Volume 13 Number 11 1985 Nucleic Acids Research

The *ilvB* locus of *Escherichia coli* K-12 is an operon encoding both subunits of acetohydroxyacid synthase I

P.Friden, J.Donegan, J.Mullen, P.Tsui and M.Freundlich
Department of Biochemistry, SUNY, Stony Brook, NY 11794, and

L.Eoyang, R.Weber and P.M.Silverman*
Department of Molecular Biology, Division of Biological Sciences, Albert Einstein College of Medicine, Bronx, NY 10461, USA

Received 4 March 1985; Revised 18 April 1985; Accepted 19 April 1985

ABSTRACT

The ilvB locus of Escherichia coli K-12 encloses two open reading frames defining polypeptides of 60,000 and 11,200 molecular weight. The entire locus, about 2.3 kb, is co-transcribed as an operon. The molecular weights and amino acid compositions of the presumptive operon polypeptides agree with those of the large and small subunit polypeptides of acetohydroxy-acid synthase (AHAS) I, for which ilvB is the structural locus. We reserve the designation ilvB for the promoter proximal (longer) cistron and designate the promoter distal cistron ilvN. The molecular weight and amino acid sequence of the ilvB polypeptide are strikingly similar to those of the IlvI (larger subunit of AHAS III) and IlvG (larger subunit of AHAS II) polypeptides. There is less size uniformity among the IlvN, IlvH (smaller subunit of AHAS III), and IlvM (smaller subunit of AHAS II) polypeptides. Nevertheless, there is significant amino acid sequence homology among the three small subunit polypeptides. Thus, all three AHAS isozymes of E. coli K-12 probably have a common evolutionary origin.

INTRODUCTION

Bacterial acetohydroxyacid synthases catalyze two homologous reactions. One, the formation of acetolactate from two molecules of pyruvate, is part of the biosynthetic pathway to leucine and valine. The other, the formation of acetohydroxybutyrate from one molecule of pyruvate and one of 2-ketobutyrate, is part of the pathway to isoleucine (1).

Of the enzymes necessary for the synthesis of the three branched chain amino acids in K-12 strains of Escherichia coli and LT2 strains of Salmonella typhimurium, only the acetohydroxyacid synthases are represented by multiple genetic loci, each encoding a distinct isozyme (summarized in Ref. 2). In both bacterial species the synthesis of each isozyme is subject to a distinct regulatory pattern (2-4), suggesting that each isozyme has evolved to allow optimal cell growth in a different set of physiologic conditions. As yet, these conditions have not been entirely defined, nor is it evident what catalytic or regulatory features of the respective isozymes have been selected as adaptively important.

© IRL Press Limited, Oxford, England.

Previously (5), we described the purification of AHAS I from E. coli K-12. Enzyme purified to constant specific activity contained two polypeptide chains that were present in constant proportion to each other and to enzyme activity. One of these had a molecular weight of about 60,000 and presumably corresponded to an AHAS I subunit of that molecular weight described by Grimminger and Umbarger (6). The second had a molecular weight of about 10,000. Though neither biochemical (6) nor genetic (7) data had suggested that AHAS I contains two kinds of subunit, we showed that the two polypeptides were immunologically distinct and non-covalently associated with each other in crude extracts. On the basis of these data we proposed that both polypeptides are AHAS I subunits (5).

Here, we show that the ilvB locus of E. coli K-12, previously identified as the structural locus for AHAS I (7,8), is an operon encoding both AHAS I subunits. In addition, it appears that all three acetohydroxyacid synthases of E. coli K-12, including the normally cryptic AHAS II (9), comprise a family of enzymes with a common evolutionary origin.

MATERIALS AND METHODS

DNA sequencing

The $ilvB^+$ plasmid pTCN12 (10) was purified by CsCl equilibrium sedimentation. The position of the ilvB locus within the plasmid was already known from analysis of its transcriptional regulatory sequences (11). Restriction fragments were isolated as previously described (11). They were labeled at the 3' termini by the fill-in reaction catalyzed by the Klenow fragment of E. coli DNA polymerase I in the presence of appropriate [α-^{32}P]-deoxyribonucleotide triphosphates (11) or by terminal deoxynucleotidyl transferase in the presence of [α-^{32}P]dideoxy ATP (32; potassium cacodylate buffer was substituted for the sodium salt). Labeling of 5' termini was catalyzed by polynucleotide kinase in the presence of [γ-^{32}P]ATP (12). DNA sequences were determined as described by Maxam and Gilbert (13), modified for some fragments as previously described (11).

Separation and analysis of AHAS I subunits

Enzyme purified as previously described (5) was dialyzed for 20 hr against 1 L of 20 mM triethylamine bicarbonate buffer, pH 8.3; the dialysis buffer was changed once after five hr. The dialysis tubing (Spectra/Por 3 from American Scientific Products) had a molecular weight cut-off of 3500 to avoid loss of the small AHAS I subunit. After dialysis the protein was concentrated by lyophilization and dissolved in 1.5 ml of 20 mM triethylamine

bicarbonate buffer containing 2% sodium dodecylsulfate and 5% 2-mercapto-ethanol by incubation at 100°C for five min. The sample was then applied to a column (2.2 cm x 78 cm) of Sephadex G200 equilibrated with 20 mM triethyl-amine bicarbonate buffer. Column fractions (4.2 ml) containing protein were examined by SDS/polyacrylamide gel electrophoresis to identify those containing sufficiently pure large or small AHAS I subunits. These were pooled as appropriate, dried by lyophilization, and buffer salts were removed by repeated lyophilization against water.

Samples for amino acid analysis contained about 5 μg of protein and 2 nmol of norleucine as a recovery standard. The samples were vacuum dried several times from a 1% triethylamine suspension to remove ammonia. Hydrolysis was under reduced pressure in 6 N HCl/0.2% phenol at 115°C for 16 hr. The amino acids were analyzed directly by ion exchange chromatography on a Beckman 121M analyzer.

