Effect of the deletion of upstream DNA sequences on expression from the *ilvGp2* promoter of the *ilvGMEDA* operon of *Escherichia coli* K-12

Manuel J.Ortuno and Robert P.Lawther

Department of Biology, University of South Carolina, Columbia, SC 29208, USA

Received November 12, 1986; Revised January 21, 1987; Accepted January 26, 1987

ABSTRACT

Transcription *in vitro* of the regulatory region of the *ilvGMEDA* operon yields two attenuated RNAs initiated from the tandem promoters *ilvGp1* and *ilvGp2*. Both S1 nuclease analysis and the fusion of *ilvGp1* to *galK* indicate that transcription is not initiated *in vivo* from *ilvGp1*. However deletion of DNA sequences 150 to 100 bp upstream of *ilvGp2* drastically reduces expression *in vivo* from *ilvGp2*. Both the distance separating *ilvGp2* from the upstream DNA sequences and their relative orientation to each other on the DNA helix affect expression from *ilvGp2*. Deletion of DNA sequences approximately 400 bp upstream of *ilvGp2* increases expression *in vivo* from this promoter. Analysis of products of transcription *in vitro* indicates that the effects observed *in vivo* are probably not due to DNA conformation or interactions of RNA polymerase.

INTRODUCTION

In *Escherichia coli* K-12, five of the genes for the biosynthesis of isoleucine and valine form the *ilvGMEDA* operon (1-4). Preceding the structural genes is a regulatory region that includes the promoter(s), the leader, and the attenuator (2,3). Transcription of the *ilv* regulatory region *in vitro* (2,5,6) yields two transcripts, one of 186 and another of 258 nucleotides (nt). The two RNAs are initiated from a pair of tandem promoters and both transcripts end within the attenuator (2,5). Using RNA fingerprint analysis, Adams et al. (5) determined the transcription initiation sites *in vitro* and *in vivo*. Their studies demonstrated that the proposed start site *in vitro* for the 186 nt RNA was correct and that the 258 nt RNA initiated 72 base pairs (bp) upstream *in vitro* (Fig. 1). The start site for the longer RNA was designated as *ilvGp1* and the start site for the shorter RNA was designated as *ilvGp2* (designations are as described [7]). They also analyzed *ilv*-specific transcription under a variety of physiological conditions known to affect the regulation of the *ilvGMEDA* operon. Their analysis of cellular RNA demonstrated initiation of transcription *in vivo* from *ilvGp2* under all the studied growth conditions but it failed to detect transcription from *ilvGp1* *in vivo* (5).

These studies appear to indicate that RNA polymerase initiates tran-

scription in vitro from a site which it does not utilize in vivo. If the above observations are correct, then one must question whether the observed interaction in vitro of RNA polymerase with the DNA sequences designated ilvGp1 represent a site of interaction in vivo for RNA polymerase. Similar observations have been made for other genes involved in amino acid metabolism. Studies of the tyrT gene (8) and the ilvIH operon (9) indicate that RNA polymerase may interact at sequences in the vicinity of the promoter other than those at which transcription initiates in vivo. In the case of the tyrT gene, the site of the initiation of transcription is well characterized (10,11), nevertheless nuclease protection studies indicate that RNA polymerase may interact with sequences more than 45 bp upstream of the initiation site (8). Studies of transcription of the ilvIH operon indicate that the majority of transcripts formed in vitro initiate upstream of the major initiation site utilized in vivo (9). Such data indicate that RNA polymerase interacts with DNA sequences in vitro from which transcription appears not to be initiated in vivo. However, other data indicate that deletion of DNA sequences upstream of the tyrT and ilvIH reduces expression from the initiation sites utilized in vivo. Transcription (in vivo) from the tyrT promoter is decreased 80-90% by deletion of DNA sequences that reside 40-100 bp upstream (12). Similarly, deletion of DNA sequences 200 bp upstream of ilvI reduces expression from the ilvIH promoter by 80 percent (9). Thus, two observations have been made for these genes. First, interactions of RNA polymerase can be detected in vitro with DNA sequences that lie upstream of the site at which initiation occurs in vivo. Second, deletion of these upstream sequences reduces expression in vivo from the promoters for these genes. From these data it was proposed that the RNA polymerase interacted with the upstream DNA sequences to enhance transcription from the promoters.

