Volume 15 Number 5 1987 Nucleic Acids Research

The complete nucleotide sequence of the *ilvGMEDA* operon of *Escherichia coli* K-12

Robert P.Lawther, Ronald C.Wek[1], John M.Lopes, Ruth Pereira, Bruce E.Taillon and G.Wesley Hatfield[1]

Department of Biology, University of South Carolina, Columbia, SC 29208 and [1]Department of Microbiology and Molecular Genetics, California College of Medicine, University of California, Irvine, CA 92717, USA

Received August 20, 1986; Revised January 12, 1987; Accepted January 21, 1987

ABSTRACT

In this report we present the complete nucleotide sequence of the ilvGMEDA operon of Escherichia coli. This operon contains five genes encoding four of the five enzymes required for the biosynthesis of isoleucine and valine. We identify and describe the coding regions for these five structural genes and the structural and functional features of the flanking and internal regulatory regions of this operon. This new information contributes to a more complete understanding of the overall control of the biosynthesis of isoleucine and valine.

I. INTRODUCTION

The biosynthesis of isoleucine and valine in Escherichia coli and other bacteria, fungi and plants, that have been studied, occurs by the parallel pathway shown in figure 1. The pathway is parallel because the last four steps of isoleucine synthesis are catalyzed by bifunctional enzymes which also function to convert two pyruvate molecules into valine. Additionally, the α-keto acid precursor of valine, α-ketoisovalerate, is a branchpoint intermediate which can be converted to leucine by the four enzymes encoded by the genes of the leucine operon. The first step of the isoleucine biosynthetic pathway, the conversion of threonine to α-ketobutyrate, is not required for the biosynthesis of valine or leucine (for reviews see refs. 1 and 2).

Carbon flow through the isoleucine pathway is metabolically regulated by isoleucine end-product inhibition of the first enzyme of this pathway, threonine deaminase (1). Similarly, excessive carbon flow through the valine pathway is controlled by valine end-product inhibition of two of three acetohydroxy acid synthase (AHAS) isozymes, AHAS I and AHAS III, that catalyze the first step of the parallel pathway (Fig. 1) (3). The third isozyme, AHAS II is resistant to end-product inhibition by valine. However, unlike other enteric bacteria, and indeed other strains of E. coli, AHAS II is not expressed in the E. coli K-12 strain because of a frameshift mutation in the ilvG gene which encodes the large subunit of this enzyme (4-6). This explains why E. coli K-12 is growth inhibited by exogenous valine; that is, the remaining two active AHAS isozymes are both inhibited by valine which blocks the synthesis of isoleucine.

The multiplicity of the AHAS isozymes appears to provide differential regulation of the flow of carbon through the first enzymatic step of the parallel pathway (Fig. 1). Each of the AHAS isosymes are distinguished by the following catalytic and regulatory properties. First, as previously mentioned, AHAS I and III are sensitive to valine feedback inhibition, whereas AHAS II is not regulated by valine. Second, it is thought

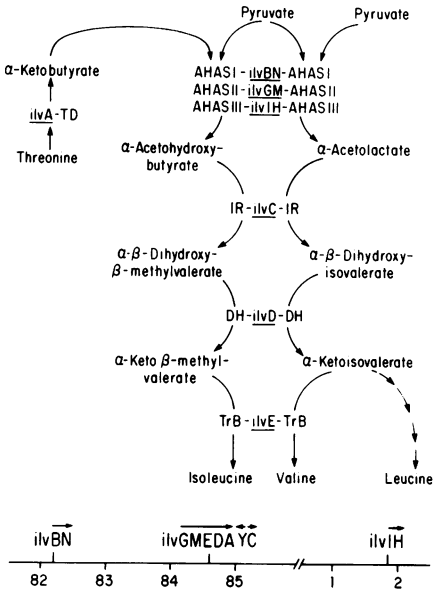

Figure 1. The pathways for the synthesis of isoleucine, leucine and valine in E. coli. The enzymes involved in the isoleucine and valine biosynthesis are abbreviated as follows: TD, threonine deaminase; AHAS, acetohydroxy acid synthase; IR, acetohydroxy acid isomeroreductase; DH, dihydroxy acid dehydrase; TrB, transaminase B. The genes encoding each of these enzymes are indicated next to the enzyme abbreviation and the chromosomal locations are shown below (7). The ilvY gene encodes a positive activator of ilvC transcription (8,9). Arrows above genes indicate the direction of transcription of these loci.

that the AHAS isozymes differ in their substrate specificities for α-ketobutyrate and pyruvate (3). Finally, the expression of genes encoding the AHAS isozymes are regulated in different manners. The ilvBN operon (AHAS I) is regulated by both catabolite repression (10) and an attenuator mechanism regulated by the intracellular levels of aminoacylated $tRNA^{Val}$ and $tRNA^{Leu}$ (11,12). The ilvIH (AHAS III) operon is repressed by leucine by a mechanism not yet fully understood (13) and the ilvGMEDA operon, which encodes the subunits of AHAS II (ilvG and ilvM), is regulated by all three branched chain amino acids via attenuation (14-16). It is not yet fully understood how these differences between the AHAS isozymes function in the regulation of the synthesis of the branched chain amino acids during different physiological conditions.

The ilvGMEDA operon of E. coli might be thought of as the "backbone" of the ilv regulon because this operon encodes four of the five enzymes required for the biosynthesis of isoleucine and valine (Fig. 1). A schematic diagram detailing the structure of this operon is presented in figure 2. The locations of the regulatory regions and structural genes are numbered relative to the in vivo transcriptional start site. In the following sections we present the complete nucleotide sequence of the ilvGMEDA operon. These sequences are discussed in the context of the regions important for the

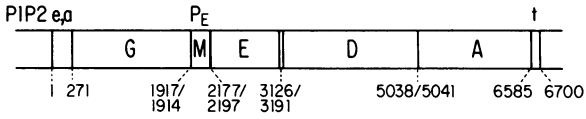

Figure 2. A schematic representation of the ilvGMEDA operon of E. coli. The tandem in vitro promoters P1P2, leader polypeptide (e), attenuator (a), internal promoter, P_E and transcriptional terminator (t) are indicated above the structural genes. Nucleotides are numbered from the in vivo transcriptional initiation site (16). The number preceding each gene corresponds to the first nucleotide of the initiation codon and the number following each gene corresponds to the last nucleotide of the termination codon.

regulation of operon expression (Section II) and the five structural genes encoded in the operon (Section III). In the last section (Section IV) we discuss future questions regarding the mechanisms of regulation of operon expression.

II. REGULATORY REGIONS OF THE ilvGMEDA OPERON

A. Transcription promoters

The nucleotide sequence of the proximal promoter region of the operon is presented in figure 3. This promoter region consists of tandem in vitro promoters P1 and P2, which are separated by 72 base pairs (14,16,18). While in vitro RNA polymerase initiates transcription at the P1 promoter at about 50% of the level of the downstream P2 promoter, RNA fingerprint analyses of in vivo transcripts from the ilvGMEDA attenuator region demonstrates that less than 1% of in vivo transcription initiates at the P1 promoter during several growth conditions known to affect operon expression (16). Therefore, it does not appear that differential expression of the tandem promoters can be a mechanism for operon regulation. However, using multicopy galK transcriptional fusions, it has been demonstrated that sequences upstream of the P2 promoter are required for maximal in vivo expression (18).

```
      -450       -440       -430       -420       -410       -400       -390       -380
AGCTTGGCTTC TTTTACCGTT GTTTCTGGTA AGCCCACCAT CGTTAAGCCG GGTAGACCTT TACTGATATG TACCTCAACA

      -370       -360       -350       -340       -330       -320       -310       -300
GTGATCGGGG GCGCATTTAC TCCCAGGGCT GCGCGGGTAT GAACAATTGA CAGTGACATA AGCCCTCCTT GAGTCACCAT

      -290       -280       -270       -260       -250       -240       -230       -220
TATGTGCATA AGATATCGCT GCTGTAGCCC GCTAATTCGT GAATTTTAGT GGCTGATTCC TGTTTATTTG TGCAAGTGAA

      -210       -200       -190       -180       -170       -160       -150       -140
GTTGAGTTGT TCTGGCGGTG GAATGATGCT CGCAAAAATG CAGCGGACAA AGGATGAACT ACGAGGAAGG GAACAACATT

      -130       -120       -110       -100        -90        -80        -70        -60
CATACTGAAA TTGAATTTTT TTCACTCACT ATTTTATTTT TAAAAAACAA CAATTTATAT TGAAATTATT AAACGCATCA
                                  -35                   -10        P1

       -50        -40        -30        -20        -10         +1         10
TAAAAATCGG CCAAAAAATA TCTTGTACTA TTTACAAAAC CTATGGTAAC TCTTTAGGCA TTCCTTCGAA
                      -35                   -10        P2
```

Figure 3. Nucleotide sequence of the promoter regions of the ilvGMEDA operon. This sequence begins at an AluI restriction site at nucleotide -460. The nucleotides are numbered corresponding to the in vivo transcriptional start site (P2 promoter). The transcriptional initiation site of the upstream in vitro promoter, P1, is indicated at nucleotide -72. The conserved -10 and -35 hexamer regions (17) of the P1 and P2 promoters are underlined.

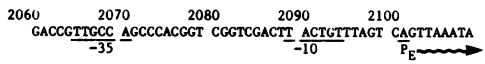

Figure 4. Nucleotide sequence of the P_E promoter located in the distal portion of the ilvM gene. The conserved -10 and -35 hexamer regions (17) of this promoter are underlined.

The nucleotide sequence of a secondary promoter, P_E, located in the distal portion of the ilvM gene is presented in figure 4. This promoter has been characterized in vivo and in vitro in both Escherichia coli and Salmonella typhimurium and, based on galK transcriptional fusions, the P_E promoter is about 50-fold weaker than the operon-proximal promoter (19-21). The P_E promoter is thought to be a weak promoter which functions to maintain a basal level of operon-distal gene expression within the ilvGMEDA operon during certain growth conditions (20,21).

In addition to the P_E promoter, there are reports of other internal promoters, P_D and P_A located upstream of the ilvD and ilvA genes of E. coli, respectively (19,22-24). Using galK transcriptional fusion plasmids and in vitro transcription techniques, we find no evidence for a P_D promoter although a very weak P_A promoter activity has been detected (unpublished observations).