For peptide isolation, purified small subunit (250 μg; 22 nmol) was digested for 24 hr at 37°C with TPCK-treated trypsin (1:25 [w/w] ratio) in 0.15 M NH_4HCO_3, pH 8/8 M urea. Peptides were separated by reverse phase HPLC (Waters C-18 μBondapak resin) as described (14). Amino acid compositions of selected fragments were determined as described above.

The NH_2-terminal amino acid sequence of the large AHAS I subunit (200 μg) was determined by automated gas-phase degradation on an Applied Biosystems sequencer followed by derivatization of the amino acid liberated at each step to its phenylthiohydantoin, which was identified by HPLC (15).

Nuclease S1 protection

Total cell RNA was isolated (34) from *E. coli* K-12 strain MF2701 in mid-logarithmic growth at 37°C in minimal glucose medium supplemented with isoleucine (100 μg/ml), valine (150 μg/ml), and limiting leucine (5 μg/ml). The *ilvB* DNA for hybridization was prepared from a 1.8 kb *Hind* III-*Eco* RI fragment of pTCN12 (10). The *Hind* III terminus of the fragment is located within the coding sequence for the large AHAS I subunit (see ref. 10 and RESULTS). The 3' termini of the DNA fragments were labeled with ^{32}P by the fill-in reaction catalyzed by the Klenow fragment of DNA polymerase I (11). The radioactive DNA was then digested with *DdeI*, which liberates a 1 kb fragment labeled at the *Hind* III terminus in the *ilvB* coding strand. The fragment includes the 3' end of the cistron for the large AHAS I subunit, the intercistronic region, the entire cistron for the small AHAS I subunit, and the transcription terminator for the operon (10). That fragment was purified by polyacrylamide gel electrophoresis (16) and used in nuclease S1

mapping experiments (17,18). Cell RNA (110 μg; see above) was hybridized to the radioactive probe fragment. Nuclease S1 (120 units) treatment was for 20 min at 37°C. Digestions were terminated by addition of EDTA to a final concentration of 4 mM. Nucleic acids were precipitated with ethanol, dissolved in a solution composed of 80% (v/v) formamide/10 mM NaOH/1 mM EDTA/0.0025% bromphenol blue, and separated by electrophoresis through a 5% polyacrylamide/8 M urea gel. Radioactive species were detected by autoradiography.

Materials

Materials for DNA sequencing were obtained from sources previously described (11). Terminal deoxynucleotidyl transferase was purchased from P-L Biochemicals, and nuclease S1 from Bethesda Research Laboratories. [α-^{32}P]-dideoxy ATP (3000-5000 Ci/mmol) was obtained from Amersham/Searle. Materials for AHAS I purification were as previously described (5). TPCK-treated trypsin was purchased from Worthington Biochemicals. Urea (Ultra-pure grade) was obtained from Schwarz/Mann. Materials for protein analyses were as described (14,15).

RESULTS

The ilvB locus of E. coli K-12 is an operon encoding two polypeptide chains

DNA regulatory sequences at the ilvB locus contain the principal sites for controlling AHAS I synthesis (11,19). Protein coding sequences at the ilvB locus are summarized in Fig. 1, and the complete nucleotide sequence is shown in Fig. 2, beginning with the ATG initiation codon of the large AHAS I subunit (see below) and extending 2040 bp to a (G+C)-rich inverted repeat followed by a sequence of T residues that constitutes a possible transcription terminator of the ilvB locus (20; see below). Nucleotide 1 in Fig. 2 corresponds to nucleotide 240 in ref. 11, in which numbering begins at the site of transcription initiation. Hence, a transcript including the entire ilvB locus would be about 2280 bases long.

The region shown in Fig. 2 encloses two extended open reading frames. The first includes 562 codons from nucleotides 1 through 1686. This region is preceded by a sequence corresponding to a ribosome binding site, as previously described (11). The second includes 96 codons from nucleotides 1693 through 1980. This region is preceded by a pentaguanylate sequence (nucleotides 1680-1684) that could function as a ribosome binding site (21). The molecular weights of the two polypeptide chains derived from the two open reading frames are 60,000 and 11,200, in good agreement with the molecular

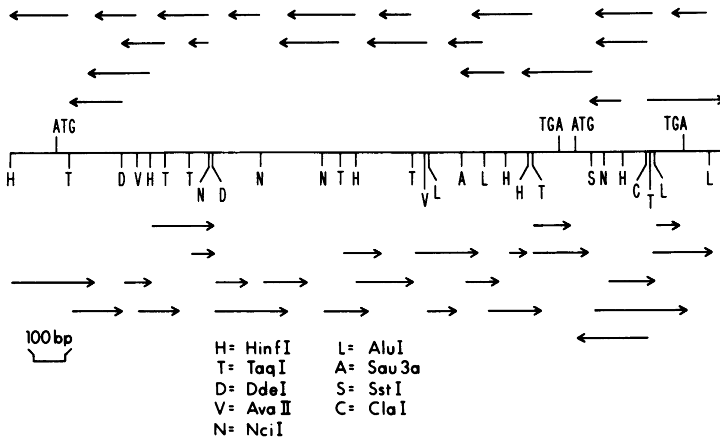

Fig. 1. Schematic structure of the E. coli K-12 ilvB locus (solid line). The single-letter designations below the line indicate restriction endonuclease cleavage sites, as indicated in the key at the bottom of the Figure. The left-hand Hinf I site begins at nucleotide +152 with respect to the site of transcription initiation (11). The ATG and TGA codons above the line indicate the boundaries of two open reading frames (see Fig. 2). The arrows above the line indicate the origin, direction, and extent of determined sequences from the non-coding DNA strand; those below the line, from the coding strand.

weights of the two AHAS I subunits determined by SDS/polyacrylamide gel electrophoresis (5).

