Regulation of *Caulobacter crescentus* *ilvBN* Gene Expression

JACK C. TARLETON,[†] JALEH MALAKOOTI, AND BERT ELY*

Department of Biological Sciences, University of South Carolina,
Columbia, South Carolina 29208

Received 1 November 1993/Accepted 4 April 1994

As part of an effort to determine the mechanisms employed by *Caulobacter crescentus* to regulate gene expression, the *ilvBN* genes encoding the two subunits of an acetohydroxy acid synthase (AHAS) have been characterized. Analysis of the DNA sequences indicated that the *C. crescentus* AHAS was highly homologous to AHAS isozymes from other organisms. S1 nuclease and primer extension studies demonstrated that transcription initiation occurred 172 bp upstream of the AHAS coding region. The region between the AHAS coding region and the transcription initiation site was shown to have the properties of a transcription attenuator. Deletion analysis of the region containing the stem-loop structure of the proposed attenuator resulted in the derepression of *ilvBN* expression. Thus, it appears that *C. crescentus* uses attenuation to regulate the expression of the *ilvBN* operon.

Caulobacter crescentus is a stalked aquatic bacterium that differentiates during its normal life cycle. An asymmetrical cell division occurs in every cell cycle and appears to be independent of environmental cues (for a review, see reference 9). This differentiation allows *C. crescentus* to have a sessile existence and yet retain the ability to move into new microenvironments, a process that may confer an enhanced adaptability to low nutrient conditions. The stalk serves as a means of attachment to underwater surfaces and increases the cellular surface area for absorption of nutrients in nutrient-poor environments. When the stalked, sessile parental cell divides, it produces a cell like itself and releases a motile, stalkless daughter cell. This motile daughter cell, termed a swarmer, responds chemotactically to nutritional gradients and can move to a more favorable environment where it will not have to compete with the parent cell. Eventually, it matures into a stalked cell and gains the ability to divide.

Recently, the initial characterization of *C. crescentus* strains that are deficient in acetohydroxy acid synthase (AHAS) production was made, and the cloning of the AHAS gene by complementation of these mutants was described (33). AHAS is a bifunctional enzyme that catalyzes the condensation of pyruvate to α -acetohydroxybutyrate in isoleucine synthesis; alternatively, two pyruvate molecules are condensed to yield α -acetolactate in valine synthesis. In enteric bacteria, two adjacent genes encode the large and small subunits of acetohydroxy acid synthases (13, 27, 32, 34). In some cases, the termination codon of the large subunit gene may overlap with the initiation codon of the small subunit gene (6, 22). It has been suggested that this juxtaposition allows the coordinate expression of these genes so that equimolar ratios of the two subunits are expressed (27). This study demonstrates that two adjacent open reading frames (ORFs) exist in the *C. crescentus* *ilvB* region, *ilvB* and *ilvN*, suggesting similar organization of large and small subunit genes in this bacterium. Furthermore, the predicted amino acid sequences of several AHASs are quite similar to that of the *C. crescentus* AHAS.

Attenuation of transcription is the major mechanism governing expression of the *ilvBN* and *ilvGMEDA* operons in the enteric bacteria (14, 21, 24). Similarly, our studies reveal a possible attenuator upstream of the *C. crescentus* *ilvBN* structural genes. Transcription begins upstream of a potential leader peptide sequence, and an mRNA containing the leader sequence appears to have the capacity to form various secondary structures and a *rho*-independent transcription terminator. While overall there is little similarity between the *C. crescentus* and the *Escherichia coli* leader DNA sequences, the predicted secondary structures of the two regions appear to be similar. Deletion of this region results in an increased level of expression of the downstream genes. Thus, the regulation of the *ilvBN* genes appears to be similar to that of the enteric bacteria.

MATERIALS AND METHODS

Bacterial growth conditions. *E. coli* strains containing recombinant plasmids were grown in Luria-Bertani broth (7) supplemented with ampicillin (50 mg/ml) at 37°C. *C. crescentus* strains were grown in peptone yeast extract (PYE) (19) or in minimal medium (28) alone or supplemented with isoleucine (25 mg/ml) and valine (26 mg/ml) at 33°C.

DNA sequencing. The DNA sequence was determined for both strands by a double-stranded template method using Sequenase (U.S. Biochemical Co., Cleveland, Ohio). 7-DeazadGTP often was substituted for dGTP in sequencing reactions to aid in overcoming compressions. On some occasions, formamide was used as a denaturing agent (along with 7 M urea) at a concentration of 20 to 30% or was substituted for urea completely at a concentration of 50% to help resolve compressions particular to the high G+C content of *C. crescentus* DNA (29). DNA sequences were analyzed by using the Genetics Computer Group program (8).

Molecular techniques. Synchronous cultures of CB15N were obtained essentially as described by Bender et al. (2). RNA was isolated from *C. crescentus* cultures (CB15 or CB15 carrying pPLG389) by a modification of the method described by Coleman et al. (5), in which guanidinium isothiocyanate is substituted for urea. S1 nuclease protection studies for determining the approximate 5' end of an mRNA transcript were performed as described by Ross and Winkler (26), except that 1,000 U of S1 nuclease instead of 120 U was added after

* Corresponding author. Mailing address: Department of Biological Sciences, University of South Carolina, Columbia, SC 29208. Phone: (803) 777-2768. Fax: (803) 777-4002.

[†] Present address: Greenwood Genetics Center, Greenwood, SC 29646.