An analogous phenomenon appears to influence transcription of the ilvGMEDA operon. Adams and Hatfield have reported that upstream DNA sequences enhance transcription of this operon from ilvGp2 (13). Together, the data from their RNA fingerprint analysis and the observation that the upstream DNA region enhances expression from the ilv promoter, suggest that nontranscriptional interactions of RNA polymerase with ilvGp1 may account for the apparent enhancement. The experiments presented in this report were undertaken to ascertain the basis of these observations. In order to clarify the role of the upstream DNA sequences on expression from the ilvGMEDA regulatory region, transcription in vivo has been analyzed by S1 nuclease analysis and fusion of restriction fragments containing portions of the ilv regulatory region to the

galK gene. The data indicate that under the physiological conditions examined, ilvGp1 does not function as a promoter in vivo. However, DNA sequences in the vicinity of ilvGp1 significantly influence expression from ilvGp2. This effect is dependent upon the distance between ilvGp2 and the upstream region as well as their relative angular orientation. Furthermore, the data indicate that expression from the ilvGMEDA promoter is modulated by a previously unrecognized DNA domain that resides more than 350 bp upstream of the site of initiation.

MATERIALS AND METHODS

Bacteria, Plasmids and Media.

The E. coli K-12 strain M152 [galK2, recA3, rpsL200, IN(rrnD-rrnE)1], obtained from the E. coli Genetic Stock Center, served as the recipient for the plasmids used in this study. The plasmids are described in Table 1 with reference to Fig. 1. M63 minimal medium was prepared as described by Miller (17), and contained 100 µg/ml of ampicillin.

Enzymes and Biochemicals.

Restriction endonucleases, T4 DNA ligase, E. coli DNA-dependent RNA polymerase, and the large fragment of E. coli DNA polymerase I were obtained from New England Biolabs. S1 nuclease and [α - 32 P]UTP were obtained from New England Nuclear Corp. Recombinant DNA linkers were obtained from Collaborative Research Inc. D-[1- 14 C] galactose was obtained from Amersham Corporation. All other reagents came from the Sigma Chemical Company.

Recombinant DNA Techniques.

Plasmid DNA was isolated as described previously (18). Restriction endonuclease digestions were performed as suggested by New England Biolabs. Other recombinant DNA techniques were as described by either Davis et al. (19) or Maniatis et al. (20). Each plasmid fusion was confirmed by DNA sequence analysis as described by Maxam and Gilbert (21).

S1 Nuclease Analysis

Cellular RNA was extracted from M152 using the procedure described by Jones et al. (22). The RNA concentration was determined by measuring the absorbance at 260 nm. RNA was prepared by transcription in vitro of 6 µg pRL5 (1 pmole) as described previously (6). Transcription was monitored by the incorporation of 3 H-UTP; the products were purified by extraction with phenol and chloroform and the nucleic acids were precipitated with ethanol. The 416 bp (Fig. 1, -355 to 61 bp), BstNI, restriction fragment was 5' end-labelled with T4 polynucleotide kinase and treated with 4000 units of exonuclease III for 30 min at 37°C in order to generate single stranded probes (23).

