

Transcription Initiation Sites of the Leucine Operons of *Salmonella typhimurium* and *Escherichia coli*

ROBERT M. GEMMILL†, JUDITH W. JONES, GEORGE W. HAUGHN
AND JOSEPH M. CALVO

*Section of Biochemistry, Molecular and Cell Biology
Cornell University, Ithaca, N. Y. 14853, U.S.A.*

(Received 13 September 1982, and in revised form 15 June 1983)

Evidence for a transcription attenuation site downstream from the *leu* promoter was obtained by transcription experiments *in vitro*. Most transcription initiated *in vitro* from *leuP* is terminated prematurely, resulting in the synthesis of a 160 nucleotide leader RNA. We define here the point at which transcription is initiated *in vitro* and *in vivo* and demonstrate that the site of premature termination is between the promoter and the first structural gene (*leuA*). Additional nucleotide sequences are presented that extend the known sequence 200 base-pairs upstream and 300 base-pairs downstream from *leuP*.

The location of the promoter-proximal end of cistron *leuA* was deduced by comparing nucleotide sequence data with the sequence of the ten amino acids at the N-terminus of α -isopropylmalate synthase. To facilitate the isolation of quantities of material for sequencing experiments, the enzyme was isolated from a plasmid-containing strain, CV605, grown under conditions of leucine limitation. Under such conditions, about 20% of the total soluble protein of strain CV605 is α -isopropylmalate synthase and another 20% is β -isopropylmalate dehydrogenase (*leuB* product).

1. Introduction

Leucine biosynthesis in enteric bacteria is catalyzed by three enzymes whose levels are co-ordinately regulated by the intracellular concentration of leucine (Calvo *et al.*, 1969a). These three enzymes are encoded by four genes which are clustered into an operon on the chromosomes of *Salmonella typhimurium* and *Escherichia coli*. *In vitro*, transcription initiated at the *leu* promoter terminates with high efficiency at a site approximately 160 nucleotides downstream (Gemmill *et al.*, 1979). The resulting transcript, termed the *leu* leader RNA, is similar in many respects to leader RNAs transcribed from other promoters (Yanofsky, 1981). It can code for a 28 amino acid peptide that includes four adjacent Leu residues and it has the potential for forming several mutually exclusive secondary structures (Keller & Calvo, 1979). On the basis of such similarities and of the transcription studies *in vitro*, we suggested that the *leu* operon is controlled by a

† Present address: Department of Zoology, Arizona State University, Tempe, Ariz. 85287, U.S.A.

transcription attenuation mechanism. Additional studies, involving analysis of regulatory mutations and of the rates of synthesis and degradation of *leu* messenger RNA, firmly establish attenuation as an important mechanism by which the *leu* operon is regulated (Searles *et al.*, 1983). Moreover, these latter studies suggest that for cells grown in a minimal medium, transcription attenuation may be the only mechanism by which leucine regulates expression of the *leu* operon.

The studies reported here establish some of the basic parameters of this system. Nucleotide sequences from *S. typhimurium* that extend the known sequences 200 base-pairs upstream and 300 base-pairs downstream from the *leu* promoter are presented. These latter sequences contain information for part of the sequence of α -isopropylmalate synthase and thereby define the promoter-proximal end of cistron *leuA*. We further define here the exact point at which transcription is initiated *in vitro* and *in vivo* and demonstrate that the site of premature termination is located between the promoter and the first structural gene.

2. Materials and Methods

(a) Bacterial strains, media and enzyme assays

The strains used in this study are shown in Table 1. L-broth or L-agar served as rich media (Miller, 1972). When required, L-agar was supplemented with ampicillin (100 μ g/ml) or tetracycline (20 μ g/ml). Minimal media included M9 for *E. coli* strains (Miller, 1972), SSA for *S. typhimurium* strains (Calvo *et al.*, 1969a) and these two solidified with 1.5% (w/v) agar. Minimal media contained 0.2% (w/v) glucose and were supplemented with Difco Casamino acids (0.2%), thiamine (1 μ g/ml) or amino acids (50 μ g/ml) when necessary.

α -Isopropylmalate α -ketoisovalerate lyase (CoA acetylating) (EC 4.1.3.12: α -IPM synthase) was assayed as described by Kohlhaw *et al.* (1969) as modified by Hertzberg *et al.* (1980). 2-Hydroxy-4-methyl-3-carboxyvalerate: NAD⁺ oxidoreductase (EC 1.1.1.85; β -IPM dehydrogenase) was assayed as described by Burns *et al.* (1963). Protein was determined by the Coomassie blue method of Bradford (1976).

(b) Purification of DNA and restriction endonuclease digestions

Cultures used for plasmid isolation were grown in M9 minimal medium supplemented with 0.5% Casamino acids and 0.2% glucose to an A_{550} of 1.0. Plasmid DNA was isolated by the CsCl density gradient method of Kupersztoch-Portnoy *et al.* (1974) as modified by Hertzberg *et al.* (1980) and chromosomal DNA by the procedure described by Smith & Calvo (1980).

Restriction endonuclease digestions were performed in the buffers recommended by the manufacturer (New England Biolabs) at DNA concentrations of 0.1 to 1 mg/ml. Preparative amounts of plasmid (0.2 to 0.5 mg) were digested with 25 units of enzyme at 37°C for 12 h. Digestion products were separated by electrophoresis on analytical (Wu *et al.*, 1976) or preparative (Smith & Calvo, 1979) agarose gels or on polyacrylamide gels (4% (w/v) acrylamide/0.2% (w/v) bisacrylamide) containing 90 mM-Tris-borate (pH 8.3), 2 mM-EDTA. DNA fragments were recovered from preparative agarose gels by the method of Zain & Roberts (1978) and from polyacrylamide gels by the method of Smith & Calvo (1980).

(c) Cloning of *leu* operon mutations

Chromosomal DNAs isolated from deletion-containing strains were digested with *EcoRI* and ligated into the *EcoRI* site of plasmid pMB9. The ligation products were used to

TABLE 1
Bacterial strains and plasmids used in this study

Strain or plasmid	Genotype	Source
A. <i>S. typhimurium</i> LT2		
CV468	<i>araB9 gal-205</i>	Margolin (1963)
<i>leuA124 araB9</i>		Margolin (1963)
<i>leuA475 araB9</i>		Margolin (1963)
CV173	<i>araB9 gal-205</i> <i>leu02007</i>	Calvo <i>et al.</i> (1969 <i>b</i>)
<i>leuA3097 araB9 gal-205</i>		Calvo & Worden (1970)
<i>leuC4138 araB9 gal-205</i>		Calvo & Worden (1970)
<i>leu-4168 araB9 gal-205</i>		Calvo & Worden (1970)
<i>leu-5058 araB9 gal-205</i>		Calvo & Worden (1970)
<i>leu-5149 araB9 gal-205</i>		Calvo & Worden (1970)
<i>leu-4017 araB9 gal-205</i>		Calvo & Worden (1970)
B. <i>E. coli</i> K12		
M94	F ⁻ <i>thr-1 leuB6</i> <i>thi-1 lacY1</i> <i>tonA21 supE44</i> <i>hsd5</i>	B. Bachmann
CSH73	HfrH Δ <i>lac</i> Δ (<i>ara-leu</i>) <i>thi</i>	Miller (1972)
CV492	M94/pCV12	Hertzberg <i>et al.</i> (1980)
CV497	CSH73/pCV14	Hertzberg <i>et al.</i> (1980)
CV591	CSH73/pMB9	This work
CV605	CSH73/pCV14	This work
CV618	M94/pCV20	This work
CV620	M94/pCV15	This work
CV659	Δ (<i>ilvHI-leu</i>) <i>ilvB thi</i> /pCV5	Wessler & Calvo (1981)
CV673	M94/pCV17	This work
CV674	M94/pCV18	This work
CV675	M94/pCV19	This work
CV676	M94/pCV21	This work
CV679	M94/pCV45	This work
C. Plasmids		
pCV12	pMB9 containing <i>leu</i> operon from CV468	Hertzberg <i>et al.</i> (1980)
pCV14	Same as pCV12 except <i>leu</i> in opposite orientation	Hertzberg <i>et al.</i> (1980)
pCV15	pMB9 containing <i>leu</i> operon from <i>leu-4168</i>	Gemmill <i>et al.</i> (1979)
pCV17	pMB9 containing <i>leu</i> operon from <i>leuC4138</i>	This work
pCV18	pMB9 containing <i>leu</i> operon from <i>leuA475</i>	This work
pCV19	pMB9 containing <i>leu</i> operon from <i>leuA124</i>	This work
pCV20	pMB9 containing <i>leu</i> operon from <i>leu-500</i>	Gemmill <i>et al.</i> (1979)
pCV21	pBR322 containing <i>leuP</i> and part of <i>leuA</i> on an <i>EcoRI/BglII</i> fragment	This work
pCV45	pMB9 containing <i>leu</i> operon from <i>leuA3097</i>	This work
pCV5	pBR322 containing <i>leu</i> operon from <i>E. coli</i>	Wessler & Calvo (1981)