Since the two open reading frames are so close together, both are likely to be cotranscribed as an operon. We established this by nuclease S1 mapping to locate the 3' terminus of the in vivo transcript(s) originating from the ilvB locus. Total cell RNA was isolated from a strain grown under conditions of limiting leucine to derepress AHAS I synthesis (22). The RNA was hybridized to a 1 kb DNA fragment from the ilvB locus and labeled at the 3' terminus of the coding strand. The labeled terminus originates from a Hind III site about 400 bp from the end of the ilvB cistron (nucleotides 1199-1204; see Fig. 2), and ends about 200 bp past a presumptive transcription terminator at the end of ilvN (see Fig. 2). As shown in Fig. 3, the predominant radioactive DNA species protected from nuclease S1 by DNA:RNA hybridization is about 850 bases. This observation implies that ilvB and ilvN are cotranscribed, with transcription termination occurring at the presumptive terminator located 840 bp from the Hind III site in ilvB (Fig. 2). If transcription terminated at a significant frequency within the ilvB-ilvN intercistronic region, then a radioactive, S1-resistant frag-

```
1          .          .          .          .          .          60         .          .          .          .          .          120
ATGGCAAGTTCGGGCACAACATCGACGCGTAAGCGCTTTACCGGCGCAGAATTTATCGTTCATTTCTTGGAACAGCAGGGCATTAAGATTGTGACAGGCATTCCGGGCGGTTCTATCCTG
 M  A  S  S  G  T  T  S  T  R  K  R  F  T  G  A  E  F  I  V  H  F  L  E  Q  Q  G  I  K  I  V  T  G  I  P  G  G  S  I  L
121        .          .          .          .          .          180        .          .          .          .          .          240
CCTGTTTACGATGCCTTAAGCCAAAGCACGCAAATCCGCCATATTCTGGCCCGTCATGAACAGGGCGCGGGCTTTATCGCTCAGGGAATGGCGCGCACCGACGGTAAACCGGCGGTCTGT
 P  V  Y  D  A  L  S  Q  S  T  Q  I  R  H  I  L  A  R  H  E  Q  G  A  G  F  I  A  Q  G  M  A  R  T  D  G  K  P  A  V  C
241        .          .          .          .          .          300        .          .          .          .          .          360
ATGGCCTGTAGCGGACCGGGTGCGACTAACCTGGTGACCGCCATTGCCGATGCGCGGCTGGACTCCATCCCGCTGATTTGCATCACTGGTCAGGTTCCCGCCTCGATGATCGGCACCGAC
 M  A  C  S  G  P  G  A  T  N  L  V  T  A  I  A  D  A  R  L  D  S  I  P  L  I  C  I  T  G  Q  V  P  A  S  M  I  G  T  D
361        .          .          .          .          .          420        .          .          .          .          .          480
GCCTTCCAGGAAGTGGAACCCTACGGCATCTCTATCCCCATCACCAAACACAACTATCTGGTCAGACATATCGAAGAACTCCCGCAGGTCATGAGCGATGCCTTCCGCATTGCGCAATCA
 A  F  Q  E  V  D  T  Y  G  I  S  I  P  I  T  K  H  N  Y  L  V  R  H  I  E  E  L  P  Q  V  M  S  D  A  F  R  I  A  Q  S
481        .          .          .          .          .          540        .          .          .          .          .          600
GGCCGCCCAGGCCCGGTGCTGATAGACATTCCTAAGGATGTGCAAACGGCAGTTTTTGAGATTGAAACACAGCCCGCTATGGCAGAAAAAGCCGCCGCCCCCGCCTTTAGCGAAGAAAGC
 G  R  P  G  P  V  W  I  D  I  P  K  D  V  Q  T  A  V  F  E  I  E  T  Q  P  A  M  A  E  K  A  A  A  P  A  F  S  E  E  S
601        .          .          .          .          .          660        .          .          .          .          .          720
ATTCGTGACGCAGCGGCGATGATTAACGCTGCCAAACGCCCGGTGCTTTATCTGGGCGGCGGTGTGATCAATGCGCCCGCACGGGTGCGTGAACTGGCGGAGAAAGCGCAACTGCCTACC
 I  R  D  A  A  A  M  I  N  A  A  K  R  P  V  L  Y  L  G  G  G  V  I  N  A  P  A  R  V  R  E  L  A  E  K  A  Q  L  P  T
721        .          .          .          .          .          780        .          .          .          .          .          840
ACCATGACTTTAATGGCGCTGGGCATGTTGCCAAAAGCGCATCCGTTGTCGCTGGGTATGCTGGGGATGCACGGCGTGCGCAGCACCAACTATATTTTGCAGGAGGCGGATTTGTTGATA
 T  M  T  L  M  A  L  G  M  L  P  K  A  H  P  L  S  L  G  M  L  G  M  H  G  V  R  S  T  N  Y  I  L  Q  E  A  D  L  L  I
841        .          .          .          .          .          900        .          .          .          .          .          960