B. Multivalent attenuation

The ilvGMEDA operon is regulated by a multivalent repression mechanism (25). That is, in the presence of all three of the branched chain amino acids, isoleucine, leucine and valine, the expression of these enzymes is inhibited, but when the growth rate of the cell is limited by the availability of any one of these amino acids the rate of synthesis of these enzymes increases. During the last several years the molecular mechanism responsible for this multivalent regulation has been elucidated (14-16). It is now known that the ilvGMEDA operon is regulated by attenuation. Attenuation may be generally defined as the regulation of transcription termination at a site (the attenuator) preceding the first structural gene of an operon. The region of the ilvGMEDA operon between the site of transcription initiation and the attenuator (the leader region) is very similar to the leader region of other amino acid biosynthetic operons regulated by attenuation (for review see ref. 26). The RNA transcripts of each of these leader regions contain the same common features: 1) a coding region for a short polypeptide containing codons for the regulating end-product amino acid(s), 2) an early stem-loop structure (stem 1:2) that functions as an RNA polymerase pause site in the leader polypeptide coding region to ensure coupling of translation and transcription for proper attenuator regulation (27), 3) a stem-loop structure at the 3' end of the leader RNA transcript (stem 3:4) followed by several uridine residues which defines a rho-independent transcription termination site, the attenuator, and 4) a third stem-loop structure, (stem 2:3) the antiterminator, which forms when stem 1 is masked by a ribosome stalled on a regulatory codon. The formation of the antiterminator precludes the subsequent formation of the attenuator stem-loop which allows the RNA polymerase to proceed into the structural genes of the operon. Regulation of the ilvGMEDA attenuator under conditions of excess or limiting branched chain amino acid availability is shown schematically in figure 5.

The 186 nucleotide leader RNA of the ilvGMEDA operon encodes a 32 residue

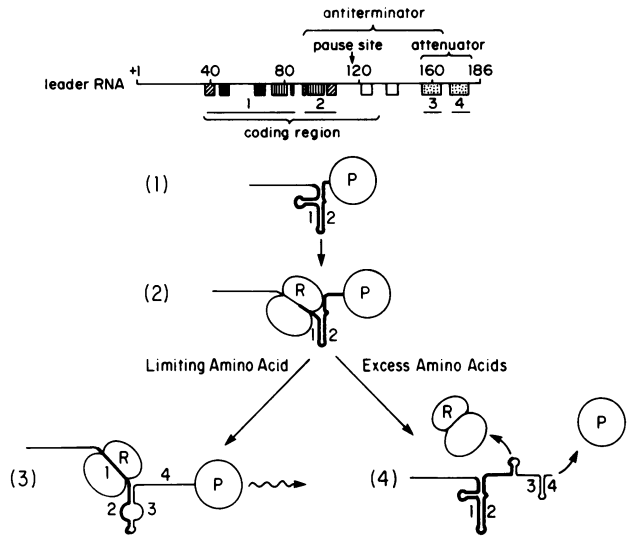

Figure 5. Attenuator control of the ilvGMEDA operon. Areas of pattern identity along the leader RNA represent regions of complementary sequences. The bold portion of the leader transcript depicts the leader polypeptide coding region. P represents an RNA polymerase and R represents a ribosome. The following events of attenuator regulation are illustrated. 1) RNA polymerase initiates transcription of leader RNA, stem-loop 1:2 forms and RNA polymerase pauses. 2) A translating ribosome enters stem 1, disrupts stem-loop 1:2 and releases the paused RNA polymerase. 3) During growth conditions limiting for leucine, valine or isoleucine, the translating ribosome stalls on a regulatory codon in stem 1:2, thereby facilitating the formation of the antiterminator, stem-loop 2:3. As a result, the terminator stem-loop 3:4 cannot form, the termination signal is ignored and RNA polymerase continues transcription into the structural genes of the operon (deattenuation). 4) During growth conditions not limiting for leucine, valine or isoleucine, the translating ribosome and pause released RNA polymerase remain coupled. The ribosome clears region 2 and releases at the end of the leader polypeptide coding region before stem 4 is synthesized. Stem-loop 1:2 reforms, stem 4 is synthesized and the terminator, stem-loop 3:4 forms resulting in transcription termination at the attenuator (attenuation).

polypeptide containing five isoleucine, five leucine and six valine residues (Fig. 6). Based on secondary structure predictions of the nucleotide sequence of the leader RNA, it is proposed that the leader RNA can form alternative structures which control the level of transcription into the ilvGMEDA structural genes (14,15) (Fig. 6B). In order to accommodate multivalent attenuation by three amino acids, tandem codons for isoleucine, valine and leucine must be located in the stem 1 region of the leader transcript. It is possible that the requirement for this much coding sequence in the stem 1 region necessitates the rather complicated bifurcated 1:2 stem-loop structure shown in figure 6B.

The regulatory codons for leucine are at positions four and five in the polypeptide coding region of the leader RNA (Fig. 6). Assuming that a ribosome covers

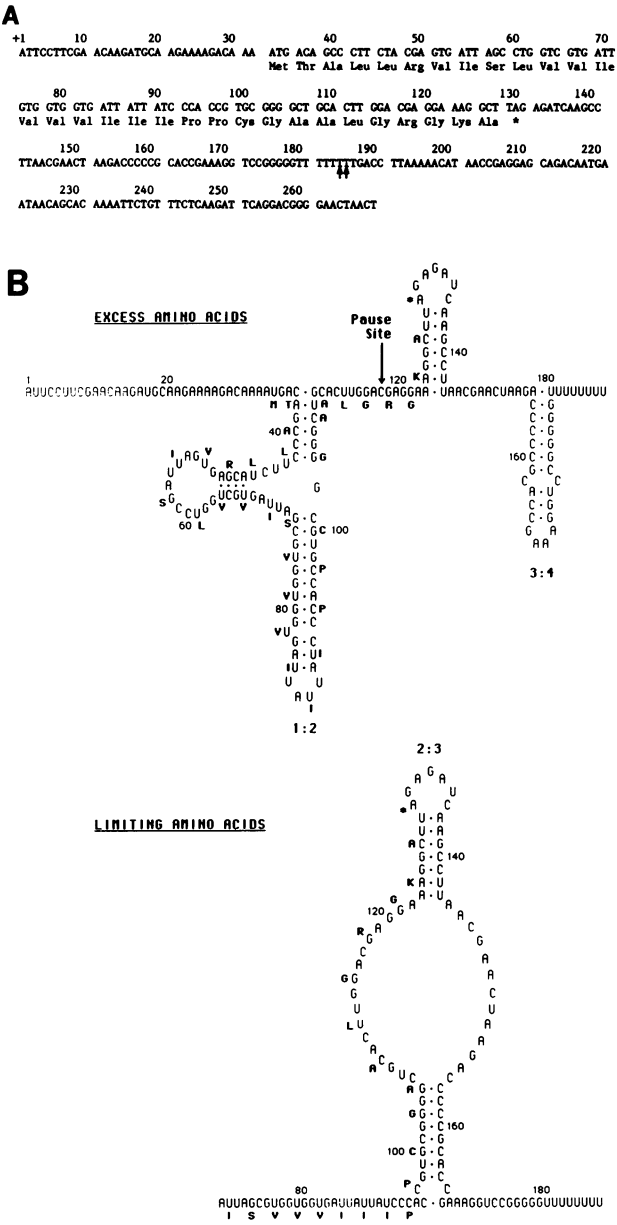

Figure 6. The leader region of the ilvGMEDA operon. (A) Nucleotide sequence of the nontranscribed DNA strand of the leader region and the deduced sequence of the leader polypeptide. The arrows indicate the nucleotides at which the leader transcript terminates. (B) The alternative RNA secondary structures that are believed to form during excess and limiting branched chain amino acids. The sequence of the encoded leader polypeptide is shown using the one letter amino acid symbols.

approximately 9 nucleotides on either side of a translated codon (28,29), a ribosome stalled on one of these leucine codons would cover a region of the leader transcript from approximately nucleotides 33 to 56. The position of this stalled ribosome precludes this covered nucleotide sequence from base pairing with its complementary sequence in stem 2, thus facilitating pairing of the stem 2 sequence with complementary sequences in stem 3 to form the antiterminator (Fig. 6B).

The role of the valine, isoleucine and leucine codons in the bifurcated stem-loop (nucleotides 43-72, Fig. 6) is not clear. Regulatory codons in attenuators are usually observed in tandem. The only exception is the single leucine regulatory codon in the leader transcript of the ilvGMEDA operon of Serratia marcescens (30). This suggests that the single valine, isoleucine and leucine codons specifying amino acids at positions 7,8 and 10 in the leader polypeptide, respectively, might be unimportant for regulation of attenuation. Consistent with this point, these codon positions within the bifurcated stem-loop of stem 1 are located 9 nucleotides or more away from RNA sequences in stem 1 which are involved in base pairing with stem 2. The tandem valine codons at positions 11 and 12, however, are close enough to disrupt the 1:2 stem-loop base pairing if a ribosome stalled on them. A ribosome stalled at one of the three tandem valine codons specifying amino acids at positions 15-17 of the leader polypeptide should also facilitate deattenuation. However, if it is accepted that a ribosome covers only 21 nucleotides of the RNA transcript, the second and third tandem isoleucine codons (specifying amino acids at positions 19-20 of the leader polypeptide) at the bottom of the 1:2 stem-loop are not situated to effect deattenuation (Fig. 6). In order to account for the fact that the ilvGMEDA operon is deattenuated by isoleucine limitation, it has been postulated that a second ribosome could queue behind a ribosome stalled at either of these isoleucine codons at the bottom of stem-loop 1:2 (14). This second ribosome would preclude 1:2 stem-loop formation and thus facilitate deattenuation. However, the recent documentation of a transcriptional pause site at nucleotide 117 near the base of stem 1:2 casts doubt on the validity of this model (31). If the first ribosome entering stem 1 releases an RNA polymerase molecule paused at nucleotide 117 then it is not likely that there would be sufficient time to load and position a second ribosome before the released RNA polymerase completed the remainder of the leader transcript resulting in termination of transcription (27). A more plausible explanation is that a ribosome stalled at an isoleucine codon at the bottom of the 1:2 stem-loop has a longer range effect on destabilization of RNA base pairing than the 21 or so nucleotides that are directly masked by the ribosome. The denaturing effect of the physical interaction of the ribosome with the 21 or so bases around the Ile codon would disrupt the bottom half of stem-loop 1:2 and spatially separate the remaining complementary base pairing regions of the stems 1 and 2. Therefore, a ribosome at this position would favor the base pairing of the unmasked downstream sequences leading to the formation of the antiterminator stem-loop 2:3.