transform *E. coli* strain M94 (*leuB6*) to leucine prototrophy and transformants were screened for those that were resistant to tetracycline. The plasmids generated during the course of this work are listed in Table 1.

(d) *Blot transfer and hybridization analysis*

Blot transfers were prepared as described by Southern (1975). Nick-translation was performed following the procedures of Rigby *et al.* (1977) as modified by Smith & Calvo (1980).

(e) *DNA sequence analysis*

DNA sequencing reactions were carried out as described by Maxam & Gilbert (1980). Portions of the cleavage reactions were resolved on 0.4 mm thick, 5% or 15% polyacrylamide gels containing 50% urea (Sanger & Coulson, 1978). 3' End-labeled DNA fragments were prepared by treating restriction fragments with AMV reverse transcriptase (gift from J. Beard) and the appropriate [α -³²P]dNTP (Smith & Calvo, 1980). 5' End-labeled DNA fragments were prepared by treatment with bacterial alkaline phosphatase followed by polynucleotide kinase and [α -³²P]ATP (2000 to 3000 Ci/mmol; Amersham: Maxam & Gilbert, 1980). Labeled ends were separated by cleavage with a second restriction endonuclease followed by gel electrophoresis to purify the resulting fragments.

(f) *Transcription in vitro*

Transcription of DNA templates *in vitro* with purified RNA polymerase was performed essentially as described by Lee & Yanofsky (1977). DNA (1 μ g) was added to a reaction mixture containing in a total volume of 100 μ l: 20 mM-Tris-acetate (pH 7.9), 0.1 mM-EDTA, 0.1 mM-dithiothreitol, 100 mM-KCl, 100 μ M-GTP, 100 μ M-ATP, 75 μ M-CTP, 75 μ M-UTP and 10 μ Ci [α -³²]UTP (400 Ci/mmol; Amersham). One unit of purified *E. coli* RNA polymerase (New England Biolabs) was added followed by the addition of magnesium acetate to 4 mM to initiate transcription. Reactions were incubated at 37°C for 30 min and terminated by the addition of 300 μ l of 0.3 M-sodium acetate, 400 μ l of buffer-equilibrated phenol and 20 μ g of carrier transfer RNA. The aqueous phase from the phenol extraction was precipitated twice with ethanol, dried *in vacuo*, dissolved in 6 μ l of 50% urea/0.04% bromophenol blue/0.04% xylene cyanol, and loaded on a 6% polyacrylamide/50% urea gel. Electrophoresis was performed at 1100 V for 4 to 5 h.

For preparing transcripts labeled at the 5' end with ³²PppA, reactions were performed in the same way except the UTP and CTP concentrations were 100 μ M, [γ -³²P]ATP (200 to 300 Ci/mmol) was 50 μ M, the total volume was 25 μ l and the reaction was incubated for 2 h at 37°C.

Transcripts labeled at the 5' end by treatment with polynucleotide kinase were prepared as follows. Plasmid pCV21 (4 pmol) was transcribed with 0.5 pmol of RNA polymerase (a gift from L. Lau) in a total volume of 0.5 ml and unlabeled *leu* leader RNA was isolated by electrophoresis through 6% acrylamide/50% urea gels. Size standards were provided by samples of transcription reactions containing [α -³²P]UTP. After elution from the gel, the sample was treated sequentially with bacterial alkaline phosphatase (Bethesda Research Labs; 650 units, 37°C, 1 h) and T4 polynucleotide kinase (Boehringer-Mannheim; 5 units, 37°C, 20 min, 120 μ Ci [γ -³²P]ATP, 3000 Ci/mmol) following procedures outlined by Maxam & Gilbert (1980). Labeled *leu* leader was isolated following electrophoresis through 5% acrylamide/50% urea gels.

(g) *RNA sequence analysis*

Portions of a gel were excised and RNA extracted from them as follows. A gel slice was ground to a fine paste with a Teflon-coated rod in a siliconized 10 mm \times 75 mm tube and

incubated for 12 h at 37°C with 0.4 ml of elution medium (0.3 M-sodium acetate, 0.1% (w/v) sodium dodecyl sulfate, and 0.5 mM-EDTA) containing 20 µg of carrier tRNA. After removal of the liquid phase the gel fragments were re-extracted for 1 h with fresh elution medium. The combined extracts were filtered through siliconized glass wool and precipitated with ethanol.

RNA sequence analyses were performed by the chemical method of Peattie (1979) and by enzymatic means. For the latter case, portions of 5' end-labeled RNA were dissolved in 25 µl of TE buffer (10 mM-Tris·HCl, pH 8, 1 mM-EDTA) and incubated 1.5 h at 37°C with either ribonuclease (RNase) T₁ (12 units; Bethesda Research Labs), RNase A (0.04 µg; Mann Research Labs), RNase T₁ plus pancreatic RNase, or RNase U₂ (14 units; Bethesda Research Labs, plus 40 µg of *E. coli* tRNA). Products following chemical or enzymatic cleavage were resolved on 12.5% polyacrylamide/50% urea gels (3400 V, 4 h) or 20% polyacrylamide/50% urea gels (2000 V, 1 h), respectively.

(h) *S*₁ mapping experiments

RNA was isolated from *Salmonella* strain CV173 grown to an *A*₅₅₀ of 0.8 in SSA minimal medium by methods described by Smith & Calvo (1980). *leu* leader RNA synthesized *in vitro* from 0.5 pmol of plasmid pCV12 was prepared and isolated as described in the last paragraph of the previous section. A 126 base-pair *Hinc*II/*Hpa*II DNA fragment, isolated as described by Smith & Calvo (1980), was labeled by sequential treatment with bacterial alkaline phosphatase and T4 polynucleotide kinase (Maxam & Gilbert, 1980) in the presence of [γ -³²P]ATP. Strands were separated and isolated by acrylamide gel electrophoresis as described by Smith & Calvo (1980). Conditions for hybridization and nuclease *S*₁ (Sigma Biochemicals) digestion are described by Favaloro *et al.* (1980).