GTGCTCGGTGCGCGTTTTGATGACCGGGCGATTGGCAAAACCGAGCAGTTCTGTCCGAATGCCAAAATCATTCATGTCGACATCGACCGTGCACAGCTGGGTAAAATCAAGCAGCCGCAC
 V  L  G  A  R  F  D  D  R  A  I  G  K  T  E  Q  F  C  P  N  A  K  I  I  H  V  D  I  D  R  A  Q  L  G  K  I  K  Q  P  H
961        .          .          .          .          .          1020       .          .          .          .          .          1080
GTGGCGATTCAGGCGGATGTTGATGACGTGCTGGCGCAGTTGATCCCGCTGGTGGAAGCGCAACCGCGTGCAGAGTGGCACCAGTTGGTAGCGGATTTGCAGCGTGAGTTTCCGTGTCCA
 V  A  I  Q  A  D  V  D  D  V  L  A  Q  L  I  P  L  V  E  A  Q  P  R  A  E  W  H  Q  L  V  A  D  L  Q  R  E  F  P  C  P
1081       .          .          .          .          .          1140       .          .          .          .          .          1200
ATCCCGAAAGCGTGCGATCCGTTAAGCCATTACGGCCTGATCAACGCCGTTGCCGCCTGTGTCGATGACAATGCAATTATCACCACCGACGTTGGTCAGCATCAGATGTGGACCGCGCAA
 I  P  K  A  C  D  P  L  S  H  Y  G  L  I  N  A  V  A  A  C  V  D  D  N  A  I  I  T  T  D  V  G  Q  H  Q  M  W  T  A  Q
1201       .          .          .          .          .          1260       .          .          .          .          ..         1320
GCTTATCCGCTCAATCGCCCACGCCAGTGGCTGACCTCCGGTGGGCTGGGCACGATGGGTTTTGGCCTGCCTGCGGCGATTGGCGCTGCGCTGGCGAACCCGGATCGCAAAGTGTTGTGT
 A  Y  P  L  N  R  P  R  Q  W  L  T  S  G  G  L  G  T  M  G  F  G  L  P  A  A  I  G  A  A  L  A  N  P  D  R  K  V  L  C
1321       .          .          .          .          .          1380       .          .          .          .          .          1440
TTCTCCGGCGACGGCAGCCTGATGATGAATATTCAGGAGATGGCGACCGCCAGTGAAAATCAGCTGGATGTCAAAATCATTCTGATGAACAACGAAGCGCTGGGGCTGGTGCATCAGCAA
 F  S  G  D  G  S  L  M  M  N  I  Q  E  M  A  T  A  S  E  N  Q  L  D  V  K  I  I  L  M  N  N  E  A  L  G  L  V  H  Q  Q
1441       .          .          .          .          .          1500       .          .          .          .          .          1560
CAGAGTCTGTTCTACGAGCAAGGCGTTTTTGCCGCCACCTATCCGGGCAAAATCAACTTTATGCAGATTGCCGCCGGATTCGGCCTCGAAACCTGTGATTTGAATAACGAAGCCGATCCG
 Q  S  L  F  Y  E  Q  G  V  F  A  A  T  Y  P  G  K  I  N  F  M  Q  I  A  A  G  F  G  L  E  T  C  D  L  N  N  E  A  D  P
1561       .          .          .          .          .          1620       .          .          .          .          .          1680
CAGGCTTCATTGCAGGAAATCATCAATCGCCCTGGCCCGGCGCTGATCCATGTGCGCATTGATGCCGAAGAAAAAGTTTACCCGATGGTGCCGCCAGGTGCGGCGAATACTGAAATGGTG
 Q  A  S  L  Q  E  I  I  N  R  P  G  P  A  L  I  H  V  R  I  D  A  E  E  K  V  Y  P  M  V  P  P  G  A  A  N  T  E  M  V
1681       .          .          .          .          .          1740       .          .          .          .          .          1800
GGGGAATAAGCCATGCAAAACACAACTCATGACAACGTAATTCTGGAGCTCACCGTTCGCAACCATCCGGGCGTAATGACCCACGTTTGTGGCCTTTTTGCCCGCCGCGCTTTTAACGTT
 G  E  *     M  Q  N  T  T  H  D  N  V  I  L  E  L  T  V  R  N  H  P  G  V  M  T  H  V  C  G  L  F  A  R  R  A  F  N  V
1801       .          .          .          .          .          1860       .          .          .          .          .          1920
GAAGGCATTCTTTGTCTGCCGATTCAGGACAGCGACAAAAGCCATATCTGGCTACTGGTCAATGACGACCAGCGTCTGGAGCAGATGATAAGCCAAATCGATAAGCTGGAAGATGTCGTG
 E  G  I  L  C  L  P  I  Q  D  S  D  K  S  H  I  W  L  L  V  N  D  D  Q  R  L  E  Q  M  I  S  Q  I  D  K  L  E  D  V  V
1921       .          .          .          .          .          1980       .          .          .          .          .          2040
AAAGTGCAGCGTAATCAGTCCGATCCGACGATGTTTAACAAGATCGCGGTGTTTTTTCAGTAACCGCTCAAGGCTTGAACAACATCGCGCTTATCGTTAAGGTAACGCGTATTTTTTTT
 K  V  Q  R  N  Q  S  D  P  T  M  F  N  K  I  A  V  F  F  Q  *
                                                                                               ------->             <-------    --------
2041       .          .          .          .  2085
ACCCGCAGGACAAGACCATGATCACCGTTGCCCTTATAGACGA
```