C. Termination of transcription at the end of the ilvGMEDA operon

The nucleotide sequence following the ilvA gene is presented in figure 7. This sequence contains a possible rho-independent terminator (32) centered at nucleotide 6686. There is a G+C rich inverted repeat of seven nucleotides followed by 5 T residues in the ilvA nontranscribed DNA strand. In vitro transcription experiments using a

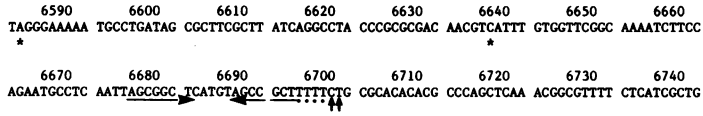

Figure 7. Transcriptional termination at the end of the ilvGMEDA operon. Nucleotide sequence of the nontranscribed DNA strand of the region distal to the ilvA gene. The sequence begins with the ilvA translational stop codon. The ilvY gene is transcribed in a direction opposite to the ilvGMEDA operon. The stop codon of the ilvY product is centered at nucleotide 6638 (designated by asterisk) on the DNA strand complementary to the sequence presented. Centered at nucleotide 6686 is an inverted repeat followed by 5 T residues. The arrows indicate the site of the rho-independent transcription termination at the end of the ilvGMEDA operon.

A

ilvG

```
        280         290         300         310         320         330         340         350         360
ATG AAT GGC GCA CAG TGG GTG GTA CAT GCC TTG CGG GCA CAG GGT GTG AAC ACC GTT TTC GGT TAT CCG GGT GGC GCA ATT ATG CCG GTT
Met Asn Gly Ala Gln Trp Val Val His Ala Leu Arg Ala Gln Gly Val Asn Thr Val Phe Gly Tyr Pro Gly Gly Ala Ile Met Pro Val
  1                          10                          20                          30
        370         380         390         400         410         420         430         440         450
TAC GAT GCA TTG TAT GAC GGC GGC GTG GAG CAC TTG CTA TGC CGA CAT GAG CAG GGT GCG GCA ATG GCG GCT ATC GGT TAT GCT CGT GCT
Tyr Asp Ala Leu Tyr Asp Gly Gly Val Glu His Leu Leu Cys Arg His Glu Gln Gly Ala Ala Met Ala Ala Ile Gly Tyr Ala Arg Ala
                           40                          50                          60
        460         470         480         490         500         510         520         530         540
ACC GGC AAA ACT GGC GTA TGT ATC GCC ACG TCT GGT CCG GGC GCA ACC AAC CTG ATA ACC GGG CTT GCG GAC GCA CTG TTA GAT TCC ATC
Thr Gly Lys Thr Gly Val Cys Ile Ala Thr Ser Gly Pro Gly Ala Thr Asn Leu Ile Thr Gly Leu Ala Asp Ala Leu Leu Asp Ser Ile
                           70                          80                          90
        550         560         570         580         590         600         610         620         630
CCT GTT GTT GCC ATC ACC GGT CAA GTG TCC GCA CCG TTT ATC GGC ACT GAC GCA TTT CAG GAA GTG GAT GTC CTG GGA TTG TCG TTA GCC
Pro Val Val Ala Ile Thr Gly Gln Val Ser Ala Pro Phe Ile Gly Thr Asp Ala Phe Gln Glu Val Asp Val Leu Gly Leu Ser Leu Ala
                          100                         110                         120
        640         650         660         670         680         690         700         710         720
TGT ACC AAG CAC AGC TTT CTG GTG CAG TCG CTG GAA GAG TTG CCG CGC ATC ATG GCT GAA GCA TTC GAC GTT GCC TGC TCA GGT CGT CCT
Cys Thr Lys His Ser Phe Leu Val Gln Ser Leu Glu Glu Leu Pro Arg Ile Met Ala Glu Ala Phe Asp Val Ala Cys Ser Gly Arg Pro
                          130                         140                         150
        730         740         750         760         770         780         790         800         810
GGT CCG GTT CTG GTC GAT ATC CCA AAA GAT ATC CAG TTA GCC AGC GGT GAC CTG GAA CCG TGG TTC ACC ACC GTT GAA AAC GAA GTG ACT
Gly Pro Val Leu Val Asp Ile Pro Lys Asp Ile Gln Leu Ala Ser Gly Asp Leu Glu Pro Trp Phe Thr Thr Val Glu Asn Glu Val Thr
                          160                         170                         180
        820         830         840         850         860         870         880         890         900
TTC CCA CAT GCC GAA GTT GAG CAA GCG CGC CAG ATG CTG GCA AAA GCG CAA AAA CCG ATG CTG TAC GTT GGC GGT GGC GTG GGT ATG GCG
Phe Pro His Ala Glu Val Glu Gln Ala Arg Gln Met Leu Ala Lys Ala Gln Lys Pro Met Leu Tyr Val Gly Gly Gly Val Gly Met Ala
                          190                         200                         210
        910         920         930         940         950         960         970         980         990
CAG GCA GTT CCG GCT TTG CGT GAA TTT CTC GCT GCC ACA AAA ATG CCT GCC ACC TGT ACG CTG AAA GGG CTG GGC GCA GTA GAA GCA GAT
Gln Ala Val Pro Ala Leu Arg Glu Phe Leu Ala Ala Thr Lys Met Pro Ala Thr Cys Thr Leu Lys Gly Leu Gly Ala Val Glu Ala Asp
                          220                         230                         240
       1000        1010        1020        1030        1040        1050        1060        1070        1080
TAT CCG TAC TAT CTG GGC ATG CTG GGG ATG CAC GGC ACC AAA GCG GCA AAC TTC GCG GTG CAG GAG TGT GAC CTG CTG ATC GCC GTG GGC
Tyr Pro Tyr Tyr Leu Gly Met Leu Gly Met His Gly Thr Lys Ala Ala Asn Phe Ala Val Gln Glu Cys Asp Leu Leu Ile Ala Val Gly
                          250                         260                         270
       1090        1100        1110        1120        1130        1140        1150        1160        1170
GCA CGT TTT GAT GAC CGG GTG ACC GGC AAA CTG AAC ACC TCC GCG CCA CAC GCC AGT GTT ATC CAT ATG GAT ATC GAC CCG GCA GAA ATG
Ala Arg Phe Asp Asp Arg Val Thr Gly Lys Leu Asn Thr Ser Ala Pro His Ala Ser Val Ile His Met Asp Ile Asp Pro Ala Glu Met
                          280                         290                         300
       1180        1190        1200        1210        1220        1230        1240        1250        1260
AAC AAG CTG CGT CAG GCA CAT GTG GCA TTA CAA GGT GAT TTA AAT GCT CTG TTA CCA GCA TTA CAG CAG CCG TTA AAT CAA TAT GAC TGG
Asn Lys Leu Arg Gln Ala His Val Ala Leu Gln Gly Asp Leu Asn Ala Leu Leu Pro Ala Leu Gln Gln Pro Leu Asn Gln Tyr Asp Trp
                          310                         320                         330
       1270        1280        1290        1300        1310        1320        1330        1340        1350
CAG CAA CAC TGC GCG CAG CTG CGT GAT GAA CAT TCC TGG CGT TAC GAC CAT CCC GGT GAC GCT ATC TAC GCG CCG TTG TTG TTA AAA CAA
Gln Gln His Cys Ala Gln Leu Arg Asp Glu His Ser Trp Arg Tyr Asp His Pro Gly Asp Ala Ile Tyr Ala Pro Leu Leu Lys Gln
                          340                         350                         360
       1360        1370        1380        1390        1400        1410        1420        1430        1440
CTG TCG GAT CGT AAA CCT GCG GAT TGC GTC GTG ACC ACA GAT GTG GGG CAG CAC CAG ATG TGG GCT GCG CAG CAC ATC GCC CAC ACT CGC
Leu Ser Asp Arg Lys Pro Ala Asp Cys Val Val Thr Thr Asp Val Gly Gln His Gln Met Trp Ala Ala Gln His Ile Ala His Thr Arg
                          370                         380                         390
       1450        1460        1470        1480        1490        1500        1510        1520        1530
CCG GAA AAT TTC ATC ACC TCC AGC GGT TTA GGT ACC ATG GGT TTT GGT TTA CCG GCG GCG GTT GGC GCA CAA GTC GCG CGA CCG AAC GAT
Pro Glu Asn Phe Ile Thr Ser Ser Gly Leu Gly Thr Met Gly Phe Gly Leu Pro Ala Ala Val Gly Ala Gln Val Ala Arg Pro Asn Asp
                          400                         410                         420
       1540        1550        1560        1570        1580        1590        1600        1610        1620
ACC GTT GTC TGT ATC TCC GGT GAC GGC TCT TTC ATG ATG AAT GTG CAA GAG CTG GGC ACC GTA AAA CGC AAG CAG TTA CCG TTG AAA ATC
Thr Val Val Cys Ile Ser Gly Asp Gly Ser Phe Met Met Asn Val Gln Glu Leu Gly Thr Val Lys Arg Lys Gln Leu Pro Leu Lys Ile
                          430                         440                         450
       1630        1640        1650        1660        1670        1680        1690        1700        1710
GTC TTA CTC GAT AAC CAA CGG TTA GGG ATG GTT CGA CAA TGG CAG CAA CTG TTT TTT CAG GAA CGA TAC AGC GAA ACC ACC CTT ACT GAT
Val Leu Leu Asp Asn Gln Arg Leu Gly Met Val Arg Gln Trp Gln Gln Leu Phe Phe Gln Glu Arg Tyr Ser Glu Thr Thr Leu Thr Asp
                          460                         470                         480
```

```
        1720        1730        1740        1750        1760        1770        1780        1790        1800
AAC CCC GAT TTC CTC ATG TTA GCC AGC GCC TTC GGC ATC CAT GGC CAA CAC ATC ACC CGG AAA GAC CAG GTT GAA GCG GCA CTC GAC ACC
Asn Pro Asp Phe Leu Met Leu Ala Ser Ala Phe Gly Ile His Gly Gln His Ile Thr Arg Lys Asp Gln Val Glu Ala Ala Leu Asp Thr
                                490                                 500                                 510
        1810        1820        1830        1840        1850        1860        1870        1880        1890
ATG CTG AAC AGT GAT GGG CCA TAC CTG CTT CAT GTC TCA ATC GAC GAA CTT GAG AAC GTC TGG CCG CTG GTG CCG CCT GGC GCC AGT AAT
Met Leu Asn Ser Asp Gly Pro Tyr Leu Leu His Val Ser Ile Asp Glu Leu Glu Asn Val Trp Pro Leu Val Pro Pro Gly Ala Ser Asn
                                520                                 530                                 540
        1900        1910
TCA GAA ATG TTG GAG AAA TTA TCA TGA
Ser Glu Met Leu Glu Lys Leu Ser  *
```