(i) Purification of α -IPM synthase

Strain CV605 (CSH73/pCV14) was grown under conditions of partial leucine limitation as follows. A culture was grown to mid-log phase in minimal medium (SSA salts) containing 10 µg leucine/ml, an amount supporting a final yield of 5×10^8 cells/ml. The absorbance at 450 nm was followed to determine the time at which growth began to be limited by the availability of leucine. Within 60 min of that time (longer delays resulted in death of cells), fresh medium containing 10 µg leucine/ml was introduced with a peristaltic pump such that the volume of the culture doubled at the end of 4 h. At that time, cells were harvested by centrifugation and cell pellets frozen at -20°C. One to 3 g of cells were disrupted by sonication for 60 s in 0.1 M-K₂HPO₄ (pH 6.8), containing 5% glycerol, 1 mM-dithiothreitol, and 0.5 mM-phenylmethanesulfonyl fluoride (2 ml buffer/g cells). After removing debris by centrifugation for 15 min at 19,000 g, the extract was treated with streptomycin sulfate (Bartholomew & Calvo, 1971) and the precipitate discarded. The supernatant from a 40% ammonium sulfate fractionation (Bartholomew & Calvo, 1971) was diluted to 20% ammonium sulfate with sonication buffer and transferred to a 2.5 cm × 40 cm Biogel 0.5 m column equilibrated with 20% ammonium sulfate in sonication buffer. The column was eluted with sonication buffer containing 20% ammonium sulfate. Fractions containing high activity were pooled and brought to 65% ammonium sulfate. After centrifugation the pellet was redissolved in 0.01 M-Tris·HCl (pH 7.5), dialyzed for 6 h against 2 changes of distilled water, lyophilized, and stored at -20°C. Just prior to analysis it was dissolved in 88% (v/v) formic acid and dialyzed overnight at 4°C against several changes of 22% formic acid.

(j) Amino acid sequence analysis

Amino acid sequence analysis of 0.85 mg of purified α -IPM synthase was performed with a Beckman model 890C sequencer with the volatile buffer (*N,N*-dimethyl allylamine)

program (Beckman λ 122974). Horse heart cytochrome *c* (type VI; Sigma Biochemicals), which has a blocked N-terminal group, was used as a carrier. Fractions were dried under nitrogen, converted to phenylthiohydantoin derivatives in 1 M-HCl containing 1 μ l ethanethiol/ml, extracted twice with ethyl acetate and analyzed by high pressure liquid chromatography. Residues were also determined by amino acid analysis following hydrolysis in constant-boiling HCl containing 0.1% SnCl₂, as described by Mendez & Lai (1975).

High pressure liquid chromatography separations of derivatives were performed on a C₁₈ μ Bondapak reverse-phase column (Water Associates, Milford, Mass.) eluted over 20 min at room temperature with a linear gradient prepared from 5% solvent A (1:9, CH₃OH : H₂O, 0.25% (v/v) acetic acid, pH adjusted to 4.1 with NaOH) to 45% solvent B (9:1, CH₃OH : H₂O, 0.025% (v/v) acetic acid).

3. Results

(a) Correlation of genetic and restriction endonuclease maps of the *leu* operon of *S. typhimurium*

A 5.1 kb \dagger *EcoRI* fragment of plasmid pCV12 carries the majority of the *S. typhimurium* LT2 *leu* operon including intact *leuA*, *B* and *C* genes (Herzberg *et al.*, 1980). A restriction endonuclease map of this DNA is shown in Figure 1(a) and (b) together with a summary of the genetic map (Fig. 1(c) and (d)). Alignment of the two maps was achieved by analyzing the results of two kinds of experiments. In one approach, chromosomal DNA from deletion-containing strains was analyzed by hybridization. DNAs from strains *leu-4168*, *leu-5058*, *leu-5149* and *leu-4017* were separately cleaved with several different restriction endonucleases and the products were fractionated by agarose gel electrophoresis, transferred to nitrocellulose, and hybridized with plasmid pCV12 DNA labeled to high specific activity by nick-translation. Each of these deletions has one endpoint outside and the other within the *leu* operon. An example of this type of analysis using the endonuclease *PstI* is shown in Figure 2. This enzyme cleaves the wild-type *leu* operon (CV468) at a single site generating two fragments that hybridize with the pCV12 probe (4.2 kb and 3.8 kb : C. Squires, unpublished data). The fragment pattern obtained from the four deletion DNAs is altered in a manner dependent on the position of the endpoint within the *leu* operon. One of the two fragments present in wild-type DNA (3.8 kb fragment) is missing from all four deletion strains and is replaced by a fusion fragment of variable size. DNA from strain *leu-4017* is missing both the 3.8 and the 4.2 kb fragments. This deletion is therefore the only one that crosses the *PstI* site. Analysis of these DNAs with a number of other restriction endonucleases positioned the deletion endpoints as shown in Figure 1(c) and (d).

Additional information for aligning genetic and restriction maps was obtained by cloning DNA fragments generated by *EcoRI* cleavage of DNAs from strains *leu-4168*, *leuA124*, *leuA475* and *leuC4138* into plasmid pMB9. Plasmids containing the *leu* operon were selected by transformation of strain M94 (*leuB6*) to leucine prototrophy. In the case of the plasmid containing *leu* DNA from strain *leu-4168* (deletion covering part of the *leu* promoter), transcription of *leu* genes

\dagger Abbreviations used: kb, 10³ bases or base-pairs; PTH, phenylthiohydantoin.

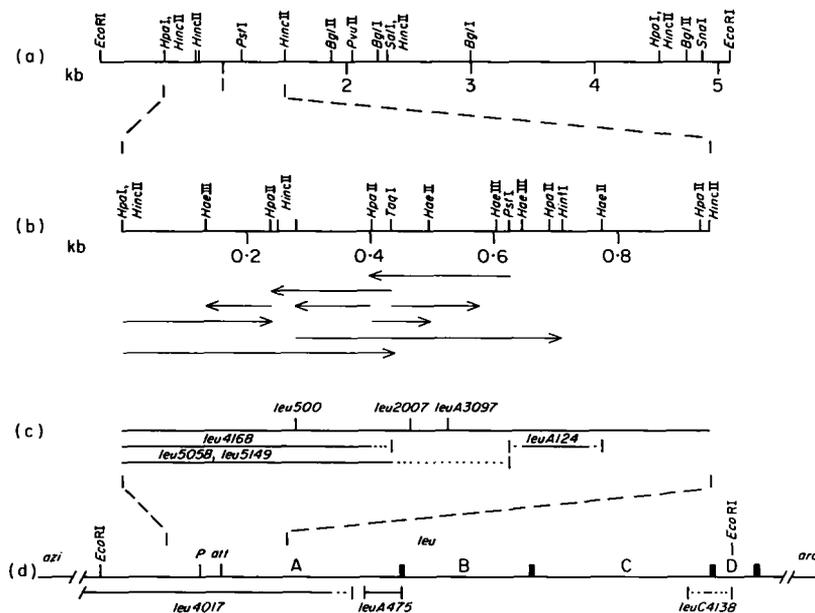


FIG. 1. Correlation between restriction and genetic maps of the *leu* operon.

(a) Restriction endonuclease cleavage map of the *Salmonella* leucine operon contained on the 5.1 kb *EcoRI* fragment from plasmid pCV12.

(b) An expanded 1 kb section of the restriction map with arrows denoting fragments that were used for sequence analysis. Sequence data were obtained for both the coding strand (arrow points leftward) and the non-coding strand (arrow points rightward) over most of this region. The sequence shown in Fig. 3 extends from the leftmost *HaeIII* site to the *PstI* site.

(c) Detailed genetic map of the *Salmonella leu* operon control region, aligned with and drawn to the same scale as the restriction map in (b). The locations of point mutations *leu-500*, *leu-2007* and *leuA3097* are known exactly from sequencing studies (this work and R. Gemmill & L. Searles, unpublished data). The extents of 4 deletions are indicated (uncertainty in endpoint indicated by dots).