TABLE 1. PLASMIDS

<u>Plasmid</u>	<u>Description</u>	<u>Reference</u>
pBR322	cloning vector	(14)
pK01, pK04, pK06	<u>galK</u> fusion vectors	(15)
pUC8	<u>lac</u> cloning vector	(16)
pRL5	4.5 Kb <u>HindIII</u> fragment that includes promoter proximal portion of <u>ilvGMEDA</u> operon through <u>ilvM</u> inserted into pBR322	(6)
pRL120	1,102 bp <u>AluI</u> fragment (-459 bp to 643 bp), denoted in Fig. 1, inserted into the <u>HindIII</u> site of pBR322 using <u>HindIII</u> linkers.	This study
pRL137	693 bp <u>HaeIII</u> (-50 bp)- <u>AluI</u> (643 bp) fragment from pRL120 inserted into pK06 using a <u>BamHI</u> linker ligated to the <u>HaeIII</u> site and <u>HindIII</u> linker ligated to the <u>AluI</u> site.	This study
pM0139	1,102 bp <u>HindIII</u> fragment from pRL120 (-459 bp to 643 bp) inserted into the <u>HindIII</u> site of pK06.	This study
pM0158	703 bp <u>BamHI</u> - <u>HindIII</u> fragment from pRL137 inserted into pUC8.	This study
pM0164	409 bp <u>AluI</u> (-459 bp)- <u>HaeIII</u> (-50 bp) fragment from pRL5 inserted using linkers <u>EcoRI</u> (<u>AluI</u>)- <u>BamHI</u> (<u>HaeIII</u>) into pK04.	This study
pM0168	419 bp <u>EcoRI</u> - <u>BamHI</u> fragment from pM0164 inserted into pM0158.	This study
pM0171	741 bp <u>DraI</u> (-98 bp)- <u>AluI</u> (643 bp) fragment from pRL120 inserted using linkers <u>BamHI</u> (<u>DraI</u>)- <u>HindIII</u> (<u>AluI</u>) into pK06.	This study
pM0173	1,112 bp <u>EcoRI</u> - <u>HindIII</u> fragment from pM0168 inserted into pK01. This inserts the 1,102 bp <u>AluI</u> fragment in Fig. 1 with a 10 bp insert (<u>BamHI</u> linker), at the <u>HaeIII</u> site at -50 bp (Fig. 1).	This study
pDM196	pM0173 was cleaved with <u>BamHI</u> and the <u>BamHI</u> site was disrupted by backfilling with the large fragment of <u>E. coli</u> DNA polymerase I. The plasmid was recircularized to yield a 14 bp insertion at -50 bp (Fig. 1).	This study

<u>Plasmid</u>	<u>Description</u>	<u>Reference</u>
pDM197	pM0173 was treated as described for pDM196 except that an <u>XhoI</u> linker was inserted prior to recircularization of the plasmid. This yielded a 22 bp insertion at -50 bp (Fig. 1).	This study
pM0202	324 bp <u>Sau3AI</u> fragment from pM0164 (which corresponds to the fragment from the <u>Sau3AI</u> site at -374 bp to the <u>HaeIII</u> site at -50 bp in Fig 1) inserted into pRL137.	This study
pM0219	pM0173 was cleaved with <u>BamHI</u> and the <u>BamHI</u> site was disrupted by <u>S1</u> nuclease digestion to create a 6 bp insertion at the -50 bp <u>HaeIII</u> site (Fig. 1).	This study
pM0226	191 bp <u>HinfI</u> (-241 bp)- <u>HaeIII</u> (-50 bp) fragment from pM0164 was inserted using <u>BamHI</u> linkers into pRL137 (Fig. 1).	This study
pM0236	101 bp <u>FokI</u> (-151 bp)- <u>HaeIII</u> (-50 bp) fragment from pM0164 was inserted using <u>BamHI</u> linkers into pRL137 (Fig. 1).	This study
pCR268	913 bp <u>EcoRV</u> (-268 bp)- <u>AluI</u> (643 bp) fragment inserted using <u>HindIII</u> linkers into the <u>HindIII</u> site of pK06.	This study

Approximately 5 µg of probe was co-precipitated with either 50 µg of cellular RNA or 25% of the RNA transcribed in vitro from pRL5 by the addition of ethanol. The pellet was dried and dissolved in 50 µl of hybridization buffer (100 mM Tris-HCl, pH 8.0, 300 mM NaCl and 2.5 mM EDTA). Hybridization mixtures were sealed in microcapillary pipettes and incubated 16 hrs at 65°C. The nucleic acids were transferred to a 1.5 ml microcentrifuge tube and digested with 50 units of S1 nuclease for one hour at 20°C in 30 mM NaOAc, pH 4.5, 125 mM NaCl, 1 mM ZnSO₄, 5% glycerol, 10 µg/ml sonicated calf thymus DNA. The digestion was terminated by precipitation with ethanol. The nucleic acids were dissolved in 20 µl of sequencing dye (21), fractionated on an 8.3 M urea, 20% polyacrylamide gel, and the S1 digestion products were visualized by autoradiography.

Galactokinase Assays and Relative Plasmid Copy Number Determinations.

Galactokinase was assayed by a method similar to that previously described

Figure 1. Partial restriction site map and DNA sequence of the DNA from -459 bp to 643 bp relative to the initiating nucleotide (designated as +1) for transcription *in vivo* for the *ilvGMEDA* operon. The two promoters *ilvGp1* and *ilvGp2* are designated as p1 and p2 and the attenuator is designated as t. The restriction sites are designated as follows: A, *AluI*; B, *BstNI*; D, *DraI*; F, *FokI*; He, *HaeIII*; Hf, *HinfI*; S, *Sau3AI*; and V, *EcoRV*.