Fig. 2. DNA nucleotide sequence of the ilvB locus of E. coli K-12. The sequence of the non-coding strand is shown. The sequence begins at the promoter-proximal ATG initiation codon of the large AHAS I subunit (IlvB); these are nucleotides 240-242 with respect to the site of transcription initiation (11). The single-letter designations below the sequence indicate the amino acids assigned to each codon of the ilvB and ilvN cistrons. In-frame translation termination codons (see Fig. 1) are indicated by an asterisk. The facing arrows between nucleotides 2008 and 2030 indicate a (G+C)-rich inverted repeat, and the underlined nucleotides 2033-2040 indicate an adjacent oligothymidylate sequence that constitutes the transcription terminator for the ilvB ilvN operon (see Fig. 3). Wek, Hauser, and Hatfield, in the accompanying communication, have obtained essentially the same sequence.

ment of only 480 nucleotides would have been observed.

Additional evidence that synthesis of the IlvB and IlvN polypeptides is coordinate, and that both are subunits of AHAS I (see below), is that both polypeptides are overproduced by E. coli K-12 strain MF2361, which contains the multicopy ilvB ilvN plasmid pTCN12 (10) and is constitutive for AHAS I

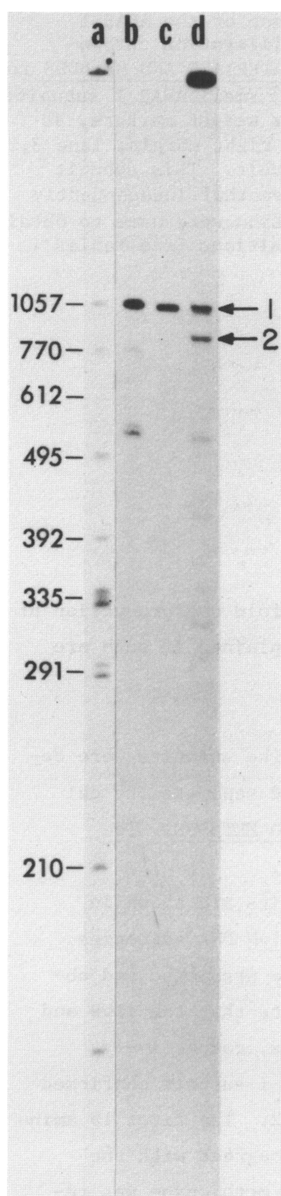

Fig. 3. Nuclease S1 mapping of the 3' terminus of in vivo mRNA from the ilvB locus. Total cell RNA was isolated from strain MF2701 grown under conditions of limiting leucine. The RNA was hybridized to a Hind III-Dde I fragment labeled with ^{32}P at its Hind III 3' end. The Hind III terminus of the fragment lies within the ilvB coding sequence, 840 nucleotides from a presumptive transcription terminator at the end of the ilvN cistron (see Fig. 2 and the text for details). The hybridization mixture was treated with nuclease S1 and analyzed by electrophoresis on 5% polyacrylamide-8 M urea gels. Lane a, ^{32}P-labeled ØX174 DNA cut with Hinc II and used as DNA size markers. The size of the markers, in number of nucleotides, is indicated at the left of the bands. Lane b, 3' end-labeled Hind III-Dde I probe; lane c, tRNA control. E. coli tRNA used in place of total cellular RNA in the S1 mapping procedure. Lane d, 110 μg of total cellular RNA from MF2701. Arrows: (1), Hind III-Dde I probe; (2) nuclease S1-resistant DNA of approximately 850 nucleotides.

synthesis owing to a chromosomal mutation, possibly a valyl-tRNA synthetase (10). Consequently, AHAS I comprises about 10% of the soluble protein from MF2361 (10; L. Eoyang, unpublished observation). AHAS I purified from MF2361 had the same polypeptide composition as previously reported for enzyme from a strain with only about 3% of the initial specific activity (5;

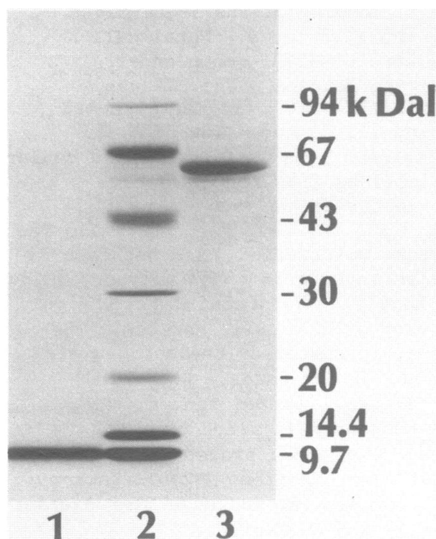

Fig. 4. Separation of the AHAS I subunits by gel filtration chromatography. See MATERIALS AND METHODS for details. Lane 1, small AHAS I subunit; lane 2, molecular weight markers, as indicated at the right margin; lane 3, large AHAS I subunit. This subunit preparation and another independently isolated preparation were used to obtain amino acid compositions (see Tables 1 and 2).

L. Eoyang, unpublished observation). The fact that 30-fold overproduction of the enzyme includes both polypeptides is most easily explained if both are enzyme subunits encoded in a single operon.

ilvB and ilvN encode AHAS I subunits

AHAS I was purified as previously described (5). The subunits were denatured by incubation in SDS at elevated temperature and separated by gel filtration chromatography, as described in MATERIALS AND METHODS. The separation was essentially complete, as shown in Fig. 4.

The amino acid compositions of the separated subunits are shown in Table 1, along with those predicted from the ilvB and ilvN DNA sequences. These comparisons, along with the agreements between the predicted and observed molecular weights of the AHAS I subunits, indicate that the ilvB and ilvN cistrons encode the large and small AHAS I subunits, respectively.

Amino-terminal sequence analysis of the large AHAS I subunit confirmed the ilvB translation initiation site indicated in Fig. 2. The first 19 amino acids, for which unambiguous assignments could be made, agreed with the sequence shown in Fig. 2 except that the amino-terminal methionine was removed in vivo, so that the amino-terminus of the subunit is alanine (data not shown).

We were unable to obtain an amino-terminal sequence for the small AHAS I subunit, perhaps owing to a blocking group. We therefore determined the amino acid compositions of tryptic fragments of the protein isolated by

Table 1. Amino Acid Compositions of AHAS I Subunits from E. coli K-12

Amino acid	Abundance					
	From protein[a] (mol/mol large subunit)		From DNA	From protein[a] (mol/mol small subunit)		From DNA
	1	2		1	2	
ASX	49.9	48.8	48	14.8	15.3	15
THR	28.4	28.7	29	5.2	5.5	5
SER	21.6	21.6	23	4.8	4.9	4
GLX	65.0	67.7	64	13.8	14.0	12
PRO	40.0	36.0	37	4.1	3.9	3
GLY	45.3	46.3	45	5.9	5.9	3
ALA	71.2	72.7	71	6.6	5.7	3
VAL	29.4	33.4	34	9.1	9.4	10
MET	17.0	20.6	21[c]	3.5	3.3	4[c]
ILE	41.6	43.3	47	6.9	6.7	7
LEU	48.5	50.0	47	10.0	9.9	9
TYR	9.5	10.6	10	0.5	0.5	0
PHE	16.6	18.4	17	5.1	5.1	5
HIS	10.9	13.8	14	2.8	3.1	4
LYS	23.6	20.4	18	5.3	4.4	4
ARG	25.4	25.7	24	5.7	5.6	5
TRP	ND[b]	ND[b]	4	ND	ND	1
CYS	ND	ND	9	ND	ND	2

[a]Each set of values is the average of triplicate determinations. The two sets of values for each subunit represent analyses from separate purifications. Based on molecular weights of 60,000 and 11,200.