1 K
NNGGTACCGTCGCGCGCCACATTTTCGCGACGACCGCACTGGCGGGTGAACCTGCCCTT 60

61 CGGACGGGCTGATCCTGGCAAGGTACACCGCGGAAAAGGCACAACCTTGGCGCAAAGA 120

121 GCGCCTTGAGCGCCCATCTTACGCATGCCATGACCCTTTCGCTAGCTTTGAGCAAGACT 180
+1

181 CGTGCGCAAAACCATGATCGTACTTATGGAGCGGCGCTTCGCGGGTGACCTCGAAGTCA 240
ValArgLysThrMetIleValLeuMetGluArgProPheAlaGlyEnd

241 CGTCGTAGCTCGGTGCGCGCGCAAGGTCCTTCGGGACCTTTTTCTTTTCCAGCCCCC 300

301 TCTTCGCGCTAATCCCTCCACCGGATCGTCCCATGACCGCCACCAGACCATCGAGAGC 360
11vB
MetThrAlaHisGlnThrIleGluSer

361 CAAGCTCCCGCGACACCGCGCATGCGGCCATGACCGGCGCGAGATCGTGGTCCGCGGC 420
GlnAlaProAlaAspThrAlaAspArgAlaMetThrGlyAlaGluIleValValArgGly

421 CTCGTGGACCAAGGAGTCGAGGTGCTGTTGCGCTATCCGGGCGCGCGGTCTGCGCGATC 480
LeuValAspGlnGlyValGluValLeuPheGlyTyrProGlyGlyAlaValLeuProIle

481 TATGACGCGCTGTTCACGAGCGCGCCTGCAGCACATTCTGGTCCGCGCAGGACGAGGGC 540
TyrAspAlaLeuPheHisGluProArgLeuGlnHisIleLeuValArgHisGluGlnGly

541 GCGGCGCACGCGCGCAGAAGGCTATGCGCGCTCGTCCGGCAAGCGGGCGTGGTCTGGTC 600
AlaAlaHisAlaAlaGluGlyTyrAlaArgSerSerGlyLysProGlyValValLeuVal

601 ACCTCGGGCCCGCGCGACCAAGCCATCACCGCATCATGACGCGCTGATGGACTCG 660
ThrSerGlyProGlyAlaThrAsnAlaIleThrGlyIleMetAspAlaLeuMetAspSer

661 ATCCCGATGGTCTCATACCGGCCAGGTTCCGACGCACCTGATCGGCACCGACGCTTTC 720
IleProMetValValIleThrGlyGlnValProThrHisLeuIleGlyThrAspAlaPhe

721 CAGGAAGCCGATACAGTCGGCATGACCGCTCGTGCACCAAGCACAACCTATCTGGTCAAA 780
GlnGluAlaAspThrValGlyMetThrArgSerCysThrLysHisAsnTyrLeuValLys

781 GACGTCGCGACCTGCGCGCAGATCATCCACGAGGGGTTCAAGATCGCCACCACCGCGCGC 840
AspValArgAspLeuProGlnIleIleHisGluAlaPheLysIleAlaThrThrGlyArg

841 CCTGGCCCCGACTGATCGACATCCCAAGGACGTCCAGTTCGCCAAGGGCGAGTACTTC 900
ProGlyProValLeuIleAspIleProLysAspValGlnPheAlaLysGlyGluTyrPhe

901 GGTCCCGCGAGGTGCGCTCGACCCACGCTATGCGCGCGCACCAAGGGCGACGCGCGC 960
GlyProGlyGluValAlaSerThrHisAlaTyrAlaProArgThrLysGlyAspAlaGly

961 CGCATCGCGAGGCGCGCAGGATGATCGCCCAAGCGCGCGACCGATCTTCTACACCGGC 1020
ArgIleAlaGluAlaAlaArgMetIleAlaGlnAlaArgArgProIlePheTyrThrGly

1021 GGCGGCGTCATCAACGCGCGCCCAAGGCCAGCGCGGCGCTCGCGAGTTCGCGCGCTG 1080
GlyGlyValIleAsnAlaGlyProLysAlaSerAlaGlyLeuArgGluPheAlaAlaLeu

1081 ACGGGCGCGCGGTACCTCGACGCTGATGGGCCTGGGCGCCTTCCGCGCGCGGATCCG 1140
ThrGlyAlaProValThrSerThrLeuMetGlyLeuGlyAlaPheProAlaAlaAspPro

1141 GCGTGGCTGGGCATGCTGGGCATGCACGGCACGTTGAGGCCAACAAGCCATGCACGAC 1200
AlaTrpLeuGlyMetLeuGlyMetHisGlyThrPheGluAlaAsnAsnAlaMetHisAsp

1201 TGGCAGCTGATGATCTGCGTCGCGCGCGGTTGATGACCGGGTCACGGCGCGCTCGAC 1260
CysAspValMetIleCysValGlyAlaArgPheAspAspArgValThrGlyArgLeuAsp

1261 GCCTTCTCGCGCGGACGAAGAAGATCCACATCGACATCGACCGCTCTCGATCAACAAA 1320
AlaPheSerProGlySerLysLysIleHisIleAspIleAspProSerSerIleAsnLys

1321 AACGTCGCGGTGACCTGCCCATCGTCGCGCACGCGGTTTCGTAAGACCTGATC 1380
AsnValArgValAspLeuProIleValGlyAspAlaGlySerValLeuGluAspLeuIle

1381 GCCGCTGGAAGCGCGCAACCTACAGCCGAACAAGCGCGCTGACCGACTGGTGGGCC 1440
AlaAlaTrpLysAlaAlaAsnLeuGlnProAsnLysAlaAlaLeuThrAspTrpTrpAla

1441 CAGATCGACCAAGTGGCGCGCGCTCAGTGCCTGTCTATCGCGCTCGGACTCGGTCATC 1500
GlnIleAspGlnTrpArgAlaArgGlnCysLeuSerTyrArgArgSerAspSerValIle

1501 AAGCGCAGTACGCCATCGAGCGCTGTACGAGCTGACCAAGGACAAGGACGTCTACATC 1560
LysProGlnTyrAlaIleGluArgLeuTyrGluLeuThrLysAspLysAspValTyrIle

FIG. 1. The complete DNA sequence of the *ilvBN* locus and the predicted amino acid sequence of the *ilvB* and *ilvN* gene products. Numbering of the nucleotide sequence begins at the *KpnI* site. Restriction enzyme sites are noted: K, *KpnI*; S, *SmaI*. The transcription initiation site is denoted at position 162 with “+1.” The –35 and –10 regions are underlined with broken lines, while potential RBSs are underlined with solid lines. The initial positions of the coding regions are noted above the sequence.