(13,15). A 10 ml bacterial culture (3×10^8 cells per ml) was chilled and collected by centrifugation (3,000 x g for 10 min) and the bacterial pellet was resuspended in 2.0 ml of galactokinase sonication buffer (20 mM Tris-HCl, pH 7.2, 8 mM $MgCl_2$ and 2.2 mM dithiothreitol). The sample was separated into 1.0 ml aliquots in each of two 1.5 ml microcentrifuge tubes. The contents of one tube was collected by centrifugation (15,000 x g for 10 min) for subsequent determination of plasmid copy number as described below. The second 1.0 ml aliquot was sonically disrupted using three 5 sec pulses at a microtip setting of 2 (Heat Systems-Ultrasonics, Inc., W-220F Sonicator). The supernatant was separated from cell debris by centrifugation (15,000 x g for 10 min) and then transferred to a 1.5 ml microcentrifuge tube. Each sample, for the determination of galactokinase, was appropriately diluted with sonication buffer that contained 0.1 mg/ml of bovine serum albumin. Each 50 μ l reaction for the determination of galactokinase was prepared from stock solutions as described (13,15) and contained 100 mM Tris-HCl pH 7.2, 4 mM $MgCl_2$, 5 mM ATP, pH 7.0, 1 mM dithiothreitol, 3.2 mM NaF and 1.6 mM [$1-^{14}C$]galactose at 2 μ Ci/ μ mole. The reaction was initiated by the addition of extract and 10 μ l aliquots were removed at 5, 10 and 15 min. The aliquots were spotted onto 2.4 cm Whatman DE81 filters and plunged immediately into 1 l of ice cold distilled water. The

filters were washed 5X with 20 ml distilled water per filter, 2X with 95% ethanol, dried, and quantitated by liquid scintillation counting. The level of galactokinase attributed to each plasmid is an average of two or more assays. As a control, a culture of M152 containing pRL137 was assayed with each set of cultures. The value indicated for pRL137 is an average of 10 separate experiments and overall individual assays for each plasmid were within 20% of the averages indicated in Fig. 3.

Plasmid copy number was determined by a method similar to that described by Haughn et al. (9). Plasmid DNA was isolated from each sample by the method described by Davis et al. (19) except that each sample was treated with RNase A and subsequently extracted with a mixture of phenol:chloroform:isoamyl alcohol (25:24:1 by volume) prior to precipitation with ethanol. The resulting nucleic acid pellet was resuspended in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) and an appropriate aliquot was digested with PstI. The plasmids and PstI digested pM0139 (which served as a standard), were fractionated on an 0.8% agarose gel. The gel was stained with ethidium bromide and photographed while illuminated with an UV transilluminator. The photograph was scanned using a Helena Laboratories Densitometer and the peaks were quantitated by weighing the paper beneath each peak. The number of moles of plasmid were determined by comparison of each plasmid to the standard present on each gel.

Transcription of ilv-galK fusions in vitro.

Transcription in vitro was performed in a 20 μ l volume as described by Lawther et al. (6). Unlabelled nucleoside triphosphates were present at a final concentration of 150 μ M and [α - 32 P]UTP was at 50 μ M at 5 mCi/ μ mole. Each reaction contained 1 μ g of plasmid (based upon the absorbance at 260 nm) and a 10 fold molar excess of RNA polymerase. The template plasmid, RNA polymerase, and buffer (20 mM Tris-Acetate pH 7.9, 100 mM KCl, 4 mM Mg(OAc) $_2$, 0.1 mM EDTA and 20 mM β mercaptoethanol) were incubated at 37°C for 5 min and 0.5 μ l of 4 mg/ml heparin added. After an additional 5 min at 37°C, the transcription mixture was transferred to a second tube at 37°C containing the nucleoside triphosphates. The reaction was terminated after 15 min as described (6). The products of transcription were fractionated on a 7 M urea, 6% polyacrylamide gel and the transcripts were visualized by autoradiography. To quantitate the products of transcription, the autoradiograph was scanned using a Beckman DU-8 Spectrophotometer and the peak areas were converted to cpm by comparison to a known standard. The moles of transcripts were calculated by normalizing the cpm for the percentage of UMP present in each transcript.