B

```
ilvM
        1920        1930        1940        1950        1960        1970        1980        1990        2000
ATG ATG CAA CAT CAG GTC AAT GTA TCG GCT CGC TTC AAT CCA GAA ACC TTA GAA CGT GTT TTA CGC GTG GTG CGT CAT CGT GGT TTC CAC
Met Met Gln His Gln Val Asn Val Ser Ala Arg Phe Asn Pro Glu Thr Leu Glu Arg Val Leu Arg Val Val Arg His Arg Gly Phe His
  1                                 10                                  20                                  30
    2010        2020        2030        2040        2050        2060        2070        2080        2090
GTC TGC TCA ATG AAT ATG GCC GCC GCC AGC GAT GCA CAA AAT ATA AAT ATC GAA TTG ACC GTT GCC AGC CCA CGG TCG GTC GAC TTA CTG
Val Cys Ser Met Asn Met Ala Ala Ala Ser Asp Ala Gln Asn Ile Asn Ile Glu Leu Thr Val Ala Ser Pro Arg Ser Val Asp Leu Leu
                                40                                  50                                  60
    2100        2110        2120        2130        2140        2150        2160        2170
TTT AGT CAG TTA AAT AAA CTG GTG GAC GTC GCA CAC GTT GCC ATC TGC CAG AGC ACA ACC ACA TCA CAA CAA ATC CGC GCC TGA
Phe Ser Gln Leu Asn Lys Leu Val Asp Val Ala His Val Ala Ile Cys Gln Ser Thr Thr Thr Ser Gln Gln Ile Arg Ala  *
                                70                                  80
```

C

```
      2180      2190
SPACER  GCGCAAAAGG  AATATAAAA
ilvE
 2200        2210        2220        2230        2240        2250        2260        2270        2280
ATG ACC ACG AAG AAA GCT GAT TAC ATT TGG TTC AAT GGG GAG ATG GTT CGC TGG GAA GAC GCG AAG GTG CAT GTG ATG TCG CAC GCG CTG
Met Thr Thr Lys Lys Ala Asp Tyr Ile Trp Phe Asn Gly Glu Met Val Arg Trp Glu Asp Ala Lys Val His Val Met Ser His Ala Leu
  1                                 10                                  20                                  30
 2290        2300        2310        2320        2330        2340        2350        2360        2370
CAC TAT GGC ACT TCG GTT TTT GAA GGC ATC CGT TGC TAC GAC TCG CAC AAA GGA CCG GTT GTA TTC CGC CAT CGT GAG CAT ATG CAG CGT
His Tyr Gly Thr Ser Val Phe Glu Gly Ile Arg Cys Tyr Asp Ser His Lys Gly Pro Val Val Phe Arg His Arg Glu His Met Gln Arg
                                40                                  50                                  60
 2380        2390        2400        2410        2420        2430        2440        2450        2460
CTG CAT GAC TCC GCC AAA ATC TAT CGC TTC CCG GTT TCG CAG AGC ATT GAT GAG CTG ATG GAA GCT TGT CGT GAC GTG ATC CGC AAA AAC
Leu His Asp Ser Ala Lys Ile Tyr Arg Phe Pro Val Ser Gln Ser Ile Asp Glu Leu Met Glu Ala Cys Arg Asp Val Ile Arg Lys Asn
                                70                                  80                                  90
 2470        2480        2490        2500        2510        2520        2530        2540        2550
AAT CTC ACC AGC GCC TAT ATC CGT CCG CTG ATC TTC GTC GGT GAT GTT GGC ATG GGA GTA AAC CCG CCA GCG GGA TAC TCA ACC GAC GTG
Asn Leu Thr Ser Ala Tyr Ile Arg Pro Leu Ile Phe Val Gly Asp Val Gly Met Gly Val Asn Pro Pro Ala Gly Tyr Ser Thr Asp Val
                                100                                 110                                 120
 2560        2570        2580        2590        2600        2610        2620        2630        2640
ATT ATC GCT GCT TTC CCG TGG GGA GCG TAT CTG GGC GCA GAA GCG CTG GAG CAG GGG ATC GAT GCG ATG GTT TCC TCC TGG AAC CGC GCA
Ile Ile Ala Ala Phe Pro Trp Gly Ala Tyr Leu Gly Ala Glu Ala Leu Glu Gln Gly Ile Asp Ala Met Val Ser Ser Trp Asn Arg Ala
                                130                                 140                                 150
 2650        2660        2670        2680        2690        2700        2710        2720        2730
GCA CCA AAC ACC ATC CCG ACG GCG GCA AAA GCC GGT GGT AAC TAC CTC TCT TCC CTG CTG GTG GGT AGC GAA GCG CGC CGC CAC GGT TAT
Ala Pro Asn Thr Ile Pro Thr Ala Ala Lys Ala Gly Gly Asn Tyr Leu Ser Ser Leu Leu Val Gly Ser Glu Ala Arg Arg His Gly Tyr
                                160                                 170                                 180
 2740        2750        2760        2770        2780        2790        2800        2810        2810
CAG GAA GGT ATC GCG CTG GAT GTG AAC GGT TAT ATC TCT GAA GGC GCA GGC GAA AAC CTG TTT GAA GTG AAA GAT GGT GTG CTG TTC ACC
Gln Glu Gly Ile Ala Leu Asp Val Asn Gly Tyr Ile Ser Glu Gly Ala Gly Glu Asn Leu Phe Glu Val Lys Asp Gly Val Leu Phe Thr
                                190                                 200                                 210
 2830        2840        2850        2860        2870        2880        2890        2900        2910
CCA CCG TTC ACC TCC TCC GCG CTG CCG GGT ATT ACC CGT GAT GCC ATC ATC AAA CTG GCG AAA GAG CTG GGA ATT GAA GTA CGT GAG CAG
Pro Pro Phe Thr Ser Ser Ala Leu Pro Gly Ile Thr Arg Asp Ala Ile Ile Lys Leu Ala Lys Glu Leu Gly Ile Glu Val Arg Glu Gln
                                220                                 230                                 240
 2920        2930        2940        2950        2960        2970        2980        2990        3000
GTG CTG TCG CGC GAA TCC CTG TAC CTG GCG GAT GAA GTG TTT ATG TCC GGT ACG GCG GCA GAA ATC ACG CCA GTG CGC AGC GTA GAC GGT
Val Leu Ser Arg Glu Ser Leu Tyr Leu Ala Asp Glu Val Phe Met Ser Gly Thr Ala Ala Glu Ile Thr Pro Val Arg Ser Val Asp Gly
                                250                                 260                                 270
 3010        3020        3030        3040        3050        3060        3070        3080        3090
ATT CAG GTT GGC GAA GGC CGT TGT GGC CCG GTT ACC AAA CGC ATT CAG CAA GCC TTC TTC GGC CTC TTC ACT GGC GAA ACC GAA GAT AAA
Ile Gln Val Gly Glu Gly Arg Cys Gly Pro Val Thr Lys Arg Ile Gln Gln Ala Phe Phe Gly Leu Phe Thr Gly Glu Thr Glu Asp Lys
                                280                                 290                                 300
 3100        3110        3120
TGG GGC TGG TTA GAT CAA GTT AAT CAA TAA
Trp Gly Trp Leu Asp Gln Val Asn Gln  *
                                310
```