1561	ACCACCGAAGTCGGCCAGCACCAGATGTGGCCCGCAGTTCTTCGCTTCGAGGAGCCG	1620
	ThrThrGluValGlyGlnHisGlnMetTrpAlaAlaGlnPhePheArgPheGluGluPro	
1621	AACCGCTGGATGACATCCGGGGGCTCGGCACCATGGGCTACGGGCTGCCCCGCCCTC	1680
	AsnArgTrpMetThrSerGlyGlyLeuGlyThrMetGlyTyrGlyLeuProAlaAlaLeu	
1681	GGTGTGCAACTGGCGCACCGAACAGCCTGGTGGTGGACATCGCCGGCGAAGCCTCGATC	1740
	GlyValGlnLeuAlaHisProAsnSerLeuValValAspIleAlaGlyGluAlaSerIle	
1741	CAGATGTGCATCCAGGAGCTGTGACGGCGATCCAGTTCGATCTGCCGGTCAAGATCTTC	1800
	GlnMetCysIleGlnGluLeuSerThrAlaIleGlnPheAspLeuProValLysIlePhe	
1801	ATCCTGAACAACGAGTGGATGGGCATGGTCCGCCAGTGGCAACAGCTGCTGCACGGCGAG	1860
	IleLeuAsnAsnGluTrpMetGlyMetValArgGlnTrpGlnGlnLeuLeuHisGlyGlu	
1861	CGCTATAGCCACTCCTATTCCGACAGCCTGCCCGACTTCGTGAAGCTGGCCGAGGCCTAT	1920
	ArgTyrSerHisSerTyrSerAspSerLeuProAspPheValLysLeuAlaGluAlaTyr	
1921	GGCGCGGTCCGATCCGTTCGCACAACCCCGCGAGCTGGACGCCAAGATCCTGGAGATG	1980
	GlyAlaValGlyIleArgCysAspAsnProAlaGluLeuAspAlaLysIleLeuGluMet	
1981	GTCAACAGCGACAAGCCGGTGTCTTCGACTGCCGCGTGGAGAAGCACGAGAACTGCCTG	2040
	ValAsnSerAspLysProValIlePheAspCysArgValGluLysHisGluAsnCysLeu	
2041	CCGATGATCCCGTCAGGCAAGGCGCACACGAGATGATCATGGCGAGGTTCGAGGACATC	2100
	ProMetIleProSerGlyLysAlaHisAsnGluMetIleMetGlyGluValGluAspIle	
2101	GGCAACGTATCCGCGAGGACGGCGGGGGCTGGTCTAAGCCCCCTCGCCCTCCACAAC	2160
	GlyAsnValIleGlyGluAspGlyAlaGlyLeuValEnd	
	<i>ilvN</i>	
2161	CCTTTGGCTTAGAAGCGATCCATGACGCCAACGTGCAACCCGCTCCCGCTCGGCCTAC	2220
	MetThrAlaAsnValGlnProAlaProAlaSerAlaTyr	
2221	GACCTTAGCCCCAAGGACAGGCCGAACAGTCCGCCACCTTCGCCCTCCTCGTCGACAAC	2280
	AspLeuSerProLysAspGlnAlaGluGlnSerAlaThrPheAlaLeuLeuValAspAsn	
2281	GAGCCCGCGCTCCTGCATCGCGTGGTGGGCTGTTCCGCGCGCGGGGTACAAATATCGAG	2340
	GluProGlyValLeuHisArgValValGlyLeuPheAlaAlaArgGlyTyrAsnIleGlu	
2341	AGCCTCACCGTCGCGAGACGGACCGCAAGGCGCACACCAGCCGTATCACCGTGGTGACC	2400
	SerLeuThrValAlaGluThrAspArgLysAlaHisThrSerArgIleThrValValThr	
2401	CGCGGCACCCGCCATGTGCTGGACCATCGAGGCCAGCTAAACAAGGTGGTCAATGTC	2460
	ArgGlyThrArgHisValLeuAspGlnIleGluAlaGlnLeuAsnLysValValAsnVal	
2461	CGCCGCGTCCACGACGTGACCCGCGATCCGAACGGCGTGGAGCGCGAGCTGGCCCTGGTG	2520
	ArgArgValHisAspValThrArgAspProAsnGlyValGluArgGluLeuAlaLeuVal	
2521	AAGGTGCGCGCTCTGGCGTCGATCGCCTGGAACGCTGCGGATCGCCGAGATCTTCGGC	2580
	LysValArgGlySerGlyValAspArgLeuGluAlaLeuArgIleAlaGluIlePheArg	
2581	GCCAAGCCGGTCGACACCACCTTGAGAGCTTCGTCTTCGAGATCTCCGGTGGCGCTCC	2640
	AlaLysProValAspThrThrLeuGluSerPheValPheGluIleSerGlyAlaProSer	
2641	AAGATCGACAAGTTCCTCGACCTGATCGGCCCTGGGCTGGTGGAGCTGTCCCGTACG	2700
	LysIleAspLysPheLeuAspLeuMetArgProLeuGlyLeuValGluLeuSerArgThr	
2701	GGCGTTCTCTCCATCGAGCGCGGCTTCGAAGGCATGTAGCCGAGGTGGTTGGGGACAAC	2760
	GlyValLeuSerIleGluArgGlyPheGluGlyMetEnd	
2761	GCGCCCCCTCACCCCTTCATTGGCTGGCGAAGTCCCTATATTGGATCTGTCCGGCTTCG	2820
	<u>S</u>	
2821	GCCGAACATATGGGATAAACGCGCACGTAATAAGCGGACTGGACCCGGG	2869

FIG. 1—Continued.