Figure 2. Autoradiograph of the products of S1 analysis of *ilv* specific transcription *in vitro* and *in vivo*. RNA isolated from M152 (lane 1) or RNA prepared by transcription *in vitro* of pRL5 (lane 2) was hybridized with 416 bp BstN1 restriction fragment (-355 bp to 61 bp, Figure 1). The probe was 5' end-labelled with [γ^{32} P]ATP (using T4 polynucleotide kinase) and treated with

exonuclease III to generate single stranded probe (23) prior to hybridization. The resulting DNA:RNA hybrids were digested with S1 nuclease and the products fractionated on an 8.3 M urea/8% polyacrylamide gel. Lanes G and A are the products of G and AG chemical sequencing reactions (21) of the BstNI restriction fragment.

RESULTS

Analysis of Transcription from *ilvGp1* in vivo.

Transcription of the *ilvGMDA* regulatory region in vitro yields both a 258 and a 186 nt RNA (2,6), transcribed from the two initiation sites *ilvGp1* and *ilvGp2* (5). Analysis of cellular RNA by fingerprint analysis demonstrated transcription from *ilvGp2* in vivo, while no transcription could be detected from *ilvGp1* (5). To further assess the role of the *ilvGp1* as a transcription start site in vivo, two additional assays were used to analyze transcription from *ilvGp1*.

Figure 2 presents the results of an S1 nuclease analysis designed to detect transcription from the *ilv* regulatory region. A 416 bp BstNI restriction fragment (from -355 bp to 61 bp, Fig. 1) was 5' labelled with $^{32}\text{PO}_4$ and converted to single-stranded probe by digestion with exonuclease III (23). The DNA probe was hybridized with either RNA extracted from M152 (grown in M63 minimal media) or with RNA prepared by transcription of pRL5 in vitro. After treatment with S1 nuclease, the products were fractionated on an 8.3 M urea, 20% polyacrylamide gel with the products of the Maxam and Gilbert G and A+G sequencing reactions (Fig. 2, lanes G and A respectively) of the BstNI restriction fragment serving as a size standard. The size standard and products of S1 nuclease digestion were visualized by autoradiography. Lane 1 presents the products of S1 nuclease analysis of RNA extracted from M152. A major band appears at approximately 61 nt. This corresponds to the initiation site determined for *ilvGp2* by RNA fingerprint analysis (5). A similar band appears at 61 nt in lane 2 containing the products of S1 analysis of the RNA obtained by transcription of pRL5 in vitro. Transcription from *ilvGp1* should yield an S1 product of 133 nt. As can be seen in Fig. 2, lane 1, no 133 nt product is observed when the probe is hybridized with the cellular RNA, whereas hybridization with the RNA transcribed from pRL5 in vitro results in a 133 nt product (lane 2). As with the RNA fingerprint analysis, S1 nuclease treatment fails to detect transcription in vivo from *ilvGp1*.

As a further assay for transcription in vivo from *ilvGp1*, the 409 bp AluI-HaeIII restriction fragment (Fig. 1, -459 to -50 bp) was inserted into the plasmid pK04 of the galactokinase expression vector system (15). Since the initiating nucleotide for *ilvGp1* is at -72 bp (Fig. 1), this fragment extends

Figure 3. Deletion analysis of the role of the upstream DNA sequences on expression of *ilv-galK* fusions. Designations are the same as indicated for Figure 1. Distances are in bp from the initiating nucleotide *in vivo*. The promoters *ilvGp1* and *ilvGp2* are designated as p1 and p2 respectively, while the attenuator is indicated by the t. The restriction sites upstream of the transcription start site are: A, AluI (-643 bp); S, Sau3AI (-374 bp); V, EcoRV (-268 bp); Hf, HinfI (-241 bp); F, FokI (-151 bp); D, DraI (-98 bp); and He, HaeIII (-50 bp). The open bars represent the inserted BamHI linker.

387 bp 5' and 22 bp 3' relative to the *ilvGp1* initiating nucleotide (Fig. 1), and based upon promoter consensus analysis (24-26) and transcription *in vitro* contains all the DNA sequence elements necessary for transcription from *ilvGp1*. The specific activity for galactokinase of M152 containing pK04 is 1.0 nmol/min/mg protein/fmol of plasmid, while the presence of pM0164 yields a galactokinase activity of 0.25. Thus the *ilvGp1* plasmid pM0164 results in about 25% of the level of galactokinase as the parental vector. Two further observations are consistent with limited transcription from *ilvGp1* *in vivo*. First, M152 containing pM0164 yields white colonies on MacConkey Agar containing galactose. This indicates either very little or no expression of galactokinase from pM0164. Second, M152 containing pM0164 fails to grow on minimal medium with galactose as the sole carbon source, while M152 containing the parental plasmid pK04 does grow slowly on minimal medium containing galactose.