D

```
        3130        3140        3150        3160        3170        3180        3190
SPACER  ATACAAAAAAAT  GGGACGGCAC  GCACGGTCCC  ATTTACGAGA  CAGACACTGG  GAGTAAATAA  AGT

ilvD
        3200        3210        3220        3230        3240        3250        3260        3270        3280
ATG CCT AAG TAC CGT TCC GCC ACC ACC ACT CAT GGT CGT AAT ATG GCG GGT GCT CGT GCG CTG TGG CGC GCC ACC GGA ATG ACC GAC GCC
Met Pro Lys Tyr Arg Ser Ala Thr Thr Thr His Gly Arg Asn Met Ala Gly Ala Arg Ala Leu Trp Arg Ala Thr Gly Met Thr Asp Ala
  1                                 10                                  20                                  30
        3290        3300        3310        3320        3330        3340        3350        3360        3370
GAT TTC GGT AAG CCG ATT ATC GCG GTT GTG AAC TCG TTC ACC CAA TTT GTA CCG GGT CAC GTC CAT CTG CGC GAT CTC GGT AAA CTG GTC
Asp Phe Gly Lys Pro Ile Ile Ala Val Val Asn Ser Phe Thr Gln Phe Val Pro Gly His Val His Leu Arg Asp Leu Gly Lys Leu Val
                                40                                  50                                  60
        3380        3390        3400        3410        3420        3430        3440        3450        3460
GCC GAA CAA ATT GAA GCG GCT GGC GGC GTT GCC AAA GAG TTC AAC ACC ATT GCG GTG GAT GAT GGG ATT GCC ATG GGC CAC GGG GGG ATG
Ala Glu Gln Ile Glu Ala Ala Gly Gly Val Ala Lys Glu Phe Asn Thr Ile Ala Val Asp Asp Gly Ile Ala Met Gly His Gly Gly Met
                                70                                  80                                  90
```

```
        3470        3480        3490        3500        3510        3520        3530        3540        3550
CTT TAT TCA CTG CCA TCT CGC GAA CTG ATC GCT GAT TCC GTT GAG TAT ATG GTC AAC GCC CAC TGC GCC GAC GCC ATG GTC TGC ATC TCT
Leu Tyr Ser Leu Pro Ser Arg Glu Leu Ile Ala Asp Ser Val Glu Tyr Met Val Asn Ala His Cys Ala Asp Ala Met Val Cys Ile Ser
                                100                                 110                                 120
        3560        3570        3580        3590        3600        3610        3620        3630        3640
AAC TGC GAC AAA ATC ACC CCG GGG ATG CTG ATG GCT TCC CTG CGC CTG AAT ATT CCG GTG ATC TTT GTT TCC GGC GGC CCG ATG GAG GCC
Asn Cys Asp Lys Ile Thr Pro Gly Met Leu Met Ala Ser Leu Arg Leu Asn Ile Pro Val Ile Phe Val Ser Gly Gly Pro Met Glu Ala
                                130                                 140                                 150
        3650        3660        3670        3680        3690        3700        3710        3720        3730
GGG AAA ACC AAA CTT TCC GAT CAG ATC ATC AAG CTC GAT CCG GTT GAT GCG ATG ATC CAG GGC GCA GAC CCG AAA GTA TCT GAC TCC CAG
Gly Lys Thr Lys Leu Ser Asp Gln Ile Ile Lys Leu Asp Pro Val Asp Ala Met Ile Gln Gly Ala Asp Pro Lys Val Ser Asp Ser Gln
                                160                                 170                                 180
        3740        3750        3760        3770        3780        3790        3800        3810        3820
AGC GAT CAG GTT GAA CGT TCC GCG TGT CCG ACC TGC GGT TCC TGC TCC GGG ATG TTT ACC GCT AAC TCA ATG AAC TGC CTG ACC GAA GCG
Ser Asp Gln Val Glu Arg Ser Ala Cys Pro Thr Cys Gly Ser Cys Ser Gly Met Phe Thr Ala Asn Ser Met Asn Cys Leu Thr Glu Ala
                                190                                 200                                 210
        3830        3840        3850        3860        3870        3880        3890        3900        3910
CTG GGC CTG TCG CAG CCG GGC AAC GGC TCG CTG CTG GCA ACC CAC GCC GAC CGT AAG CAC CTG TTC CTT AAT GCT GGT AAA CGC ATT GTT
Leu Gly Leu Ser Gln Pro Gly Asn Gly Ser Leu Leu Ala Thr His Ala Asp Arg Lys Gln Leu Phe Leu Asn Ala Gly Lys Arg Ile Val
                                220                                 230                                 240
        3920        3930        3940        3950        3960        3970        3980        3990        4000
GAA TTG ACC AAA CGT TAT TAC GAG CAA AAC GAC GAA AGT GCA CTG CCG CGT AAT ATC GCC AGT AAG GCG GCG TTT GAA AAC GCC ATG ACG
Glu Leu Thr Lys Arg Tyr Tyr Glu Gln Asn Asp Glu Ser Ala Leu Pro Arg Asn Ile Ala Ser Lys Ala Ala Phe Glu Asn Ala Met Thr
                                250                                 260                                 270
        4010        4020        4030        4040        4050        4060        4070        4080        4090
CTG GAT ATC GCG ATG GGT GGA TCG ACT AAC ACC GTA CTT CAC CTG CTG GCG GCG GCG CAG GAA GCG GAA ATC GAC TTC ACC ATG AGT GAT
Leu Asp Ile Ala Met Gly Gly Ser Thr Asn Thr Val Leu His Leu Leu Ala Ala Ala Gln Glu Ala Glu Ile Asp Phe Thr Met Ser Asp
                                280                                 290                                 300
        4100        4110        4120        4130        4140        4150        4160        4170        4180
ATC GAT AAG CTT TCC CGC AAG GTT CCA CAG CTG TGT AAA GTT GCG CCG AGC ACC CAG AAA TAC CAT ATG GAA GAT GTT CAC CGT GCT GGT
Ile Asp Lys Leu Ser Arg Lys Val Pro Gln Leu Cys Lys Val Ala Pro Ser Thr Gln Lys Tyr His Met Glu Asp Val His Arg Ala Gly
                                310                                 320                                 330
        4190        4200        4210        4220        4230        4240        4250        4260        4270
GGT GTT ATC GGT ATT CTC GGC GAA CTG GAT CGC GCG GGG TTA CTG AAC CGT GAT GTG AAA AAC GTA CTT GGC CTG ACG TTG CCG CAA ACG
Gly Val Ile Gly Ile Leu Gly Glu Leu Asp Arg Ala Gly Leu Leu Asn Arg Asp Val Lys Asn Val Leu Gly Leu Thr Leu Pro Gln Thr
                                340                                 350                                 360
        4280        4290        4300        4310        4320        4330        4340        4350        4360
CTG GAA CAA TAC GAC GTT ATG CTG ACC CAG GAT GAC GCG GTA AAA AAT ATG TTC CGC GCA GGT CCT GCA GGC ATT CGT ACC ACA CAG GCA
Leu Glu Gln Tyr Asp Val Met Leu Thr Gln Asp Asp Ala Val Lys Asn Met Phe Arg Ala Gly Pro Ala Gly Ile Arg Thr Thr Gln Ala
                                370                                 380                                 390
        4370        4380        4390        4400        4410        4420        4430        4440        4450
TTC TCG CAA GAT TGC CGT TGG GAT ACG CTG GAC GAC GAT CGC GCC AAT GGC GTA TCC GCT CGC TGG AAC ACG CCT ACA GCA AAG ACG GCG
Phe Ser Gln Asp Cys Arg Trp Asp Thr Leu Asp Asp Asp Arg Ala Asn Gly Val Ser Ala Arg Trp Asn Thr Pro Thr Ala Lys Thr Ala
                                400                                 410                                 420
        4460        4470        4480        4490        4500        4510        4520        4530        4540
GCC TGG CGG TGC TCT ACG GTA ATC TTG CGG AAA CGG CTG CAT CGT GAA ACG GCA GGC GTC GAT GAC AGC ATC CTC AAA TTC ACC GGC CCG
Ala Trp Arg Cys Ser Thr Val Ile Leu Arg Lys Arg Leu His Arg Glu Thr Ala Gly Val Asp Asp Ser Ile Leu Lys Phe Thr Gly Pro
                                430                                 440                                 450
        4550        4560        4570        4580        4590        4600        4610        4620        4630
GCG AAA GTG TAC GAA AGC CAG GAC GAT GCG GTA GAA GCG ATT CTC GGC GGT AAA GTT GTC GCC GGA GAT GTG GTA GTA ATT CGC TAT GAA
Ala Lys Val Tyr Glu Ser Gln Asp Asp Ala Val Glu Ala Ile Leu Gly Gly Lys Val Val Ala Gly Asp Val Val Val Ile Arg Tyr Glu
                                460                                 470                                 480
        4640        4650        4660        4670        4680        4690        4700        4710        4720
GGC CCG AAA GGC GGT CCG GGG ATG CAG GAA ATG CTC TAC CCA ACC AGC TTC CTG AAA TCA ATG GGT CTC GGC AAA GCC TGT GCG CTG ATC
Gly Pro Lys Gly Gly Pro Gly Met Gln Glu Met Leu Tyr Pro Thr Ser Phe Leu Lys Ser Met Gly Leu Gly Lys Ala Cys Ala Leu Ile
                                490                                 500                                 510
        4730        4740        4750        4760        4770        4780        4790        4800        4810
ACC GAC GGT CGT TTC TCT GGT GGC ACC TCT GGT CTT TCC ATC GGC CAC GTC TCA CCG GAA GCG GCA AGC GGC GGC AGC ATT GGC CTG ATT
Thr Asp Gly Arg Phe Ser Gly Gly Thr Ser Gly Leu Ser Ile Gly His Val Ser Pro Glu Ala Ala Ser Gly Gly Ser Ile Gly Leu Ile
                                520                                 530                                 540
        4820        4830        4840        4850        4860        4870        4880        4890        4900
GAA GAC GGT GAC CTG ATC GCT ATC GAC ATC CCG AAC CGT GGC ATT CAG TTA CAG GTA AGC GAT GCC GAA CTG GCG GCG CGT CGT GAA GCG
Glu Asp Gly Asp Leu Ile Ala Ile Asp Ile Pro Asn Arg Gly Ile Gln Leu Gln Val Ser Asp Ala Glu Leu Ala Ala Arg Arg Glu Ala
                                550                                 560                                 570
        4910        4920        4930        4940        4950        4960        4970        4980        4990
CAG GAC GCT CGA GGT GAC AAA GCC TGG ACG CCG AAA AAT CGT GAA CGT CAG GTC TCC TTT GCC CTG CGT GCT TAT GCC AGC CTG GCA ACC
Gln Asp Ala Arg Gly Asp Lys Ala Trp Thr Pro Lys Asn Arg Glu Arg Gln Val Ser Phe Ala Leu Arg Ala Tyr Ala Ser Leu Ala Thr
                                580                                 590                                 600
        5000        5010        5020        5030
AGC GCC GAC AAA GGC GCG GTG CGG CAT AAA TCG AAA CTG GGG GGT TAA
Ser Ala Asp Lys Gly Ala Val Arg His Lys Ser Lys Leu Gly Gly  *
                                610
```