DNA-RNA hybridization. Primer extension experiments using cloned murine Maloney virus reverse transcriptase were carried out according to the procedure of Pereira et al. (25) by using the primer BE6, GGGTGGAGGGATTAGGCGAA.

In vitro mutagenesis. Deletions of the *ilvBN* leader region were obtained by PCR amplification of plasmid pJM11 as described by Hemsley et al. (17). Primer I, TTCCAGCCCC CTCTTCCGCC, corresponded to residues 290 to 309 and primer II, TCCCGAAGGATCCTTTGCGCG, was comple-

mentary to residues 278 to 259 except for an extra T added after position 269 to create a *Bam*HI site. Since the two primers were in opposite orientations, amplification and subsequent ligation resulted in an 11-bp deletion of the region between the two primers. A second, larger deletion was constructed by removing all the DNA downstream of the newly constructed *Bam*HI site. The extent of both deletions was confirmed by DNA sequencing.

Plasmid constructions. To construct a promoter test vector

FIG. 2. Comparison of the *C. crescentus* (*Ce.*) IlvB amino acid sequence with the sequences of the three *E. coli* (*Ec.*) isozymes and the corresponding yeast (*Sc.*) AHAS. Amino acid residues common to four of the five proteins are shaded. An asterisk above a particular position indicates that the amino acid present at that position in *C. crescentus* is also present in one of the others.

Cc.B	E A N N A M H D C D V M I C V G A R F F D D R V T G R L D A F S P G S K K	
Ec.B	S T N Y I L H E A D L L I C V L G A R F F D D R A I G K T E Q F C P N A K I	
Ec.G	A A N F A V Q E C D L L I A V G A R F F D D R V T G K L N T F A P N H A S V	
Ec.I	E A N M T M H N A D V I F A V G V R F F D D R T T N N L A K Y C P N A T V	
Sc.2	T A N L A V Q N A D L I I A V G A R F F D D R V T G N I S K F A P E A R R	
Cc.B	- - - - -	I H I D I D P S S I N K N V R V D L P I V G D A G S V
Ec.B	- - - - -	I H V D I D R A Q L G K I K Q P H V A I Q A D V D D V
Ec.G	- - - - -	I H M D I D P A E M N K L R Q A H V A L Q G D L N A L
Ec.I	- - - - -	L H I D I D P T S I S K T V T A D I P I V G D A R Q V
Sc.2	A A A E G R G G I I H F E V S P K N I N K V V Q T Q I A V E G D A T T N	
Cc.B	I E D L I A A W K A A N L Q P N K A A L T D W W A Q I D Q W R A R Q C L	
Ec.B	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	
Ec.G	L P A L Q Q P L N Q Y D W Q Q H C A - - - - - Q L R D E H S W	
Ec.I	L E Q M L E L L S Q E S A H Q P L D E I R D W W Q Q I E Q W R A R Q C L	
Sc.2	I G K M M - - - - S K I F P V K E R - S E W F A Q I N K W K K E Y P Y	
Cc.B	S Y R R - - S D S V I K P Q Y A I E R L Y E L T K D - - K D V Y I T T E	
Ec.B	C P 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	
Ec.G	R Y D H - - P G D A I Y A P L L L K Q L S D - R K P - - A D C V V T T D	
Ec.I	K Y D T - - H S E K I K P Q A V I E T L W R L T K G - - D A Y V T S D	
Sc.2	A Y M E E T P G S K I K P Q T V I K K L S K V A N D T G R H V I V T T G	
Cc.B	V G Q H Q M W A A Q F F R F E E P N R W M T S G G L G T M G Y G L P A A	
Ec.B	V G Q H Q M W T A Q A Y P L N R P P R Q W L T S G G L G T M G F G L P A A	
Ec.G	V G Q H Q M W A A Q H I A H T R P E N F I T S G G L G T M G F G L P A A	
Ec.I	V G Q H Q M F A A L Y Y P F D K P R R W I N S G G L G S M G F G L P A A	
Sc.2	V G Q H Q M W A A Q H W T Q R N P H T F I T S G G L G T M G Y G L P A A	
Cc.B	L G V Q L A H P N S L V V D I A G E A S I Q N C I Q E L S T A S I Q F D L	
Ec.B	I G A A L A N P D R K V L C F S G D G S L M M N I Q E M A T A S I S E N Q L	
Ec.G	V G A Q V A R P N D T V V C I S G D G S F M M N V Q E L G T V K R K Q L	
Ec.I	L G V K M A L P E E T V V C V T G D G S I Q N N I Q E L S T A L Q Y E L	
Sc.2	I G A Q V A K P E S L V I D I D G D A S F N M T L T E L S S A V Q A G T	
Cc.B	P V K I F I L N N E W M G N V R Q Q Q Q L L H G E R Y S H S Y S D S L P	
Ec.B	D V K I I L M N N E A L G L V H Q Q Q Q L F F Y E Q G V F A A T Y P G K I	
Ec.G	P L K I V L L D N Q R L G M V R Q Q Q Q L F F Q E R Y S E T T L T D N P	
Ec.I	P V L V V N L N N R Y L G M V K Q M Q D M I Y S G R H S Q S Y M Q S L P	
Sc.2	P V K I L I L N N E E Q G N V T Q W Q S I F Y E H R Y S H T H - Q L N P	
Cc.B	D F V K L A E A Y G A V G I R C D N P A E L D A K I L E M V N S D K P V	
Ec.B	N F M Q I A A G F G L E T C D L N N E A D P Q A S L Q E I I N R P G P A	
Ec.G	D F L M L A S A F G I H G Q H I T R K D Q V E A A L D T M L N S D G P Y	
Ec.I	D F V R R G - A Y G H V G I Q I S H P H G W K A N L A R R W N R C A I I	
Sc.2	D F I K L A E A M G L K G L R V K K Q E E L D A K L K E F V S T K G P V	
Cc.B	I F D C R V E K H E N C L P M I P S G K A H N E M I M G E V E D I G N V	
Ec.B	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 - G E * - - - -	
Ec.G	L L H V S I D E L E N V W P L V P P G A S N S E M L E K L S * - - - -	
Ec.I	A W C L L M L P S M A A S T S T R C R F A G A E W M K C G * - - - -	
Sc.2	L L E V E V D K K V P V L P M V A G G S G L D E F I N F D P E V E R Q Q	
Cc.B	I G E D G A G L V * - - -	
Ec.B	- - - - -	
Ec.G	- - - - -	
Ec.I	- - - - -	
Sc.2	T E L R H K R T G G K H *	

