be useful for future studies of these and other questions related to the regulation of the genes in this operon.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation (to MN) and by Public Health Service Grants from the National Institute for General Medicine (to MN) and the National Institute for Allergy and Infectious Diseases (to LL) and by a Research Career Development Award (to LL) from the National Institute for Allergy and Infectious Diseases. We thank Mr. W. Burke of the University of Rochester for helpful advice on DNA sequencing.

*Present address: SIBIA, P.O. Box 85200, San Diego, CA 92138

†Present address: The Upjohn Company, Kalamazoo, MI 49001

REFERENCES