E

```
ilvA
5040        5050        5060        5070        5080        5090        5100        5110        5120
  TA ATG GCT GAC TCG CAA CCC CTG TCC GGT GCT CCG GAA GGT GCC GAA TAT TTA AGA GCA GTG CTG CGC GCG CCG GTT TAC GAG GCG GCG
     Met Ala Asp Ser Gln Pro Leu Ser Gly Ala Pro Glu Gly Ala Glu Tyr Leu Arg Ala Val Leu Arg Ala Pro Val Tyr Glu Ala Ala
       1                                  10                                  20
5130        5140        5150        5160        5170        5180        5190        5200        5210
 CAG GTT ACG CCG CTA CAA AAA ATG GAA AAA CTG TCG TCG CGT CTT GAT AAC GTC ATT CTG GTG AAG CGC GAA GAT CGC CAG CCA GTG CAC
 Gln Val Thr Pro Leu Gln Lys Met Glu Lys Leu Ser Ser Arg Leu Asp Asn Val Ile Leu Val Lys Arg Glu Asp Arg Gln Pro Val His
  30                                  40                                  50
5220        5230        5240        5250        5260        5270        5280        5290        5300
 AGC TTT AAG CTG CGC GGC GCA TAC GCC ATG ATG GCG GGC CTG ACG GAA GAA CAG AAA GCG CAC GGC GTG ATC ACT GCT TCT GCG GGT AAC
 Ser Phe Lys Leu Arg Gly Ala Tyr Ala Met Met Ala Gly Leu Thr Glu Glu Gln Lys Ala His Gly Val Ile Thr Ala Ser Ala Gly Asn
  60                                  70                                  80
5310        5320        5330        5340        5350        5360        5370        5380        5390
 CAC GCG CAG GGC GTC GCG TTT TCT TCT GCG CGG TTA GGC GTG AAG GCC CTG ATC GTT ATG CCA ACC GCC ACC GCC GAC ATC AAA GTC GAC
 His Ala Gln Gly Val Ala Phe Ser Ser Ala Arg Leu Gly Val Lys Ala Leu Ile Val Met Pro Thr Ala Thr Ala Asp Ile Lys Val Asp
  90                                  100                                 110
5400        5410        5420        5430        5440        5450        5460        5470        5480
 GCG GTG CGC GGC TTC GGC GGC GAA GTG CTG CTC CAC GGC GCG AAC TTT GAT GAA GCG AAA GCC AAA GCG ATC GAA CTG TCA CAG CAG CAG
 Ala Val Arg Gly Phe Gly Gly Glu Val Leu Leu His Gly Ala Asn Phe Asp Glu Ala Lys Ala Lys Ala Ile Glu Leu Ser Gln Gln Gln
 120                                 130                                 140
```

```
5490          5500          5510          5520          5530          5540          5550          5560          5570
GGG TTC ACC TGG GTG CCG CCG TTC GAC CAT CCG ATG GTG ATT GCC GGG CAA GGC ACG CTG GCG CTG GAA CTG CTC CAG CAG GAC GCC CAT
Gly Phe Thr Trp Val Pro Pro Phe Asp His Pro Met Val Ile Ala Gly Gln Gly Thr Leu Ala Leu Glu Leu Leu Gln Gln Asp Ala His
150                                       160                                       170
5580          5590          5600          5610          5620          5630          5640          5650          5660
CTC GAC CGC GTA TTT GTG CCA GTC GGC GGC GGC GGT CTG GCT GCT GGC GTG GCG GTG CTG ATC AAA CAA CTG ATG CCG CAA ATC AAA GTG
Leu Asp Arg Val Phe Val Pro Val Gly Gly Gly Gly Leu Ala Ala Gly Val Ala Val Leu Ile Lys Gln Leu Met Pro Gln Ile Lys Val
180                                       190                                       200
5670          5680          5690          5700          5710          5720          5730          5740          5750
ATC GCC GTA GAA GCG GAA GAC TCC GCC TGC CTG AAA GCA GCG CTG GAT GCG GGT CAT CCG GTT GAT CTG CCG CGC GTA GGG CTA TTT GCT
Ile Ala Val Glu Ala Glu Asp Ser Ala Cys Leu Lys Ala Ala Leu Asp Ala Gly His Pro Val Asp Leu Pro Arg Val Gly Leu Phe Ala
210                                       220                                       230
5760          5770          5780          5790          5800          5810          5820          5830          5840
GAA GGC GTA GGC GTA AAA CGC ATC GGT GAC GAA ACC TTC CGT TTA TGC CAG GAG TAT CTC GAC GAC ATC ATC ACC GTC GAT AGC GAT GCG
Glu Gly Val Gly Val Lys Arg Ile Gly Asp Glu Thr Phe Arg Leu Cys Gln Glu Tyr Leu Asp Asp Ile Ile Thr Val Asp Ser Asp Ala
240                                       250                                       260
5850          5860          5870          5880          5890          5900          5910          5920          5930
ATC TGT GCG GCG ATG AAG GAT TTA TTC GAA GAT GTG CGC GCG GTG GCG GAA CCC TCT GGC GCG CTG GCG CTG GCG GGA ATG AAA AAA TAT
Ile Cys Ala Ala Met Lys Asp Leu Phe Glu Asp Val Arg Ala Val Ala Glu Pro Ser Gly Ala Leu Ala Leu Ala Gly Met Lys Lys Tyr
270                                       280                                       290
5940          5950          5960          5970          5980          5990          6000          6010          6020
ATC GCC CTG CAC AAC ATT CGC GGC GAA CGG CTG GCG CAT ATT CTT TCC GGT GCC AAC GTG AAC TTC CAC GGC CTG CGC TAC GTC TCA GAA
Ile Ala Leu His Asn Ile Arg Gly Glu Arg Leu Ala His Ile Leu Ser Gly Ala Asn Val Asn Phe His Gly Leu Arg Tyr Val Ser Glu
300                                       310                                       320
6030          6040          6050          6060          6070          6080          6090          6100          6110
CGC TGC GAA CTG GGC GAA CAG CGT GAA GCG TTG TTG GCG GTG ACC ATT CCG GAA GAA AAA GGC AGC TTC CTC AAA TTC TGC CAA CTG CTT
Arg Cys Glu Leu Gly Glu Gln Arg Glu Ala Leu Leu Ala Val Thr Ile Pro Glu Glu Lys Gly Ser Phe Leu Lys Phe Cys Gln Leu Leu
330                                       340                                       350
6120          6130          6140          6150          6160          6170          6180          6190          6200
GGC GGG CGT TCG GTC ACC GAG TTC AAC TAC CGT TTT GCC GAT GCC AAA AAC GCC TGC ATC TTT GTC GGT GTG CGC CTG AGC CGC GGC CTC
Gly Gly Arg Ser Val Thr Glu Phe Asn Tyr Arg Phe Ala Asp Ala Lys Asn Ala Cys Ile Phe Val Gly Val Arg Leu Ser Arg Gly Leu
360                                       370                                       380
6210          6220          6230          6240          6250          6260          6270          6280          6290
GAA GAG CGC AAA GAA ATT TTG CAG ATG CTC AAC GAC GGC GGC TAC AGC GTG GTT GAT CTC TCC GAC GAC GAA ATG GCG AAG CTA CAC GTG
Glu Glu Arg Lys Glu Ile Leu Gln Met Leu Asn Asp Gly Gly Tyr Ser Val Val Asp Leu Ser Asp Asp Glu Met Ala Lys Leu His Val
390                                       400                                       410
6300          6310          6320          6330          6340          6350          6360          6370          6380
CGC TAT ATG GTC GGC GGA CGT CCA TCG CAT CCG TTG CAG GAA CGC CTC TAC AGC TTC GAA TTC CCG GAA TCA CCG GGC GCG CTG CTG CGC
Arg Tyr Met Val Gly Gly Arg Pro Ser His Pro Leu Gln Glu Arg Leu Tyr Ser Phe Glu Phe Pro Glu Ser Pro Gly Ala Leu Leu Arg
420                                       430                                       440
6390          6400          6410          6420          6430          6440          6450          6460          6470
TTC CTC AAC ACG CTG GGT ACG TAC TGG AAC ATT TCT TTG TTC CAC TAT CGC AGC CAT GGC ACC GAC TAC GGG CGC GTA CTG GCG GCG TTC
Phe Leu Asn Thr Leu Gly Thr Tyr Trp Asn Ile Ser Leu Phe His Tyr Arg Ser His Gly Thr Asp Tyr Gly Arg Val Leu Ala Ala Phe
450                                       460                                       470
6480          6490          6500          6510          6520          6530          6540          6550          6560
GAA CTT GGC GAC CAT GAA CCG GAT TTC GAA ACC CGG CTG AAT GAG CTG GGC TAC GAT TGC CAC GAC GAA ACC AAT AAC CCG GCG TTC AGG
Glu Leu Gly Asp His Glu Pro Asp Phe Glu Thr Arg Leu Asn Glu Leu Gly Tyr Asp Cys His Asp Glu Thr Asn Asn Pro Ala Phe Arg
480                                       490                                       500
6570          6580
TTC TTT TTG GCG GGT TAG
Phe Phe Leu Ala Gly  *
510
```

Figure 8. The complete nucleotide sequence of the nontranscribed DNA strand of the structural genes of the ilvGMEDA operon. The nucleotide sequence is numbered relative to the in vivo transcriptional initiation site. The amino acid residues of each encoded polypeptide are numbered from the amino terminal Met. The nucleotide sequence of the intercistronic region preceding each gene is identified as spacer. The nucleotide sequence of each structural gene listed in this figure are presented as follows: A) ilvG; B) ilvM; C) ilvE; D) ilvD; E) ilvA. It is noted that two nucleotides, a C at 3257 and a G at 3260, have been inserted compared to a previously published sequence (33). These nucleotide insertions are consistent with the fact that a ApaI restriction site (GGGCCC), which was present in the previously published sequence, does not exist, while a BssHII restriction site (GCGCGC) is conserved at this site in both E. coli K-12 and S. typhimurium.

restriction fragment containing this sequence fused to the tac promoter indicate that this is a rho-independent termination site (unpublished observation). When termination factor rho (32) was added to the system, a rho-dependent termination site was detected about 100 bp downstream of the rho-independent termination site. Therefore, the arrangement of transcriptional terminators at the 3'-terminus of the ilvGMEDA operon

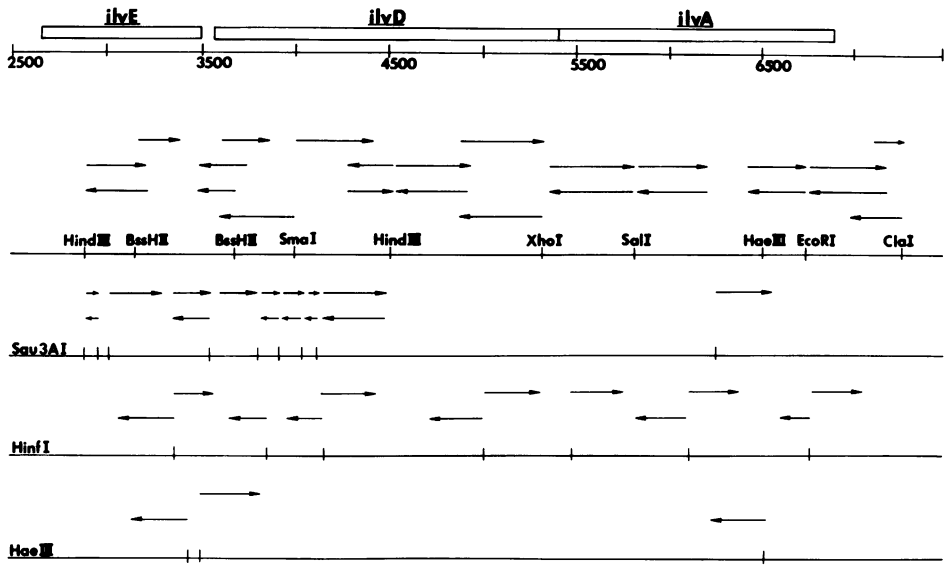

Figure 9. Nucleotide sequence strategy for the ilvE, ilvD and ilvA genes. The lengths of the arrows represent the extent of readable sequence determined by the Maxam and Gilbert chemical cleavage method (34) and dideoxy chain termination method of Sanger et al. (35). It is noted that the nucleotide sequence of the ilvE gene and the 5'-portion of the ilvD gene have been previously published (33). Nucleotide differences between the sequence determined in this strategy and that of previous work (33) are indicated in the figure 8 legend.

appears to be very similar to the well characterized trp tt' terminators of the trp operon of E. coli (32) (unpublished observations). It is interesting to note that both transcriptional termination sites are located in the ilvY coding region. This gene is transcribed in an opposite direction to the ilvA gene (9). The positions of the ilvA and ilvY stop codons are indicated by asterisks in figure 7.

III. STRUCTURAL GENES OF THE ilvGMEDA OPERON

A schematic diagram of the ilvGMEDA operon outlining the location of each structural gene is presented in figure 2. The complete nucleotide sequence of the non-transcribed DNA strand and the corresponding deduced amino acid sequence of the five structural genes are presented in figure 8. The nucleotide sequence of the ilvG and ilvM genes are from previous publications (5) and the sequence of the ilvE, ilvD and ilvA genes were determined according to the strategy shown in figure 9.