FIG. 2—Continued.

capable of replication in *Caulobacter* spp. a 0.8-kb *Hind*III fragment containing a promoterless *cat* gene from pCM7 (4) was inserted into pUC18 to create pJM2-20. A gentamicin resistance gene was added by filling in the ends of a 1.6-kb *Hind*III fragment from pRAR154 and inserting it into the *Ssp*I site of pJM2-20. The resulting plasmid, pJM2-23, was cleaved with *Sst*I and *Xmn*I. The fragment containing the multiple cloning sites in addition to the *cat* and gentamicin resistance genes was cloned into the IncQ vector pKT230 (1) that had been cleaved with *Sst*I and *Sma*I to yield pJM2-33. This construct could be readily transferred to *C. crescentus* but was unstable. Therefore, the *Sst*I-*Xho*I fragment from pJM2-33, which contained the multiple cloning site, the *cat* gene, and the gentamicin resistance gene was cloned into the IncP vector pRK290-20R cleaved with the same enzymes. The resulting plasmid, pJM2-80, contained the promoterless *cat* gene located downstream from the multiple cloning sites derived from pUC18. It conferred resistance to tetracycline and gentamicin but not to chloramphenicol and was stable in *C. crescentus*. To assay for promoter activity, transcriptional fusions in which the DNA regions harboring the wild-type or mutated promoters were inserted into the multiple cloning sites upstream from the promoterless *cat* gene were constructed. The *cat* activities of the new recombinant plasmids were determined after electroporation of the plasmids into wild-type *C. crescentus*.

Enzyme assay. Assays for AHAS activity were performed by using sonicated CB15 cellular extracts as described by Tarleton and Ely (33). Chloramphenicol acetyltransferase (CAT) activity was assayed by using [¹⁴C]acetyl coenzyme A according to the directions of the manufacturer (New England Nuclear, Boston, Mass.). Strains carrying the transcriptional fusion plasmids were grown to exponential growth phase in PYE medium supplemented with tetracycline (2 mg/liter) in 10-ml culture samples. Cell extracts were prepared by sonic disruption of the cells in 0.1× TE (10 mM Tris, 0.1 mM EDTA, pH 8.0). The cell debris was removed by centrifugation, and the supernatant was assayed for CAT activity.

Nucleotide sequence accession number. The nucleotide sequence shown in Fig. 1 has been deposited with GenBank under accession number L25317.

RESULTS

DNA sequence of *ilvBN* and analysis of translational signals. The sequences of both DNA strands were determined from the unique *Kpn*I restriction site in pPLG389 to the *Sma*I restriction site downstream of the *ilvN* termination codon. Examination of the DNA sequence revealed three possible start codons at positions 182, 334, and 391 and a fourth one at position 2182 (Fig. 1). The first start codon is 21 bases from the transcription start site (see below) and would result in a peptide of 15 amino acids. This ORF is followed by a sequence which resembles a *rho*-independent terminator and is downstream of a potential ribosome binding site (RBS) (10, 30), GGAG, at position 170. Similar ORFs encode leader peptides which function as part of an attenuation mechanism to regulate a number of prokaryotic biosynthetic operons.

The start codons at positions 334 and 391 are in the same reading frame and either one could be the initiation codon for the *ilvB* polypeptide. The first start codon is 11 nucleotides downstream from an RBS sequence GGA, while the only possible RBS for the second is a GA dinucleotide. Furthermore, analysis of codon usage and third position G+C content indicates that the region between bases 334 and 391 conforms to the pattern expected for *C. crescentus* coding regions (29). Thus, the first start codon is likely to be the actual start site for

FIG. 3. Primer extension analysis of the *ilvBN* mRNA. Lane 1, the cDNA generated by reverse transcription of total RNA isolated from CB15 (pPLG389) by using primer BE6. Lanes A to T, dideoxy sequencing reactions using pJCT31 as the template and BE6 as the primer. The cDNA migrates to a position the same as that of a dideoxycytosine-terminated fragment of 163 bases (indicated by an asterisk) and corresponding to the guanine on the complementary DNA strand which is 159 bases downstream of the *Kpn*I site (position 162 in Fig. 1).

the *ilvB* gene. The termination codon for the *ilvB* gene lies at position 2139, 40 bases upstream from the remaining ORF that encodes the *ilvN* polypeptide. The ORF for the *ilvN* gene is preceded by the RBS sequence AGAA and terminates at position 2739. Support for the coding region assignments comes from genetic complementation experiments (33) and from experiments with a *Rhodobacter sphaeroides* in vitro transcription-translation system (3). Previous experiments demonstrated that the *C. crescentus* tryptophan biosynthetic genes were expressed in this system and produced proteins of the predicted sizes (26). When cloned DNA containing the *ilvBN* region was used as a template, two proteins with molecular masses of 65 and 20.5 kDa were synthesized, corresponding to the calculated molecular masses of IlvB and IlvN, respectively (data not shown).