[b]ND = not determined.

[c]Includes NH_2-terminal methionine. The NH_2-terminal methionine of the large subunit is removed in vivo and not present in the purified protein. The NH_2-terminal methionine of the small subunit is not removed in vivo.

reverse phase HPLC, and compared the results to compositions expected from the DNA sequence of Fig. 2. The data for four peptides, encompassing half the small subunit, are shown in Table 2. Each composition could be matched perfectly with the composition of a tryptic peptide expected from the ilvN sequence. Note that T31 corresponds to the amino-terminus of the small subunit and T25 to the carboxyl-terminus. These data establish the identity of the IlvB and IlvN polypeptides as the subunits of AHAS I.

E. coli acetohydroxyacid synthases comprise a family of related isozymes

Complete DNA sequence data are now available for all three acetohydroxyacid synthase isozymes of E. coli K-12. Recently, Squires et al. (23), noting substantial amino acid sequence homology between IlvI (large subunit of AHAS III) and IlvG (large subunit of AHAS II (24)) suggested a common evolutionary origin for both polypeptides. Fig. 5 extends their analysis to include the IlvB polypeptide. Owing to the slightly different sizes of the polypeptides and to short gaps introduced to align obviously homologous

Table 2. Amino Acid Compositions of Tryptic Peptides from the Small AHAS I Subunit

Amino acid	T31: Met1-Arg16			T34: Leu62-Lys77			T17: Asn81-Lys90			T25: Ile91-Gln96		
	nmol	mol/mol peptide		nmol	mol/mol peptide		nmol	mol/mol peptide		nmol	mol/mol peptide	
		Obs.	Expected		Obs.	Expected		Obs.	Expected		Obs.	Expected
ASX	33.4	3.1	3	12.9	2.3	2	7.0	3.1	3	-	-	-
THR	26.7	2.4	3	-	-	-	13.9	0.9	1	-	-	-
SER	-	-	-	4.4	0.8	1	12.2	0.8	1	-	-	-
GLX	21.3	2.0	2	23.6	4.3	4	16.1	1.1	1	12.6	1.0	1
PRO	-	-	-	-	-	-	18.8	1.2	1	-	-	-
ALA	-	-	-	-	-	-	-	-	-	12.2	1.0	1
VAL	19.9	1.8	2	9.6	1.7	2	-	-	-	11.4	0.9	1
MET	8.1	0.7	1	4.3	0.8	1	12.7	0.8	1	-	-	-
ILE	9.1	0.8	1	10.9	2.0	2	-	-	-	11.8	0.9	1
LEU	20.7	1.9	2	11.4	2.0	2	-	-	-	-	-	-
PHE	-	-	-	-	-	-	15.2	1.0	1	22.7	1.8	2
HIS	10.8	1.0	1	-	-	-	-	-	-	-	-	-
LYS	-	-	-	11.0	2.0	2	9.1	0.6	1	-	-	-
ARG	10.9	1.0	1	-	-	-	-	-	-	-	-	-

regions, there are 540 positions where we could compare all three polypeptides. Of these, the same amino acid is found in all three polypeptides in 159 positions (29% of the total; boxed positions in Fig. 5). In another 122 positions, 23% of the total, the same amino acid is found in the IlvB polypeptide and one of the other two (underlined positions). The overall degree of homology is not equally distributed over the entire lengths of the three polypeptides; there are extended regions with >50% sequence identity among all three, and other regions show much less homology. But the degree of overall homology indicates that the *ilvB*, *ilvG*, and *ilvI* loci of *E. coli* K-12 encode a family of closely-related polypeptides.

There is less overall homology among the polypeptides encoded by the *ilvN*, *ilvM* (region K of Squires *et al.* (23)), and *ilvH* loci. These loci encode the small subunits of AHAS I (this study), AHAS II (23,24) and AHAS III (23,25), respectively. However, there is a region near the COOH-terminus of the IlvN and IlvM polypeptides and near the middle of the IlvH polypeptide that does appear to be significantly homologous among all three (Fig. 6). While the region is not itself extensive, it does represent more than 20% of the IlvN and IlvM polypeptides.

DISCUSSION

Our data provide important confirmatory evidence for the subunit composition of AHAS I, which we previously inferred from biochemical data alone (5). Whereas previous information (6,7) left no reason to believe that