(d) Genetic map of the *Salmonella leu* operon aligned with and drawn to the same scale as the restriction map in (a). The sites marked *P* (*leu* promoter) and *att* (*leu* attenuator) are defined by mutations *leu-500* and *leu-2007*, respectively. The position of boundaries between *leuA*, *B*, *C* and *D* are approximate.

was presumably *via* readthrough from a plasmid promoter. The position of each deletion, determined by comparing restriction endonuclease maps of deletion-containing DNAs with wild-type DNA, is shown in Figure 1(c) and (d). Deletion *leu-4168* is particularly important to this analysis because it covers promoter mutation *leu-500* but does not cover nearby mutations that cause constitutive expression of leucine-forming enzymes (Calvo & Worden, 1970; and our unpublished data). One endpoint of deletion *leu-4168* was localized by these experiments to a 30 base-pair region between an *HpaII* site and a *TaqI* site.

Relative to the restriction maps shown in Figure 1, these results establish the position of the *leu* regulatory region, the direction of transcription of the *leu* operon and the approximate positions of the *leuA*, *B* and *C* genes.

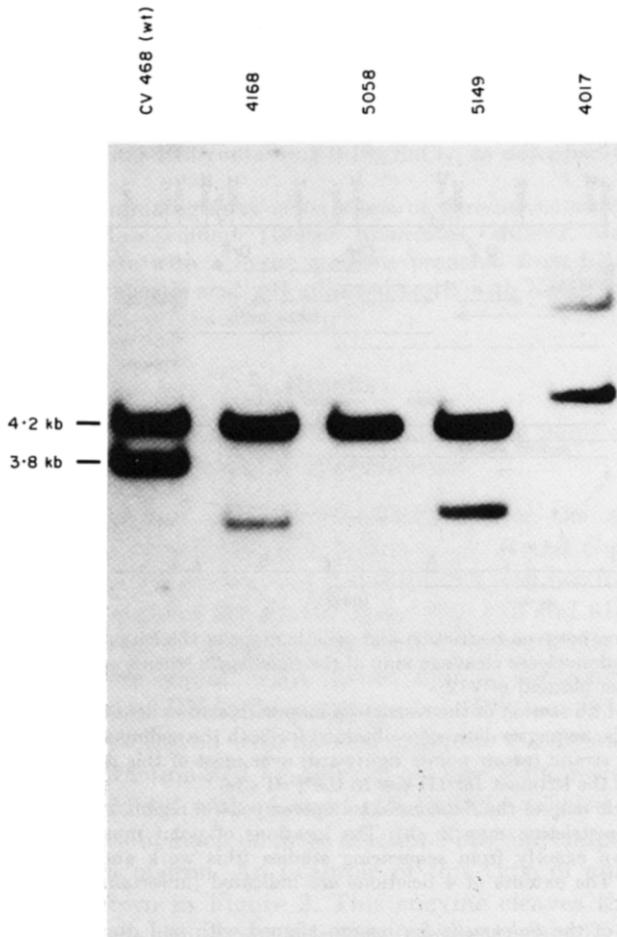


FIG. 2. Southern blot analysis defining the ends of several *leu* deletions.

DNA from each of the indicated strains was digested with *Pst*I, and 2 μ g of each were resolved on a 1% agarose gel. After transfer to nitrocellulose, samples were hybridized to nick-translated pCV12 DNA.

(b) *The nucleotide sequence of the promoter end of the leu operon*

The chemical DNA sequencing method of Maxam & Gilbert (1980) was used to sequence a 500 base-pair region surrounding and including the *leu* regulatory region. The sequencing strategy employed is shown in Figure 1(b) and the sequence in Figure 3. From results published previously (Gemmill *et al.*, 1979) we tentatively assigned the *leu* promoter to positions -36 to -7 and the *leu* attenuator to positions 132 to 157. The results presented in the next two sections confirm these assignments.



FIG. 3. Nucleotide sequence (non-coding strand) of the region including and adjacent to the *leu* promoter of *S. typhimurium*. Hyphens have been omitted for clarity.

The region from which the *leu* leader is transcribed is underlined with an arrow. Position 1 corresponds to the site at which transcription is initiated. The amino acid sequence of α -IPM synthase predicted from the DNA sequence is shown in italics and that determined by Edman degradation is shown below. The asterisk signifies that the identity of the N terminal amino acid is unknown.

(c) *Sequence analysis of leu leader RNAs synthesized in vitro from S. typhimurium and E. coli templates*

Plasmids pCV12 (*S. typhimurium*) and pCV5 (*E. coli*) give rise to *leu* leader transcripts of about 161 and 164 nucleotides, respectively, when transcribed *in vitro* with purified RNA polymerase (Fig. 4, lanes 1 and 3) (Gemmill *et al.*, 1979; Wessler & Calvo, 1981). These same transcripts are labeled when $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ was included in reaction mixtures, indicating that at least some of the transcripts begin with an A residue (Fig. 4, lanes 4 and 5). Although not readily apparent from this gel, $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ labels three similarly sized species. The same is true when labeling was carried out more generally with $[\alpha\text{-}^{32}\text{P}]\text{UTP}$.

Nucleotide sequence analyses were performed on samples of 5' end-labeled transcripts by the chemical method of Peattie (1979) using the G specific and A + G reactions (Fig. 5). As noted by Peattie, hydrazinolysis did not give specific cleavage at pyrimidine residues of 5' end-labeled samples. However, a hydrazinolysis reaction was included in each experiment because it aided in the analysis by producing a pattern of all fragment lengths up to the intact molecule. At the bottom of Figure 5, the nucleotide sequence of a portion of both the *S. typhimurium* and *E. coli leu* regulatory regions (sequences from position -37 to +71) is shown together with the pattern of A and G residues in leader transcripts deduced from RNA sequencing experiments. Because of extensive homology between the DNAs of these two organisms, the A and G patterns of the respective leader transcripts are very similar. However, the patterns are not identical and

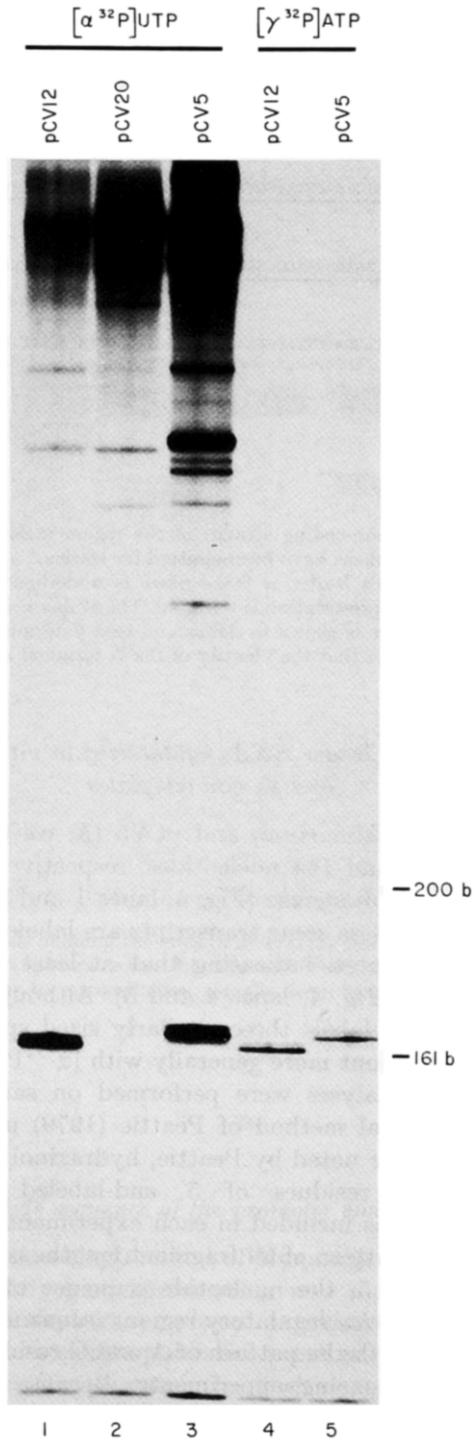


FIG. 4.