The amino acid sequence deduced for the *ilvB* gene was homologous to the sequences deduced for all three *E. coli* AHAS isozymes (13, 22, 32, 34), the plant AHASs from *Arabidopsis thaliana* and *Nicotiana tabacum* (23), and the *Saccharomyces cerevisiae* AHAS (ILV2) (12). Similarly, the amino acid sequence deduced for the *ilvN* gene was homologous to the sequences of the *E. coli* isozymes (13, 22, 32, 34). When the *C. crescentus* IlvB amino acid sequence is compared with the *E. coli* AHAS isozymes, the percent identity ranges from 43% (IlvB) to 51% (IlvI) (Fig. 2).

Determination of the *ilvBN* transcription start site. The 5' terminus of the *ilvBN* mRNA was determined by using S1 nuclease protection studies. The 511-bp *Kpn*I-*Pst*I fragment, presumed to contain the *ilvB* 5' regulatory region, was end labeled at the *Pst*I site, hybridized to total cellular RNA, and digested with S1 nuclease overnight. Acrylamide gel electrophoresis demonstrated that a 350-bp fragment was protected

when mRNA was isolated from different stages of synchronized cell cultures. Thus, transcription appears to occur from the *ilvB* region throughout the cell cycle. To determine the site of transcription more precisely, a primer extension experiment

FIG. 5. Plasmid constructs used to assess the role of attenuation in *ilvBN* expression. The relative level of CAT enzyme activity is shown for each of the *cat* gene fusions. The values are averages of two independent experiments. The DNA insert in pJM81 is a 0.5-kb *KpnI*-*PstI* fragment containing wild-type promoter sequences. In pJM82 sequences corresponding to positions 279 to 289 (Fig. 1) are deleted, and a new *Bam*HI site is created by insertion of an "A" at position 269. pJM83 contains a DNA region corresponding to positions 1 to 269, and pJM84 contains a *Sau*3AI fragment corresponding to positions 71 to 186. The *cat* gene is depicted as a heavy bar. Abbreviations: A, *Apa*I; B, *Bam*HI; B*, introduced *Bam*HI site; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; S, *Sau*3AI; Sa, *Sal*I; Sm, *Sma*I; //, deletion site.

was performed by using oligonucleotide BE-6, which is complementary to the mRNA-like DNA strand 11 bases upstream of the *ilvB* coding region. A fragment 163 bases in length was produced, indicating that the guanine at position 162 is likely to be the initial nucleotide of the *ilvB* transcript (Fig. 3). Transcription from this nucleotide to the *PstI* site would produce a transcript of 352 nucleotides. Thus, this result is consistent with the S1 nuclease protection studies.

Regulation of *ilvBN* expression. Previous experiments had demonstrated that the expression of both the *trpFBA* operon and the *hisD* gene was constitutive (11, 35). To determine whether expression of the *ilvBN* operon is regulated, AHAS activity was assayed in extracts of CB15 grown in minimal medium, in minimal medium supplemented with isoleucine and valine, and in complex medium. AHAS activity was sixfold higher in cells grown in minimal medium than in those grown in PYE (10.7 compared with 1.7 nmol/min/mg of protein; values are the averages of two independent experiments). AHAS activity in cells grown in minimal medium was 12-fold higher than in cells grown in minimal medium supplemented with isoleucine and valine (0.8 nmol/min/mg). Thus the level of AHAS was reduced below the level found in PYE-grown cells, indicating that AHAS production is regulated in response to the presence of branched-chain amino acids.

The DNA region upstream of the *ilvBN* genes functions as an attenuator. The organization and spacing of the *ilvBN* leader closely resemble those of the *E. coli ilvGMEDA* leader sequence. Furthermore, computer analysis of possible mRNA secondary structure indicates that this region has the hallmarks of a *rho*-independent terminator, i.e. a section of repeated uridine nucleotides preceded by a stem-loop structure (Fig. 4). While the possible peptides encoded within the leader regions of these genes differ in size and composition, the spacing between the termination codons and the *rho*-independent transcription terminator is similar, and three branched-chain amino acid codons are found in the appropriate place. These results suggest that the *ilvBN* genes may be regulated by attenuation in a manner similar to that described for the *E. coli ilvGMEDA* operon. To test this hypothesis, we analyzed the effect of deletions in this region on *ilvBN* expression. A series

of plasmids was constructed with wild-type or deleted regulatory regions fused to a *cat* reporter gene (Fig. 5). These plasmids were introduced into *C. crescentus* by electroporation, and expression was monitored by assaying chloramphenicol acetyltransferase activity. Deletion of nucleotides 279 through 289 (pJM82) eliminated part of the attenuator stem-loop structure and resulted in an 18-fold increase in expression of the downstream *cat* gene. Thus, the *ilvBN* operon appears to be regulated by attenuation. However, fusion of the reporter gene to a site just upstream of the stem-loop structure (pJM83 and pJM84) resulted in a more than 100-fold increase in *cat* gene expression. Thus, the *ilvBN* operon may be regulated by an additional mechanism as well.