The following criteria were used to predict each polypeptide sequence. First, each structural gene had an open translational reading frame which encodes a product with a molecular weight consistent with the electrophoretic analysis of the gene products synthesized in vivo. Second, the initiation codons for the ilvG, ilvM and ilvE deduced

amino acid sequences were selected based on the amino terminal sequences determined from the purified protein products of these genes (36-38). Finally, the nucleotide sequence of the intercistronic regions of the structural genes of the ilvGMEDA operon in S. typhimurium were determined and compared to the homologous regions in E. coli (Fig. 10). Because these two enteric bacteria are so closely related, coding sequences are usually conserved while noncoding sequences, unless they perform a regulatory role, are diverged. Based on this coding conservation the probable initiation codons for the ilvD and ilvA genes of E. coli were determined and the deduced carboxy terminal sequences for each structural gene were confirmed. These points will be addressed more fully in the description of the structural genes of the operon below.

The ilvG and ilvM genes encode the large and small subunits, respectively, of the AHAS II isozyme. The AHAS II isozyme is one of three AHAS isozymes in E. coli K-12. As mentioned in the introduction, the ilvG gene in E. coli K-12 contains a frameshift mutation which results in termination of translation approximately in the middle of the gene (4-6). Mutations have been isolated in E. coli K-12 which displace this frameshift site. These ilvG mutations (previously referred to as ilvO mutations) result in the full length expression of the ilvG gene and, therefore, a functional AHAS II isozyme. The deduced amino acid sequences of the ilvG and ilvM polypeptides are presented in figures 8A and 8B, respectively. The ilvG sequence shown contains the mutation ilvG2096 (5), which differs from the wild type E. coli K-12 sequence by the deletion of a G residue between nucleotides 1252 and 1253 in figure 8A (5,6). Several other mutations have been characterized and each mutation involves a one base pair deletion or a two base pair insertion within a ten base pair region preceding the frameshift site in the ilvG gene of E. coli K-12 (6).

Recently, the AHAS II isozyme from S. typhimurium has been purified and characterized (36). The AHAS II isozyme was shown to be a tetramer composed of two large and two small subunits with molecular weights of 59,300 and 9,700, respectively. The amino terminal sequences of the ilvG and ilvM encoded subunits were determined and these compare almost identically with the deduced E. coli subunit sequences (Fig. 8A and 8B). Of the 35 amino terminal residues determined for the S. typhimurium ilvG subunit there is only a single amino acid difference, a lysyl residue substituted for an asparaginyl residue at position 17 of the E. coli sequence. The amino terminal analysis of the ilvM subunit showed the first 25 residues are identical between S. typhimurium and E. coli. An amino acid composition analysis of the small and large AHAS II subunits of S. typhimurium were also very similar to the deduced polypeptide sequences of E. coli.

The intercistronic region between the ilvG and ilvM genes of E. coli and the homologous region of S. typhimurium is presented in figure 10. The ilvG chain termination codon overlaps the ilvM initiation codon by four nucleotides in both enteric bacteria. There are similar examples of overlapping termination and initiation codons in the trpE-trpD and trpB-trpA intercistronic regions of the trp operon of E. coli. In both examples the pairs of genes which overlap encode non-identical polypeptides which are associated in multi-subunit enzyme complexes. It has been suggested that this juxtaposition of translational signals is important in the coordinate synthesis of these subunits (39-42). It is possible, therefore, that the overlap of the ilvG and ilvM genes

Figure 10. Comparison of the nucleotide sequence of the intercistronic regions of the ilvGMEDA operon of E. coli and S. typhimurium. The nontranscribed DNA strand of the four intercistronic regions of the ilvGMEDA operon of E. coli are presented. In parentheses are the nucleotide positions of each of these intercistronic regions relative to the in vivo transcriptional initiation site. The encoded carboxy and amino terminal sequences encoded by the flanking E. coli genes are also shown. The differences in the same intercistronic regions of S. typhimurium are noted above the E. coli nucleotide sequence. A minimum number of gaps in the E. coli and S. typhimurium sequences were introduced to maximize the sequence alignment. Such gaps in the nucleotide sequence comparison are indicated as dashes. Amino acid differences in S. typhimurium as compared to E. coli are shown in parentheses below the E. coli polypeptide sequences. The proposed ribosomal binding sites preceding each of the initiation codons are underlined. In the ilvG-ilvM intercistronic region the two gene coding regions overlap by four nucleotides, therefore, the ilvM sequence is shown overlapping the ilvG sequence.

also functions in an analogous manner to insure the equimolar expression of the large and small AHAS II subunits (36).

The primary branched chain amino acid transaminase, transaminase B, is the product of the ilvE gene. Transaminase B from E. coli has been purified to apparent homogeneity and the native enzyme was characterized as a hexamer of identical subunits with molecular weights of about 32,000 (37). The amino terminal sequence and amino acid composition was determined for the ilvE encoded product and both compare favorably with the deduced amino acid sequence determined from the nucleotide sequence (Fig. 8C) (37,38). One interesting point concerning the amino terminal sequence of transaminase B is that it appears to be post-translationally modified. Two independent determinations of this sequence begin Gly Thr Lys Lys Ala. When compared to the deduced amino acid sequence in figure 8C, it appears that the amino terminal Met residue has been cleaved and the Thr residue modified to a Gly residue. In S. typhimurium, the amino terminal sequence of transaminase B is Thr Thr Lys Lys Ala (R. R. Randall, M. H. Wallis, G. J. Young and F. B. Armstrong, Fed. Am. Abstr. 38, 325, 1979). In this example it appears that only the Met residue has been cleaved.

The ilvD gene encodes dihydroxy acid dehydrase, which catalyzes the dehydration of the α,β-dihydroxy acids to yield the corresponding α-keto acids (Fig. 1). The nucleotide sequence of the nontranscribed DNA strand and corresponding deduced amino acid sequence of the ilvD product is presented in figure 8D. The encoded polypeptide has a molecular weight of approximately 66,000 which is in agreement with that determined from electrophoresis of labeled maxicell products of the ilvD gene (43,44). It is interesting to note that in the 64 bp intercistronic region between the ilvE and ilvD genes there is an 11 bp inverted repeat (Fig. 8D). Although there is no direct evidence, it is possible that this region, which ends with 3 T residues in row, could function as a weak, operon internal, rho-independent terminator (45). In the ilvE-ilvD intercistronic region of S. typhimurium this region of dyad symmetry is not present (Fig. 10); however, there is approximately 100 bp of additional sequence in this intercistronic region of S. typhimurium which is not represented in E. coli. At present we are not certain of the origin or function of this insertion/deletion in the ilvE-ilvD intercistronic region of E. coli and S. typhimurium.

The most distal gene of the operon, ilvA encodes threonine deaminase. This enzyme catalyzes the first enzymatic step of the isoleucine pathway (Fig. 1). In 1956, Umbarger made the classical observation that threonine deaminase was inhibited by isoleucine (46). This phenomena whereby the end-product of a pathway inhibits the first enzymatic step of a pathway was termed end-product inhibition. End-product inhibition has subsequently been shown to be a universal type of metabolic regulation.

Threonine deaminase has been purified to near homogeneity in many organisms, including Bacillus subtilis (47), S. typhimurium (48), E. coli (49-51) and Saccharomyces cerevisiae (52,53). It is a tetramer of identical subunits, and, in E. coli, has a monomer molecular weight of 53,000 as determined by SDS gel electrophoresis (49). The nucleotide sequence of the nontranscribed DNA strand of the ilvA gene and the corresponding deduced threonine deaminase sequence is presented in figure 8E. We reported previously the nucleotide sequence of the 3'-portion of the ilvA gene (9). The deduced 76 amino acids of this carboxy terminal region of E. coli threonine deaminase

shares 51% homology with the deduced carboxy terminus of the S. cerevisiae enzyme (9,54). With the completion of the nucleotide sequence of the ilvA gene from E. coli we compared the entire deduced threonine deaminase sequence of E. coli with the deduced sequence from S. cerevisiae (54). The gene encoding threonine deaminase from S. cerevisiae has an open translational reading frame which could encode a polypeptide of 63,700, but the purified subunit has been characterized to have a molecular weight of 48,000 (52). Kielland-Brandt et al. (54) suggested that since threonine deaminase in S. cerevisiae is located in the mitochondria a larger molecular weight precursor polypeptide is processed by proteolytic cleavage into a 48,000 molecular weight subunit. This processing is consistent with the cleaving of a signal sequence characterized for many proteins that are synthesized in the cytoplasm and transported into the mitochondria (55). As predicted by the processing model there is no detectable homology between the first 62 amino acids of threonine deaminase from S. cerevisiae and the amino terminal sequence of E. coli. Amino acids 16 through 514 from the threonine deaminase sequence of E. coli were aligned with residues 63 through 592 from the S. cerevisiae sequence. A minimum number of gaps were allowed to maximize the alignment between the two sequences. This alignment shows 50% identical amino acid homology between the paired sequences. The homology is interspersed throughout the threonine deaminase sequences with a large gap of non-homology between the E. coli residues 371-390 and S. cerevisiae residues 434-462.

IV. QUESTIONS TO BE ADDRESSED IN THE FUTURE

In the preceding sections we presented the complete nucleotide sequence and described the structural and functional features of the regulatory and coding regions of the ilvGMEDA operon of E. coli. In this section we consider some remaining questions concerning the role of the ilvGMEDA proximal promoter and attenuator in the regulation of this operon.

Attenuation, a mechanism involved in the control of several amino acid biosynthetic operons, was first described in the trp operon of E. coli. In addition to attenuation, the trp operon is also regulated by repression. The trp repressor regulates transcription initiation of the operon in response to changes in the intracellular availability of free tryptophan, while attenuation regulates transcription termination in the leader region of the trp operon in response to the intracellular levels of aminoacylated $tRNA^{Trp}$. Yanofsky et al. (56) have shown that repression and attenuation respond to different degrees of tryptophan starvation. The trp repressor regulates transcription initiation during growth with excess to moderately limiting levels of tryptophan, whereas, attenuation is not relaxed until tryptophan limitation is in the range of moderate to severe. Unlike the trp operon, however, there is no direct evidence for the involvement of a repressor in the regulation of the ilvGMEDA operon. This implies that this operon is regulated exclusively by attenuation. Such a possibility creates an apparent paradox. During increasing growth rates accompanying a shift from a poor to a good carbon source, protein synthesis rates increase. Concomitantly, the rates of synthesis of the branched chain amino acids increase. If attenuation of the ilvGMEDA operon is responsible for this growth rate regulation, and if attenuation of the ilvGMEDA operon parallels that of the trp operon, then relaxed attenuation at increased

growth rates must be accompanied by a moderate to severe starvation for aminoacylated branched chain tRNAs. However, such limitations for aminoacylated tRNAs are not observed during rapid rates of growth and protein synthesis. In order to accomodate this dilemma it has been noted that amino acid biosynthetic operons which are apparently regulated solely by attenuation (ilvGMEDA, his, thr, leu) contain several tandem regulatory codons, whereas, the attenuator region of the trp operon contains only two tandem trp codons. It has been proposed that there is a translational delay at each regulatory codon which is amplified at each successive regulatory codon (57). This "queuing effect" could make attenuation hypersensitive to relatively small changes in the intracellular levels of the regulatory aminoacylated tRNA. Additionally, the fact that rare codons recognized by minor tRNA isoacceptor species are used as regulatory codons for attenuation may further enhance this sensitivity. One last point, which is surely important, is the coding context of these tandem regulatory codons. For example, the leader polypeptide of the ilvGMEDA operon of S. marcescens encodes a single Leu regulatory codon (30). Since this codon, albeit rare, appears in structural gene coding sequences, its role as a regulatory codon in the leader transcript is probably enhanced by non-leucine flanking codons.