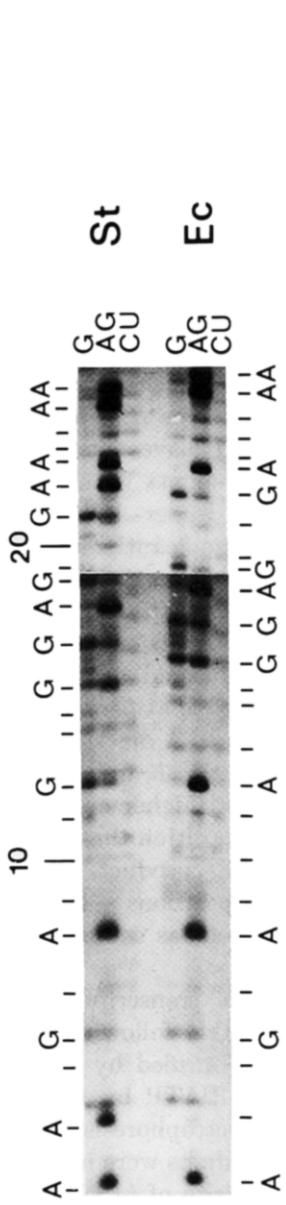
the differences serve to confirm that the patterns have been positioned correctly in Figure 5. If the leftmost distinct band in Figure 5 corresponds to cleavage at position 3, then the observed pattern of bands matches the known sequence. For example, *S. typhimurium* has an A at position 4 and a G at position 12, whereas *E. coli* has a pyrimidine at position 4 (and therefore no band in the A or A/G lanes) and an A at position 12. Similarly, both leaders have a distinctive pattern of bands corresponding to G-G-A-G (15 to 18 in *Salmonella*), but this pattern is shifted one space to the right in the case of *E. coli* because of an extra nucleotide between positions 12 and 13. Analysis of other gels from which the sequence could be read in the region +25 to +75 confirmed the alignment shown in Figure 5 (data not shown).

The sequence analysis described above identified nucleotides to within a few bases of the 5' end of the leader RNAs. The 5' terminal nucleotide was identified by an enzymatic approach. 5' End-labeled samples of the *S. typhimurium* and *E. coli* leader RNAs were digested with either RNase T₁ (cleaves after G residues) or RNase A (cleaves after pyrimidine residues). The products were resolved on a 20% polyacrylamide/50% urea gel (Fig. 6). Both leader RNAs yielded T₁ digestion products of the same size but of a size significantly larger than the RNase A digestion products. Since the RNA transcripts were labeled with [γ -³²P]ATP, they must be initiated with an A residue. Initiation at one of the four consecutive A residues at positions -3 to -6 is eliminated from consideration because the expected products of RNase A digestion would be longer than the product from RNase T₁ digestion. Similarly, the A residue at position 3 could not have been a major start site because in that case RNase A digestion would have given different products from RNAs transcribed from *E. coli* and *S. typhimurium* templates. Furthermore, A residues at positions 4 and higher are eliminated as starting sites because the sequence could be read to a position that included these residues (Fig. 5). Finally, the size of the RNase T₁ product, a hexamer, is consistent with the proposed initiation site. Between positions -10 and +35 for the *S. typhimurium* sequence, only the A at position 1 has a G five nucleotides downstream (at position 6).

To confirm these results, and to determine whether transcription is initiated with nucleotides other than A, we carried out the following experiment. *Salmonella leu* leader RNA was synthesized *in vitro*, purified by acrylamide gel electrophoresis, and labeled at the 5' end with [γ -³²P]ATP by treatment with polynucleotide kinase. After repurification by gel electrophoresis, samples were digested with ribonucleases T₁, A and U₂, and the products were identified by gel electrophoresis. The results confirmed that the major site of *in vitro* initiation is

Fig. 4. *In vitro* transcripts synthesized from *leu* templates.

Samples of the indicated plasmid (1 μ g) were incubated with purified RNA polymerase and nucleoside triphosphates, including [α -³²P]UTP or [γ -³²P]ATP as indicated. Electrophoresis was performed through 6% acrylamide gels containing 50% urea. Plasmids pCV12 (wild-type *leu* operon) and pCV20 (*leu-500* promoter mutation) contain *S. typhimurium* DNA, whereas CV5 contains *E. coli leu* DNA. Molecular weight standards (not shown) included the 6 S transcript of λ phage and end-labeled *Hinf*I fragments of pBR322.



-10 | 1 | 10 | 20 | 30 | 40 | 50 | 60 | 70 |

TAcCaCTAAaAGcATaAcGcATTcG■CTGGAGcTGaATTAATGtCaCATATcGTTcGtTTcActGGcTAcTAcTAcTAcTcAaCGcCA

St

AA-G-A---G■---GGAG--GAA--AA-G--A-A-A--G---G---A---GGG---A---A---AA-G-A
 A--G-A---A---GGAG--GA---AA-GA---A-A--G---G---A---GG---A---A---AAA-G-A

Ec

TAAcTcTAAaAGcATaTAcGcATTcATcTGGAGcTgATTTAAATGAcTcAcATcGTTcGcTTTATcGGtCTAcTAcTAcTAcTAAcCGcCA

FIG. 5.

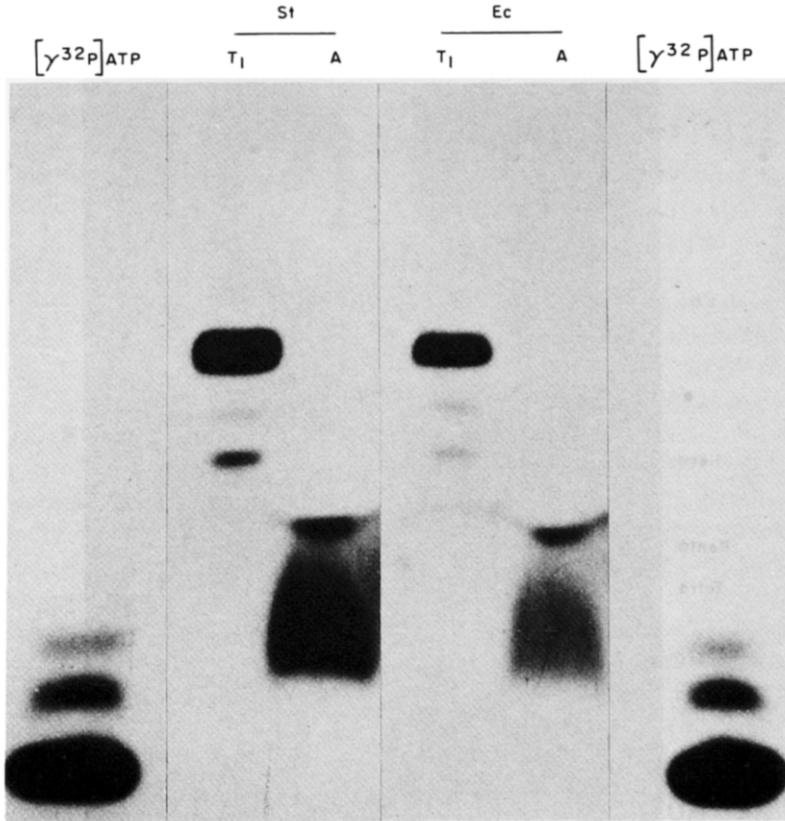


FIG. 6. Enzymatic digestion products of leader RNA labeled during transcription with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$. End-labeled leader RNAs, synthesized as described in the legend to Fig. 5, were digested to completion with either T_1 ribonuclease or ribonuclease A. The products were resolved on a 20% polyacrylamide/50% urea sequencing gel. The outer 2 lanes contain $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ as a marker.

the A at position 1: digestion with RNase T_1 gave a hexamer, RNase A a dimer, and RNase U_2 (cleaves after G and A) a mononucleotide (Fig. 7).