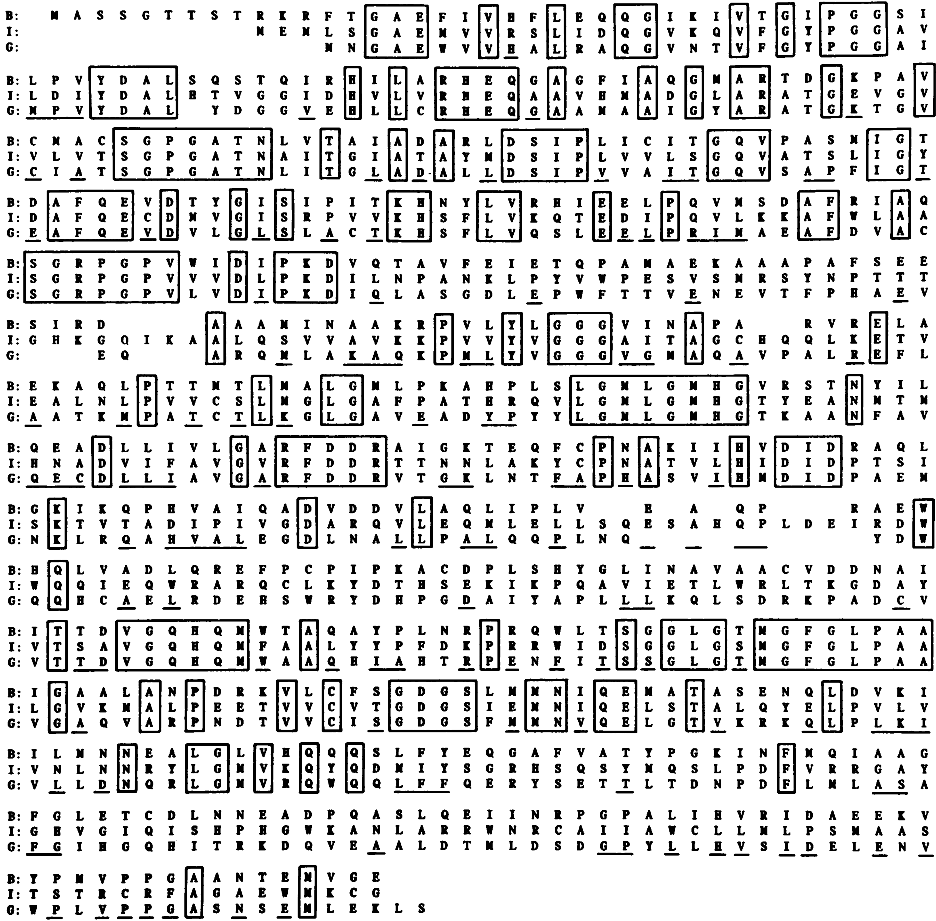

Fig. 5. Amino acid sequence comparison of the IlvB, IlvI and IlvG polypeptides. The IlvI sequence and IlvG sequence were taken from references 23 and 9, respectively. The NH_2-terminus of the IlvG polypeptide is based on a comparison with the NH_2-terminal sequence determined for the cognate polypeptide from Salmonella typhimurium (24). Boxed regions indicate positions where all three polypeptides have the same amino acid; underlined regions indicate positions where the IlvB polypeptide and either the IlvI or the IlvG polypeptide have the same amino acid. For a detailed comparison of the IlvI and IlvG polypeptides, see reference 23.

AHAS I was composed of more than one kind of subunit, both the enzyme and its genetic locus are more complex. Nucleotide sequence analysis, nuclease S1 mapping, and analysis of the AHAS I subunit polypeptides showed that the ilvB

IlvN: Q m i S Q i d K L e D V v k V q r n Q S
IlvH: Q i e k Q L h K L V D V l r V s e l g q
IlvM: l l f S Q L n K L V D V a h V a i c Q S

Fig. 6. Amino acid sequence comparison of the IlvN, IlvH and IlvM polypeptides. The IlvH and IlvM (region K) sequences are from reference 23. The IlvN sequence is from Fig. 2. The region indicated includes amino acids 64-93 (from the NH_2-terminus) of IlvN, 58-77 of IlvM, and 59-78 of IlvH. Upper-case symbols indicate amino acid identity between at least two of the three polypeptides.

locus of E. coli K-12 is an operon composed of two cistrons, each encoding an AHAS I subunit. We reserve the designation ilvB for the promoter-proximal cistron, and designate the promoter-distal cistron ilvN.

The subunit composition of E. coli K-12 AHAS I is analogous to that of AHAS III, insofar as both enzymes consist of large and small subunits (17, 20). Moreover, accumulating evidence supports an analogous structure for AHAS II. Squires et al. (23) noted that the region (region K) between ilvG, the structural locus for an AHAS II polypeptide in $IlvG^+$ strains of E. coli K-12 (9), and ilvE includes an open reading frame of 86 codons. Kutney and Schloss (24) reported that purified AHAS II from Salmonella typhimurium could be resolved into two polypeptides. The amino-terminal sequence of the smaller of the two Salmonella AHAS II polypeptides (M_r=10,000; (24)) is strikingly homologous to that predicted from the E. coli K-12 region K open reading frame noted by Squires et al. (23). It thus appears that the aceto-hydroxyacid synthases of E. coli K-12 and probably of Salmonella comprise a family of related enzymes, each composed of a large subunit with a molecular weight very close to 60,000 (Fig. 5) and a smaller subunit.

This general structure is not limited to acetohydroxyacid synthases from enteric bacteria. An AHAS from Pseudomonas aeruginosa, although different in significant respects from any of the E. coli K-12 isozymes, is similarly composed of two kinds of subunit; one with a molecular weight of 60,000, and the other, of 15,000 (21). For the Pseudomonas enzyme, as for the enteric bacterial enzymes, the roles of the different subunits in enzyme function are not entirely clear. At least for E. coli K-12 AHAS III, the catalytic activities of the enzyme reside in the larger subunit (25) and, in view of the extensive homologies among the large subunits of the different enteric bacterial isozymes, the same is likely true of AHAS I and II.

LaRossa and Smulski (29) have proposed a unique evolutionary origin for enteric bacterial AHAS I on the basis of the resistance of that enzyme and

the sensitivity of AHAS II and AHAS III to inhibition by sulfometuron methyl. The data presented here provide no support for that hypothesis, especially as it relates to the large subunits of the enteric bacterial acetohydroxyacid synthases. The large subunit of AHAS I shows as much homology to those of AHAS II and AHAS III as the last two do to each other.