Structural similarities between the promoter region of the ilvGMEDA operon and the tyrT gene, which is regulated by growth rate control, suggests an alternative explanation for the mechanism of growth rate regulation of the ilvGMEDA operon that does not involve attenuation. A deletion analysis of the promoter region of the tyrT gene of E. coli has demonstrated that, like the ilvGMEDA promoter region, nucleotide sequences upstream of this promoter (between positions -40 to -98) are required for maximal expression of the tyrT gene (58). These sequences upstream of the tyrT promoter have affinity for RNA polymerase in vitro as judged by DNAase I protection assays (59). The promoter region of the ilvGMEDA operon consists of tandem in vitro promoters P1P2 (Fig. 5) (14,16,18). In vivo, however, initiation of transcription is detected at only the downstream P2 promoter (16). Travers at al. (59) have proposed a model in which an RNA polymerase bound at an upstream binding site activates transcription from a polymerase bound at the downstream primary promoter. Such an activation could involve protein-protein contact between the adjacent RNA polymerase molecules. If this is so, then tyrT promoter activity would be proportional to the square of polymerase concentration. Since the concentration of RNA polymerase in the cell increases linearly with the growth rate, such a mechanism could explain why the synthesis of stable RNA is proportional to the square of the growth rate. An analogous model can be invoked to explain the role of the dual promoters of the ilvGMEDA operon. The cooperative action of RNA polymerase bound at the upstream P1 binding site could activate transcription, in vivo, by a second RNA polymerase bound at the downstream P2 promoter. This cooperative action could account for the growth rate regulation of the ilvGMEDA operon.

In conclusion, a fundamental problem for the future is to elucidate the physiological roles of the ilvGMEDA promoter and attenuator regions. Most certainly, the attenuation mechanism functions to increase expression of the operon during conditions of severe limitation of aminoacylated branched chain amino acid tRNAs. But this only explains why elevated levels of operon expression are observed during

nutritional shift up or auxotroph starvation experiments. The sites and mechanisms of action of growth rate control and end-product repression remain uncertain.

ACKNOWLEDGEMENTS

We thank Charles Yanofsky and Robert Landick for helpful discussions concerning the implications of RNA polymerase pausing in the attenuator region of the ilvGMEDA operon. This work was supported in part by grants GM24330 (G.W.H.) and GM28021 (R.P.L.) from the National Institutes of Health.

REFERENCES

1. Umbarger, H. E. (1978) Ann. Rev. Biochem. 47, 533-606.
2. Umbarger, H. E. (1983) in Amino Acids: Biosynthesis and Genetic Regulation, eds. Herrmann, K. M. and Somerville, R. L. (Reading, Massachusetts, Addison-Wesley), pp. 245-266.
3. Guardiola, J., DeFelice, M., Lamberti, A. and Iaccarino, M. (1977) Molec. Gen. Genet. 156, 17-25.
4. Favre, R., Wiater, A., Puppo, S., Iaccarino, M., Noelle, R. and Freundlich, M. (1976) Molec. Gen Genet. 143, 243-255.
5. Lawther, R. P., Calhoun, D. H., Adams, C. W., Hauser, C. A., Gray, J. and Hatfield, G. W. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 922-925.
6. Lawther, R. P., Calhoun, D. H., Gray, J., Adams, C. W., Hauser, C. A. and Hatfield, G. W. (1982) J. Bacteriol. 149, 294-298.
7. Bachmann, B. J. (1983) Microbiol. Rev. 47, 180-230.
8. Watson, M. D., Wild, J. and Umbarger, H. E. (1979) J. Bacteriol. 139, 1014-1020.
9. Wek, R. C. and Hatfield, G. W. (1986) J. Biol. Chem. 261, 2441-2450.
10. Sutton, A. and Freundlich, M. (1980) Molec. Gen Genet. 178, 179-183.
11. Friden, P., Newman, T. and Freundlich, M. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6156-6160.
12. Hauser, C. A. and Hatfield, G. W. (1983) Nucleic Acids Res. 11, 127-139.
13. Haughn, G. W., Squires, C. H. DeFelice, M., Largo, C. T. and Calvo, J. M. (1985) J. Bacteriol. 163, 186-198.
14. Lawther, R. P. and Hatfield, G. W. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1862-1866.
15. Nargang, F. E., Subrahmanyam, C. S. and Umbarger, H. E. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1823-1827.
16. Adams, C. W., Rosenberg, M. and Hatfield, G. W. (1985) J. Biol. Chem. 260, 8538-8544.
17. Hawley, D. K. and McClure, W. R. (1983) Nucleic Acids Res. 11, 2237-2255.
18. Adams, C. W. and Hatfield, G. W. (1984) J. Biol. Chem. 259, 7399-7403.
19. Calhoun, D. H., Wallen, J. W., Traub, L., Gray, J. E. and Kung, H. (1985) J. Bacteriol. 161, 128-132.
20. Wek, R. C. and Hatfield, G. W. (1986) Nucleic Acids Res. 14, 2763-2777.
21. Lopes, J. M. and Lawther, R. P. (1986) Nucleic Acids Res. 14, 2779-2798.
22. Berg, C. M., Shaw, K. J., Vender, J. and Boruchka-Mankiewicz, M. (1979) Genetics 93, 309-319.
23. Berg, C. M. and Shaw, K. J. (1981) J. Bacteriol. 145, 984-989.
24. Blazey, D. L. and Burns, R. O. (1979) Molec. Gen. Genet. 177, 1-11.
25. Freundlich, M., Burns, R. O. and Umbarger, H. E. (1962) Proc. Natl. Acad. Sci. U.S.A. 48, 1804-1808.
26. Kolter, R. and Yanofsky, C. (1982) Ann. Rev. Genet. 16, 113-134.
27. Landick, R., Carey, J. and Yanofsky, C. (1985) Proc. Natl. Acad. Sci. 82, 4663-4667.
28. Steitz, J. A. (1979) in Biological Regulation and Development, ed. Goldberger, R. (Plenum, New York), pp. 349-389.

29. Oxender, D. L., Zurawski, G. and Yanofsky, C. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 5524-5528.
30. Hsu, J., Harms, E. and Umbarger, H. E. (1985) J. Bacteriol. 164, 217-222.
31. Hauser, C. A., Sharp, J. A., Hatfield, L. K. and Hatfield, G. W. (1985) J. Biol. Chem. 260, 1765-1770.
32. Platt, T. (1986) Ann. Rev. Biochem. 55, 339-372.
33. Kuramitsu, S., Ogawa, T., Ogawa H. and Kagamiyama, H. (1985) J. Biochem. 97, 993-999.
34. Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
35. Sanger, F., Nicklen, S. and Coulsen, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
36. Schloss, J. V., Van Dyk, D. E. Vasta, J. F. and Kutny, R. M. (1985) Biochemistry 24, 4952-4959.
37. Lee-Peng, F., Hermodson, M. A. and Kohlhaw, G. B. (1979) J. Bacteriol. 139, 339-345.
38. Adams, C. W., Lawther, R. P. and Hatfield, G. W. (1979) Biochem. Biophys. Res. Commun. 89, 650-658.
39. Oppenheim, D. and Yanofsky, C. (1980) Genetics 95, 785-795.
40. Yanofsky, C., Platt, T., Crawford, L P., Nichols, B. P., Christie, G. E., Horowitz, H., VanCleemput, M. and Wu, A. M. (1981) Nucleic Acids Res. 9, 6647-6668.
41. Das, A. and Yanofsky, C. (1984) Nucleic Acids Res. 12, 4757-4768.
42. Aksoy, S., Squires, C. L. and Squires, C. (1984) J. Bacteriol. 157, 363-367.
43. Uzan, M., Favre, R., Gallay, E. and Caro, L. (1981) Molec. Gen. Genet. 182, 462-470.
44. Gray, J. E., Patin, D. W. and Calhoun, D. H. (1981) Molec. Gen. Genet. 183, 428-436.
45. Christie, G. E., Farnham, P. J. and Platt, T. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 4180-4184.
46. Umbarger, H. E. (1956) Science 123, 848.
47. Hatfield, G. W. and Umbarger, H. E. (1970) J. Biol. Chem. 245, 1736-1741.
48. Decedue, C. J., Hofler, J. G. and Burns, R. O. (1975) J. Biol. Chem. 250, 1563-1570.
49. Calhoun, D. H., Rimerman, R. A. and Hatfield, G. W. (1973) J. Biol. Chem. 248, 3511-3516.
50. Grimminger, H., Rahimi-Laridjani, L, Coerner, K. and Lingens, F. (1973) FEBS Lett. 35, 273-275.
51. Kagan, Z. S., Dorozhko, A. I., Kovaleva, S. V. and Yakovleva, L. I. (1975) Biochim. Biophys Acta 403, 208-220.
52. Katsumma, T., Elsasser, S. and Holzer, H. (1971) Eur. J. Biochem. 24, 83-87.
53. Ahmed, S., Bollon, A., Rogers, S., Magee, P. (1976) Biochimie 58, 225-232.
54. Kielland-Brandt, M. C., Holmberg, S., Petersen, J.G. L. (1984) Carlsberg Res. Commun. 49, 567-575.
55. Schatz, G. and Butow, R. A. (1983) Cell 32, 316-318.
56. Yanofsky, C., Kelley, R. L. and Horn, V. (1984) J. Bacteriol. 158, 1018-1024.
57. Ingraham, J. L., Maaloe, O. and Neidhardt, F. (1983) Growth of the Bacterial Cell (Sinauer Associates, Sunderland, Mass.).
58. Lamond, A. I. and Travers, A. A. (1983) Nature 305, 248-250.
59. Travers, A. A., Lamond, A. I., Mace, H. A. F. and Berman, M. L. (1983) Cell 35, 265-273.