The *in vitro* synthesized *leu* leader RNAs from both *S. typhimurium* and *E. coli* show a triplet band pattern when resolved on 6% polyacrylamide/50% urea gels (visible in a shorter exposure of the autoradiogram shown in Fig. 4). This triplet

FIG. 5. Leader RNA sequence analysis.

Leader RNAs synthesized from *S. typhimurium* (St, pCV20) and *E. coli* (Ec, pCV5) templates were labeled at the 5' ends by incorporation of $[\gamma\text{-}^{32}\text{P}]\text{ATP}$. The sequencing reactions of Peattie (1979) were carried out on gel-purified RNAs (Fig. 4) and the products were resolved on 12.5% sequencing gels. The direction of migration is from right to left. Below is shown a portion of the DNA sequence (non-coding strand) of both the *S. typhimurium* and *E. coli* *leu* regulatory regions including the first 70 nucleotides of the leader region. The numbering corresponds to that in Fig. 3. Small capital letters denote positions at which the 2 sequences differ. In between the 2 DNA strands are shown the positions for A and G residues as deduced from RNA sequence gels. A space was inserted into the *S. typhimurium* sequence between positions 12 and 13 so that the sequences from this species and from *E. coli* remain in register. Sequence hyphens have been omitted for clarity.

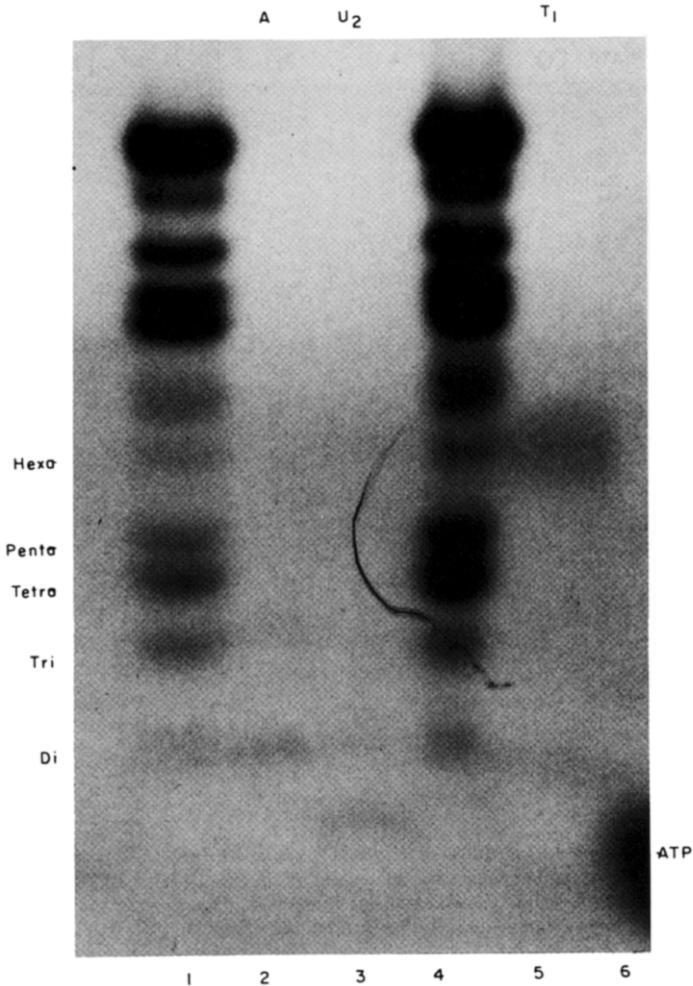


FIG. 7. Enzymatic digestion products of leader RNA labeled after transcription with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and polynucleotide kinase.

leu leader RNA was synthesized *in vitro*, purified by acrylamide gel electrophoresis, and labeled with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and polynucleotide kinase as described in Materials and Methods. Samples were digested to completion with RNase A (lane 2), RNase U_2 (lane 3), or RNase T_1 (lane 5) and resolved on a 20% acrylamide/50% urea gel. Lanes 1 and 4 contain as size standards samples of a chemically synthesized oligodeoxyribonucleotide (dodecamer) labeled at the 5' end with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ and polynucleotide kinase. The latter was treated under conditions of partial hydrolysis with DNase I (yields dinucleotides as the smallest product). Lane 6 contains $[\gamma\text{-}^{32}\text{P}]\text{ATP}$.

pattern must be due to heterogeneity at the 3' end since heterogeneity at the 5' end would have resulted in uninterpretable band patterns on the sequence gels. Figure 8 shows a portion of a sequencing gel including the largest fragments at the 3' end of both the *S. typhimurium* and *E. coli leu* leaders. The three G residues from positions 151 to 153 are seen as an unresolved region in lane G. Six bands can be seen in lane U between the G residues and the intense bands consisting of the three full-length leader transcripts. The simplest interpretation of this pattern

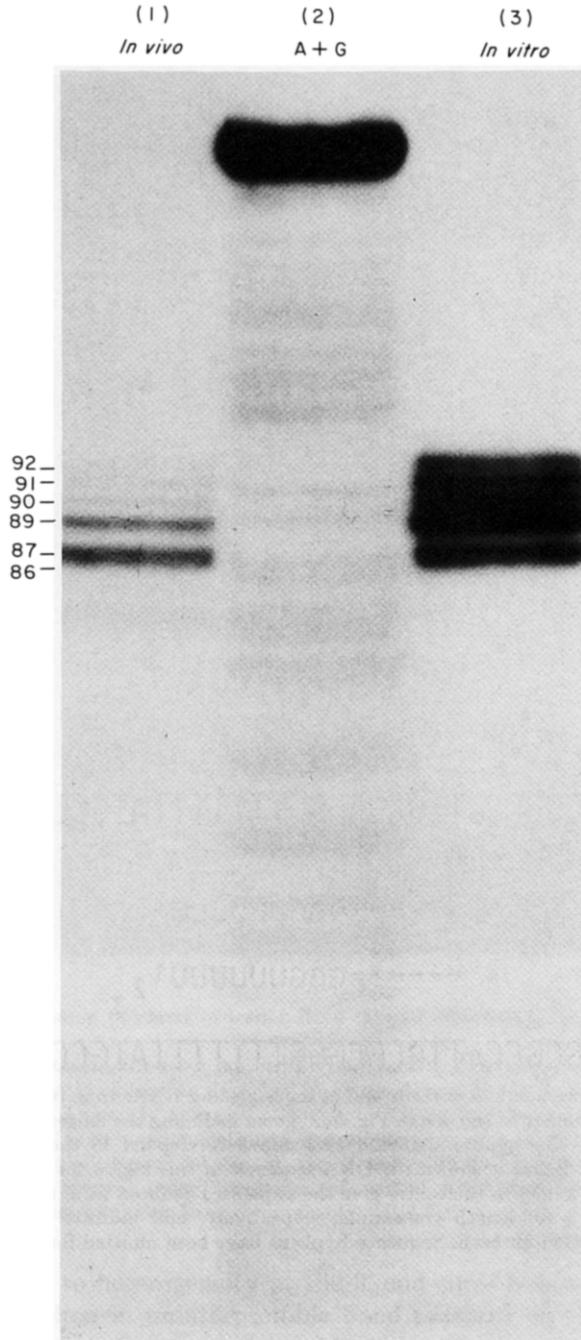


FIG. 9. S_1 mapping of 5' ends of *leu* leader transcripts.