The small subunits of the E. coli K-12 isozymes show less homology and uniformity in molecular weight than the large subunits. Nevertheless, in view of the regions of homology that do exist among the respective small subunits, it is likely that they too evolved from a common ancestor, but have diverged, along with selected regions of the large subunits, to provide the isozymes with different regulatory properties. The pairing of genes for one small and one large AHAS subunit in genetically co-regulated units is consistent with this view of the origin of the different isozymes. In addition, this arrangement provides more precise control over subunit stoichiometries and greater physiological flexibility than a system consisting of a single kind of large subunit, encoded by a single structural gene, whose activity is modulated by three different small subunits.

This general idea readily accommodates the fact that different physiological conditions favor the synthesis of different isozymes. Thus, AHAS I synthesis in E. coli K-12 (3) and in S. typhimurium LT2 (4) is regulated by cyclic AMP and cyclic AMP receptor protein. Insofar as that mechanism operates in enteric bacteria to ensure an adequate cellular carbon flow in the presence of secondary carbon/energy sources (27), AHAS I may play a specific role in directing carbon flow towards the synthesis of the branched chain amino acids. Hauser and Hatfield (28) discussed the regulation of AHAS I synthesis along broadly similar lines.

The cellular activity of AHAS I, but not of AHAS III, is prevented by mutation in two genes, cpxA and cpxB, that are unlinked to the ilvB ilvN operon or to any other ilv locus (30). We have presented evidence elsewhere that the cpx mutations alter the cellular level of a factor, perhaps a cell envelope component, that limits AHAS I activity in vivo (31). The effect of the cpx mutations on the function of AHAS I, but not that of AHAS III, suggests that at least those two isozymes are not entirely interchangeable in the cell. Whether the distinctive physiological properties of the different isozymes reflect structural differences among their large subunits, their small subunits, or, as seems most likely, both, remains to be determined.

ACKNOWLEDGEMENTS

We are grateful to K. Williams and the staff of the Protein Analysis Facility at Yale University. This work was supported at the State University of New York at Stony Brook by NIH grant GM-17152 to M. F. and at the Albert Einstein College of Medicine by NIH grants GM-11301, CA-13330 and GM-30549. P. M. S. is a Career Scientist Awardee of the Irma T. Hirschl Trust.

*To whom correspondence should be addressed

REFERENCES

1. Umbarger, H. (1978) Ann. Rev. Biochem. 47, 533-606.
2. Squires, C., De Felice, M., Lago, C. and Calvo, J. (1983) J. Bacteriol. 154, 1054-1063.
3. Sutton, A. and Freundlich, M. (1980) Molec. Gen. Genet. 178, 179-183.
4. Weinberg, R. and Burns, R. (1984) J. Bacteriol. 160, 833-841.
5. Eoyang, L. and Silverman, P. (1984) J. Bacteriol. 157, 184-189.
6. Grimminger, H. and Umbarger, H. (1979) J. Bacteriol. 137, 846-853.
7. Newman, T. and Levinthal, M. (1980) Genetics 96, 59-77.
8. Guardiola, J., De Felice, M., Lamberti, A. and Iaccarino, M. (1977) Molec. Gen. Genet. 156, 17-25.
9. Lawther, R., Calhoun, D., Adams, C., Hauser, J., Gray, J. and Hatfield, G. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 922-925.
10. Newman, T., Friden, P., Sutton, A. and Freundlich, M. (1982) Molec. Gen. Genet. 186, 378-384.
11. Friden, P., Newman, T. and Freundlich, M. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6156-6160.
12. Maniatis, T., Fritsch, E. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, N. Y.
13. Maxam, A. and Gilbert, W. (1977) Methods Enzymol. 65, 499-560.
14. Williams, K., Murphy, J. and Chase, J. (1984) J. Biol. Chem. 259, 11804-11811.
15. Merrill, B., Williams, K., Chase, J. and Konigsberg, W. (1984) J. Biol. Chem. 259, 10850-10856.
16. Friden, P., Tsui, P., Okamoto, K. and Freundlich, M. (1984) Nucl. Acids Res. 12, 8145-8160.
17. Berk, A. J. and Sharp, P. A. (1977) Cell 22, 721-732.
18. Favaloro, J., Triesman, R. and Kamen, R. (1980) Methods Enzymol. 65, 718-750.
19. Hauser, J. and Hatfield, G. (1983) Nucl. Acids Res. 11, 127-140.
20. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353.
21. Steitz, J., Sprague, K., Steege, D., Yuan, R., Laughrea, M., Moore, P. and Whaba, A. (1977) in Nucleic Acid-Protein Recognition, Vogel, H. Ed., pp. 491-508, Academic Press, New York.
22. De Felice, M., Newman, T. and Levinthal, M. (1978) Biochim. Biophys. Acta 541, 1-8.
23. Squires, C., De Felice, M., Devereux, J. and Calvo, J. (1984) Nucl. Acids Res. 11, 5299-5313.
24. Kutney, R. and Schloss, J. (1984) Fed. Proc. 43, 2630a.
25. De Felice, M., Guardiola, J., Esposito, B. and Iaccarino, M. (1974) J. Bacteriol. 120, 1068-1077.

26. Arfin, S. and Koziell, D. (1973) Biochim. Biophys. Acta 321, 356-360.
27. Botsford, J. (1981) Microbiol. Rev. 45 620-642.
28. Hauser, C. and Hatfield, G. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 76-79.
29. LaRossa, R. and Smulski, D. (1984) J. Bacteriol. 160, 391-394.
30. Mc Ewen, J. and Silverman, P. M. (1980) J. Bacteriol. 144, 68-73.
31. Sutton, A., Newman, T., Mc Ewen, J., Silverman, P. and Freundlich, M. (1982) J. Bacteriol. 151, 976-982.
32. Yousof, S., Carroll, A. and Clarke, B. (1984) Gene 27, 309-313.
33. Coleman, J., Green, P. J. and Inouye, M. (1984) Cell 37, 429-436.