DISCUSSION

We have accumulated evidence that the *ilvBN* locus of *C. crescentus* encodes two polypeptides of 67.8 and 20.5 kDa, representing the large and small subunits of the *C. crescentus* AHAS. Computer analysis of the codon usage and third position G+C content derived from DNA sequence conforms to the pattern expected for *C. crescentus* coding regions. Furthermore, the deduced amino acid sequence of the *C. crescentus* AHAS is highly conserved compared with those of AHASs from other species. In addition, genetic analysis of cloned DNA from this locus suggests that it is capable of complementing *ilv* auxotrophs deficient in AHAS activity (33). The proximity of the *ilvN* gene to the translation termination codon of *ilvB* indicates that these genes may be cotranscribed. In all prokaryotic AHAS genes characterized to date, the large and small AHAS subunits are encoded by adjacent cistrons contained within the same transcription unit (13, 21, 32, 34). In contrast, the eukaryotic AHASs described to date consist of a single polypeptide (31). In *E. coli*, the major regulatory control of the *ilvBN* and *ilvGMEDA* operons, which encode AHAS I and AHAS II, occurs by multivalent attenuation mediated through regulatory codons within a 5' leader polypeptide sequence (15, 16). The location of branched-chain amino acid codons in strategic regions of the leader peptide governs the attenuation mechanism. The results of our DNA sequencing

experiments suggest that the *C. crescentus ilvBN* leader sequence encodes a leader peptide and contains a potential *rho*-independent transcription terminator. The relatively short (15-residue) *C. crescentus ilvBN* leader contains only one codon each for isoleucine, valine, and leucine. Leader peptides shorter than 15 residues are not unusual, since the classical example for an attenuation mechanism is the *E. coli trp* operon, in which the leader peptide is only 14 amino acids in length and contains two tandem tryptophan residues (36). Furthermore, a single leucine codon, the less frequently used CUA codon, has been shown to regulate expression under leucine limitation in the *Serratia marcescens ilvGMDA* leader peptide (18). Similarly, examination of the *C. crescentus* codon usage indicates that the leader peptide contains the infrequently used valine (GUA) and leucine (CUU) codons as well as the most frequently used isoleucine codon (AUC). The presence of methionine codons flanking the three tandem branched-chain amino acid codons may also have some significance since we have shown that *ilvA* mutants can utilize methionine for growth. Presumably, this occurs through the production of α -ketobutyrate resulting from methionine degradation in the synthesis of cysteine (33).

Lawther et al. (22) have reviewed the common features of RNA transcripts of amino acid biosynthetic operons regulated by attenuation. These are (i) a coding region for a short polypeptide containing codons for the end-product amino acids; (ii) an early stem-loop structure that functions as an RNA polymerase pause site in the leader polypeptide coding region to ensure coupling of translation and transcription (20); (iii) a stem-loop structure at the 3' end of the leader RNA transcript followed by several uridine residues which define a *rho*-independent transcription termination site; and (iv) a third stem-loop structure, the antiterminator, which forms when the early stem-loop is masked by a ribosome stalled on a regulatory codon. On the basis of these criteria, the *C. crescentus ilvBN* leader appears to be regulated by attenuation. Figure 4 depicts two different mRNA structures possible within the *ilvBN* leader. In this model, five regions within the leader mRNA which may form complementary structures are described. In the presence of excess branched-chain amino acids, RNA polymerase would pause at the 3' base of the early stem-loop 1:2. A translating ribosome would enter stem-loop 1:2 and disrupt the secondary structure, releasing the paused RNA polymerase. The ribosome would pass through region 2 and be released at the end of the leader polypeptide coding region before stem 5 is synthesized. Thus, stem-loop 1:2 would reform, and stem-loop 4:5 would result in transcription termination. In contrast, when any of the branched-chain amino acids are limited during growth, the translating ribosome would stall in the region of stem-loop 1:2. The paused RNA polymerase would proceed and stem-loop 3:4 would be formed to create the antiterminator.

ACKNOWLEDGMENTS

We thank Nelida Caballero and Lilly Gallman for expert technical assistance and Betty Branham for help in preparing the manuscript.

This work was supported by Public Health Service grant GM 34765 to B. Ely.