Samples of the coding strands of 5' end-labeled *HincII/HpaII*₁₂₆ (Fig. 1(b), kb scale 0.28 to 0.4) containing 20,000 cts/min were hybridized for 18 h at 45°C with 25 µg of *E. coli* tRNA (no bands visible; data not shown), 25 µg of RNA from strain CV173 (lane (1)), or with the RNA synthesized *in vitro* from 0.5 pmol of template pCV12 (lane (3)). Samples were digested with 200 units nuclease S_1 /ml

pattern is shifted in position relative to *S. typhimurium* due to the extra nucleotides in the *E. coli* leader sequence.

(d) *S*₁ nuclease analysis of *leu* transcripts synthesized *in vivo*

*S*₁ mapping following the procedure of Favalaro *et al.* (1980) was used to determine the 5' end of *leu* leader RNA molecules synthesized *in vivo*. The coding strand of fragment *HincII/HpaII*₁₂₆ (Fig. 3, -34 to +91), labeled at the 5' end with ³²P, was hybridized with RNA isolated from *Salmonella* strain CV173. After digestion with endonuclease *S*₁, samples were denatured and fractionated on acrylamide/urea gels. To estimate the lengths of DNA molecules protected from *S*₁ nuclease digestion, samples of the end-labeled DNA fragment were subjected to chemical cleavage at A plus G residues and fractionated alongside the *S*₁-treated samples (Fig. 9). Three intense bands (86, 87 and 89 nucleotides long) and three relatively weak bands (90, 91, 92) correspond to 5' endpoints of +7, +6, and +4 (strong bands) and +3, +2, and +1 (weak bands), respectively. From this result one might imagine that the synthesis of *leu* leader RNA is initiated at multiple points *in vivo*, or that *leu* leader RNA is either normally processed at the 5' end or is partially degraded by nucleases during isolation. That none of these explanations is likely is seen from the following experiment. *leu* leader RNA synthesized *in vitro* was subjected to the identical *S*₁ nuclease analysis (Fig. 9). The pattern of bands is very similar for RNAs synthesized *in vivo* and *in vitro* except that for the latter, there is a higher proportion of the 89 nucleotide band. The simplest interpretation of these results is that the synthesis of *leu* leader is initiated at the same point both *in vitro* and *in vivo*, and that there is no significant processing at the 5' end *in vivo*. The heterogeneity of endpoints observed in these experiments are most likely due to nibbling at double-stranded ends by *S*₁ or contaminating nucleases.

The results in preceding sections establish the sites at which transcription of the *leu* leader is initiated and terminated. The position of the promoter-proximal end of the *leuA* gene relative to those sites was determined as described in the next two sections.

(e) *Strains producing large quantities of leu enzymes*

To simplify purification of α -IPM synthase, isolation was carried out from cells containing exceptionally large quantities of the enzyme. The fragment of *Salmonella* DNA cloned on plasmid pCV12 carries a functional *leu* control region and functional *leuA*, *leuB* and *leuC* genes (Fig. 1). It lacks, however, a small part of *leuD* and, as a result, does not confer prototrophy upon a strain deleted for the entire leucine operon. Plasmid pCV14 is identical to plasmid pCV12 except that

at 14°C for 2 h, denatured, and fractionated on an 8% acrylamide/50% urea gel. DNA fragment lengths were estimated by comparison with fragments of *HincII/HpaII*₁₂₆ generated by cleavage at A+G residues with pyridinium formate (lane (2)). Note that the band pattern from the A+G sequencing from this and other experiments shows an abnormally small spacing between bands 86 and 87. We assumed that the product of *S*₁ cleavage migrates as though it were 1.5 nucleotides longer than an equivalent product of the A+G cleavage reaction.

the *leu* fragment is in the opposite orientation relative to pMB9 genes (Hertzberg *et al.*, 1980). Since the parent pMB9 plasmid is under relaxed replication control, these two hybrid plasmids were expected to exist in multiple copies per cell. That this is indeed the case is suggested by the fact that plasmid-carrying strains grown under conditions of excess leucine had levels of *leu* enzymes more than 40-fold higher than in the parent strains (data not shown). Further elevation of enzyme levels was achieved by growing cells under conditions of leucine limitation in a modified chemostat. The results of one such experiment involving strain CV605($\Delta leu/pCV14$) are shown in Figure 10. Under conditions of derepression, massive synthesis of *leu* enzymes occurred. From a number of such experiments, the average specific activity for α -IPM synthase and β -IPM dehydrogenase after four hours of limitation was 237 ± 38 and 700 ± 85 mmol/hour per mg protein, respectively. Assuming specific activities of 990 and 3240 for the pure proteins (Kohlhaw *et al.*, 1969; and our unpublished data), each enzyme amounted to about 23% of the total soluble protein of derepressed cells. Similar results were obtained for strain CV497 ($\Delta leu/pCV12$) grown under the same conditions except that only about half as much of each enzyme was produced (Fig. 10).

(f) *N-Terminal amino acid sequence of α -IPM synthase*

α -IPM synthase was purified from a plasmid-containing strain grown as described above and the sequence of 11 amino acids at the N terminus was determined by sequential Edman degradation (Edman & Begg, 1967) (Fig. 3). α -IPM synthase does not have a methionine at the N terminus; presumably it is incorporated into a nascent polypeptide chain and subsequently removed. The identity of the amino-terminal amino acid, indicated in Figure 3 by an asterisk, is

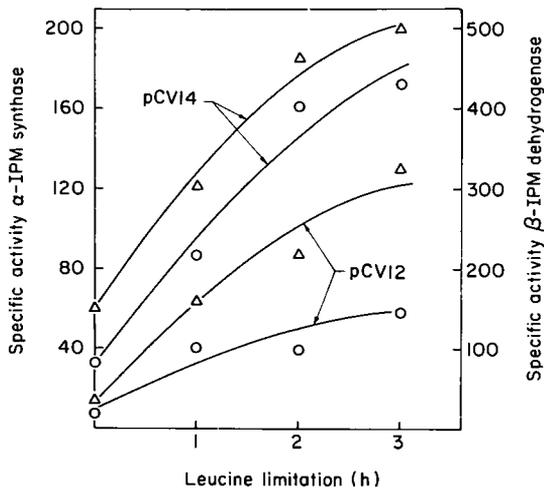


FIG. 10. Synthesis of α -IPM synthase (○) and β -IPM dehydrogenase (△) in strains CV497 ($\Delta leu/pCV12$) and CV605 ($\Delta leu/pCV14$) in response to leucine limitation.

The limitation was initiated at time zero.

not defined by this work. The nucleotide sequence predicts a serine at this position. The first cycle of Edman degradation yielded two peaks in about equal amounts, one eluting from high pressure liquid chromatography near PTH-tryptophan (but not coinciding with known standards) and the other at a position corresponding to PTH-alanine. Upon hydrolysis of the PTH derivatives, only alanine was observed, in 50% the expected yield. A cycle of Edman degradation of 16 nmol of other proteins containing N-terminal serine normally yields no PTH derivative, while back hydrolysis produces alanine. These results indicate that the N terminus of α -IPM synthase is not serine. Lawther *et al.* (1979) described a case in which the N-terminal amino acid of transaminase B is not threonine as predicted from the nucleotide sequence but rather glycine. They ascribed the difference to a post-translational modification reaction.