REFERENCES

1. Bagdasarian, M., R. Lurz, B. Ruckert, F. C. H. Franklin, M. M. Bagdasarian, J. Frey, and K. Timmis. 1981. Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number RSF1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene* **16**:237-247.
2. Bender, R. A., C. M. Refson, and E. A. O'Neill. 1989. Role of the flagellum in cell-cycle-dependent expression of bacteriophage receptor activity in *Caulobacter crescentus*. *J. Bacteriol.* **171**:1035-1040.
3. Chory, J., and S. Kaplan. 1982. The *in vitro* transcription-translation of DNA and RNA templates by extracts of *Rhodospseudomonas sphaeroides*. *J. Biol. Chem.* **257**:15110-15121.
4. Close, T. J., and R. L. Rodriguez. 1982. Construction and characterization of the chloramphenicol-resistance gene cartridge: a new approach to the transcriptional mapping of extrachromosomal elements. *Gene* **20**:305-316.
5. Coleman, J., P. J. Green, and M. Inouye. 1984. The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. *Cell* **37**:429-436.
6. Cox, J., B. Cox, V. Fidanza, and D. Calhoun. 1987. The complete nucleotide sequence of the *ilvGMDA* cluster of *Escherichia coli* K-12. *Gene* **56**:185-198.
7. Davis, R. W., D. Botstein, and J. R. Roth. 1980. Advanced bacterial genetics, p. 201. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
9. Dingwall, A., L. Shapiro, and B. Ely. Flagella biogenesis in *Caulobacter*. In J. P. Armitage and J. M. Lackie, (ed.), *Biology of the chemotactic response*. Cambridge University Press, Cambridge.
10. Ely, B. 1992. DNA sequence of the 3' end of the *Caulobacter crescentus* 16S rRNA gene. *Nucleic Acids Res.* **20**:1423.
11. Ely, B. Unpublished data.
12. Falco, S., K. Dumas, and K. Livak. 1985. Nucleotide sequence of the yeast *ILV2* gene which encodes acetolactate synthase. *Nucleic Acids Res.* **13**:4011-4026.
13. Friden, P., J. Donegan, J. T. Mullen, P. Tsui, and M. Freundlich. 1985. The *ilvB* locus of *Escherichia coli* K-12 is an operon encoding both subunits of acetohydroxy acid synthetase I. *Nucleic Acids Res.* **13**:3979-3993.
14. Friden, P., T. Newman, and M. Freundlich. 1982. Nucleotide sequence of the *ilvB* promoter-regulatory region: a biosynthetic operon controlled by attenuation and cyclic AMP. *Proc. Natl. Acad. Sci. USA* **79**:6156-6160.
15. Harms, E., J. Hsu, C. Subrahmanyam, and H. Umbarger. 1985. Comparison of the regulatory regions of the *ilvGEDA* operons from several enteric organisms. *J. Bacteriol.* **164**:207-216.
16. Hauser, C., and G. W. Hatfield. 1984. Attenuation of the *ilvB* operon by amino acids reflecting substrates or products of the *ilvB* gene product. *Proc. Natl. Acad. Sci. USA* **81**:76-79.
17. Hemsley, A., N. Arnheim, M. D. Toney, G. Cortopassi, and D. J. Galas. 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res.* **17**:6545-6551.
18. Hsu, J., E. Harms, and H. Umbarger. 1985. Leucine regulation of the *ilvGEDA* operon of *Serratia marcescens* by attenuation is modulated by a single leucine codon. *J. Bacteriol.* **164**:217-222.
19. Johnson, R. C., and B. Ely. 1977. Isolation of spontaneously derived mutants of *Caulobacter crescentus*. *Genetics* **86**:25-32.
20. Landick, R., J. Carey, and C. Yanofsky. 1985. Translation activates the paused transcription complex and restores transcription of the *trp* operon leader region. *Proc. Natl. Acad. Sci. USA* **82**:4663-4667.
21. Lawther, R. P., and G. W. Hatfield. 1980. Multivalent translational control of transcription termination by attenuation of *ilvGEDA* operon of *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. USA* **77**:1862-1866.
22. Lawther, R. P., R. C. Wek, J. M. Lopes, R. Pereira, B. E. Taillon, and G. W. Hatfield. 1987. The complete nucleotide sequence of the *ilvGMDA* operon of *Escherichia coli* K-12. *Nucleic Acids Res.* **15**:2137-2155.
23. Mazur, B., C. Chui, and J. K. Smith. 1987. Isolation and characterization of plant genes coding for acetolactate synthase, the target enzyme for two classes of herbicides. *Plant Physiol.* **85**:1110-1117.
24. Nargang, F. E., C. S. Subrahmanyam, and H. E. Umbarger. 1980. Nucleotide sequence of *ilvGEDA* operon attenuator region of

- Escherichia coli*. Proc. Natl. Acad. Sci. USA 77:1823–1827.
25. **Pereira, R. F., M. J. Ortuno, and R. P. Lawther.** 1988. Binding of integration host factor (IHF) to the *ilvGp1* promoter of the *ilvGMEDA* operon of *Escherichia coli* K-12. Nucleic Acids Res. 16:5973–5989.
 26. **Ross, C. M., and M. E. Winkler.** 1988. Structure of the *Caulobacter crescentus* *trpFBA* operon. J. Bacteriol. 170:757–768.
 27. **Schloss, J. V., D. E. V. Dyk, J. F. Vasta, and R. M. Kutny.** 1985. Purification and properties of *Salmonella typhimurium* acetolactate synthase isozyme II from *Escherichia coli* HB101/pDU9. Biochemistry 24:4952–4959.
 28. **Schoenlein, P. V., L. S. Gallman, and B. Ely.** 1989. Organization of the *flaFG* gene cluster and identification of two additional genes involved in flagellum biogenesis in *Caulobacter crescentus*. J. Bacteriol. 171:1544–1553.
 29. **Schoenlein, P. V., L. S. Gallman, M. Winkler, and B. Ely.** 1990. Nucleotide sequence analysis of the *Caulobacter crescentus* *flaF* and *flbT* genes and an analysis of codon usage in organisms with GC-rich genomes. Gene 92:17–25.
 30. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) 254:34–38.
 31. **Smith, J. K., J. V. Schloss, and B. Mazur.** 1989. Functional expression of plant acetolactate synthase genes in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 86:4179–4183.
 32. **Squires, C. H., M. Defelice, J. Devereux, and J. M. Calvo.** 1983. Molecular structure of *ilvIH* and its evolutionary relationship to *ilvG* in *Escherichia coli* K-12. Nucleic Acids Res. 11:5299–5313.
 33. **Tarleton, J., and B. Ely.** 1991. Isolation and characterization of *ilvA*, *ilvBN*, and *ilvD* mutants of *Caulobacter crescentus*. J. Bacteriol. 173:1259–1267.
 34. **Wek, R. C., C. A. Hauser, and G. W. Hatfield.** 1985. The nucleotide sequence of the *ilvBN* operon of *Escherichia coli*: sequence homologies of the acetohydroxy acid synthase isozymes. Nucleic Acids Res. 13:3995–4010.
 35. **Winkler, M. E., P. V. Schoenlein, C. M. Ross, J. T. Barrett, and B. Ely.** 1984. Genetic and physical analyses of *Caulobacter crescentus* *trp* genes. J. Bacteriol. 160:279–287.
 36. **Yanofsky, C.** 1981. Attenuation in the control of expression of bacterial operons. Nature (London) 289:751–758.