We considered the possibility that the N terminus of α -IPM synthase is a modified serine residue. Attempts were made to identify the modification by mass spectrometry. PTH samples isolated by high pressure liquid chromatography and subjected to mass spectrometry apparently did not volatilize under conditions that gave good signals from pure samples of PTH-alanine and PTH-tryptophan. From the ultraviolet light absorption spectrum of a sample of pure enzyme, there was no evidence for a bound nucleotide derivative. To determine whether the derivative contained a phosphate group, strain CV605 was grown in a medium containing ^{32}P -labeled phosphate and proteins were fractionated by two-dimensional procedures. No label was observed in the region corresponding to α -IPM synthase (S. Dunn & J. Jones, unpublished results).

Strains carrying the ochre mutation *leuA3097* (which maps very close to the control region) produce little if any of the *leu* enzymes (G. Schatz & J. Calvo, unpublished data). *leu* operon DNA was cloned from strain *leuA3097* and the nucleotide sequence of the region between positions 90 and 290 was determined. The mutation converts the fourth amino acid codon, CAA (Gln in wild-type), to nonsense codon UAA (Fig. 3). Presumably, the strong pleiotropic effect of *leuA3097* is the result of translation termination at the UAA codon which subsequently results in transcription termination.

4. Discussion

The DNA sequence data reported here include the entire *S. typhimurium leu* operon control region as well as the first nucleotides of the *leuA* cistron. More limited nucleotide sequence data were reported together with a transcription attenuation model that explains regulatory properties of the *leu* operon (Gemmill *et al.*, 1979). The work reported here establishes three important points that are relevant to this model: the site at which transcription begins *in vitro* (which in turn defines the location of *leuP*), the sites at which most transcription terminates *in vitro*, and the proximal boundary of the *leuA* cistron. The ATG translational start site is preceded by a potential ribosome-binding site (A-G-G-A) centered at position 191 (Steitz, 1979). The attenuator site (position 160 to 162) is 160 nucleotides from the point where transcription is initiated and 40 nucleotides from the point at which translation of *leuA* is initiated (Fig. 3).

The results of S_1 mapping experiments indicate that *in vivo*, transcription of the *leu* operon begins at or near the position defined in Figure 3. Does such transcription terminate *in vivo* at position 160, as it does *in vitro*? The following evidence suggests that it does. Hybridization experiments employing pulse-labeled RNA indicate that the rate of synthesis of *leu* leader RNA is constitutive and high, whereas the rate of synthesis of *leu* mRNA is regulated by leucine and is high only when cells are starved for leucine (Searles *et al.*, 1983). Since transcription of the *leu* operon occurs only from a single promoter (Gemmill *et al.*, 1979; R. Gemmill, unpublished data), these data indicate that most transcription initiated at *leuP* terminates prior to the *leu* structural genes. The position of the termination site can be deduced from an analysis of single base substitution mutations, known from nucleotide sequence studies to be in the *leu* attenuator (Searles *et al.*, 1983). Such mutations cause transcription to continue past the attenuator site *in vitro* and result in constitutive, high levels of *leu* enzymes and mRNA *in vivo* (Searles *et al.*, 1983; Hertzberg *et al.*, 1980).

By growing plasmid-containing strains under conditions of leucine limitation, we found that two *leu* enzymes amount to more than 40% of the total soluble protein of cells. This has facilitated the isolation of quantities of these two proteins, which can now be obtained pure after a fivefold purification. The consequences to cells of directing a large proportion of their carbon and energy towards the production of just two enzymes is difficult to assess because at present the only good method of achieving derepression is by limiting the growth rate of a leucine auxotroph. We note, however, that complete starvation for leucine for just a few hours resulted in the death of strain CV605 ($\Delta leu/pCV14$).

We thank Dr C. Fullmer (Cornell Amino Acid Analysis and Protein Sequencing Facility) and Dr T. Wachs (Cornell Mass Spectrometer Facility) for analyses they performed on α -IPM synthase, and Dr E. B. Keller for criticism of the manuscript. This work was supported by grant AI14340 from the National Institutes of Health. R.M.G. was supported by National Institutes of Health training grant GM07273.

REFERENCES

- Bartholomew, J. C. & Calvo, J. M. (1971). *Biochim. Biophys. Acta*, **250**, 577-587.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248-254.
Burns, R. O., Umbarger, H. E. & Gross, S. R. (1963). *Biochemistry*, **2**, 1053-1058.
Calvo, J. M. & Worden, H. E. (1970). *Genetics*, **64**, 199-214.
Calvo, J. M., Freundlich, M. & Umbarger, H. E. (1969a). *J. Bacteriol.* **97**, 1272-1282.
Calvo, J. M., Margolin, P. & Umbarger, H. E. (1969b). *Genetics*, **61**, 777-787.
Edman, P. & Begg, G. (1967). *Eur. J. Biochem.* **1**, 80-91.
Favaloro, J., Treisman, R. & Kamen, R. (1980). *Methods Enzymol.* **65**, 718-749.
Gemmill, R. M., Wessler, S. R., Keller, E. B. & Calvo, J. M. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 4941-4945.
Hertzberg, K. M., Gemmill, R. M., Jones, J. & Calvo, J. M. (1980). *Gene*, **8**, 135-152.
Keller, E. B. & Calvo, J. M. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 6186-6190.
Kohlhaw, G., Leary, T. R. & Umbarger, H. E. (1969). *J. Biol. Chem.* **244**, 2218-2225.
Kupersztoch-Portnoy, Y. M., Lovett, M. A. & Helinsky, D. R. (1974). *Biochemistry*, **13**, 5484-5490.

- Lawther, R. P., Nichols, B., Zurawski, G. & Hatfield, G. W. (1979). *Nucl. Acids Res.* **7**, 2289-2301.
- Lee, F. & Yanofsky, C. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **74**, 4365-4369.
- Margolin, P. (1963). *Genetics*, **48**, 441-457.
- Maxam, A. M. & Gilbert, W. (1980). *Methods Enzymol.* **65**, 499-560.
- Mendez, E. & Lai, C.Y. (1975). *Anal. Biochem.* **68**, 47-53.
- Miller, J. M. (1972). In *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory of Quantitative Biology, Cold Spring Harbor.
- Peattie, D. A. (1979). *Proc. Nat. Acad. Sci., U.S.A.* **76**, 1760-1764.
- Rigby, P. W. J., Dieckman, M., Rhodes, C. & Berg, P. (1977). *J. Mol. Biol.* **113**, 237-251.
- Sanger, F. & Coulson, A. R. (1978). *FEBS Letters*, **87**, 107-110.
- Searles, L. L., Wessler, S. R. & Calvo, J. M. (1983). *J. Mol. Biol.* **163**, 377-394.
- Smith, D. R. & Calvo, J. M. (1979). *Mol. Gen. Genet.* **175**, 31-38.
- Smith, D. R. & Calvo, J. M. (1980). *Nucl. Acids Res.* **8**, 2255-2274.
- Southern, E. M. (1975). *J. Mol. Biol.* **98**, 503-517.
- Steitz, J. A. (1979). In *Biological Regulation and Development* (Goldberger, R. F., ed.), vol. 1, pp 349-399, Plenum Press, New York.
- Wessler, S. R. & Calvo, J. M. (1981). *J. Mol. Biol.* **149**, 579-597.
- Wu, R., Jay, E. & Roychoudhury, R. (1976). In *Methods in Cancer Research* (Busch, H., ed.), vol. 12, pp. 87-176, Academic Press, New York.
- Yanofsky, C. (1981). *Nature (London)*, **289**, 751-758.
- Zain, B. S. & Roberts, R. J. (1978). *J. Mol. Biol.* **120**, 13-31.

Edited by M. Gottesman