Role of Upstream DNA Sequences in Expression of the *ilv* Promoter.

Figure 3 presents an analysis of the effect of DNA sequences upstream of *ilvGp2* (from -459 to -50 bp) on the expression of galactokinase fused to the *ilv* regulatory region. All plasmids are constructed to extend 643 bp downstream of the transcription initiation site *in vivo*; as our preliminary observations indicate that altering the downstream fusion site alters the level of galactokinase (K. Bell and R. P. Lawther, unpublished observations). The specific activity of galactokinase from each plasmid fusion has been normalized for relative plasmid copy number. The plasmid pMO139 includes DNA sequences from -459 to 643 bp (relative to the site of initiation of transcription *in vivo*) fused to the *galK* gene. This construct yields approximately 58 units of galactokinase. The plasmids pCR268, pMO171 and pRL137 illustrate the effect of the DNA sequences from -450 to -50 bp on expression from the *ilv* regulatory region. The plasmid pCR268 extends to the *EcoRV* restriction site at -286 bp and results in a specific activity of 93 units of galactokinase. Thus, deletion of the region upstream of the *EcoRV* site results in a 60% increase in the level of galactokinase relative to pMO139. This indicates that the deleted DNA sequences (removed by 280 bp from the site of transcription initiation) limit the expression of galactokinase. This observation is supported by experiments described subsequently. The further deletion of DNA to the *DraI* restriction site at -98 bp (pMO171) results in 13 units of galactokinase while deletion to the *HaeIII* site at -50 bp (pRL137) results in slightly less than 8 units of galactokinase. Thus, the deletion of DNA sequences between -280 and -50 bp dramatically decreases the expression of galactokinase from the *ilv* regulatory region. If the DNA sequences used to construct pRL137 (to -50 bp) defines the promoter *ilvGp2* (based upon consensus DNA sequence analysis [24-26], and our studies of transcription *in vitro*), then the upstream region can enhance expression from *ilvGp2* as much as 13-fold.

To further refine our analysis of the upstream region, additional plasmids were constructed. The previously described plasmids were constructed using restriction sites that occur uniquely 5' to *ilvGp2* in the *ilv* regulatory region (Figs. 1 & 3). Other restriction sites exist within the upstream region but homologous restriction sites also exist downstream within 645 bp of *ilvGp2*. To conserve the insert/plasmid juncture at 645 bp, the plasmids were constructed as derivatives of pRL137. These derivatives (pMO202, pMO226 and pMO236) were constructed by using *Bam*HI linkers to insert restriction fragments into the unique *Bam*HI site (-50 bp relative to *ilvGp2*) present in pRL137 (Table 1).

Because of the manner of their construction, each of these plasmids contains the 10 bp BamHI linker (CCGGATCCGG) inserted at the HaeIII restriction site at -50 bp. As a control, plasmid pM0173 was constructed identical to pM0139 except that it has the 10 bp linker inserted at the HaeIII site (-50 bp). As can be seen in Figure 3, the presence of pM0173 results in 37 units of galactokinase. Insertion of the BamHI linker resulted in a decrease in the observed specific activity of galactokinase relative to pM0139 (no insert). Thus, the subsequent plasmids need to be compared to pM0173 and pRL137. The next three plasmids, pM0202, pM0226, and pM0236 are deleted to the Sau3AI site, -374 bp; the HinFI site, -241 bp; and the FokI site, -151 bp, respectively (Figure 3). Two observations can be made with these three plasmids. First, they demonstrate that the DNA sequences that enhance expression from the ilv promoter lie within 150 bp of the site of initiation of transcription *in vivo* (compare to pRL137). Second, these three plasmids show an increase in expression over pM0173 similar to that observed with pCR268 relative to pM0139. The observed increase indicates that the sequences upstream of -375 bp in some manner limit expression of galactokinase from the ilv regulatory region. Thus, the specific activities of galactokinase observed with both sets of plasmids indicate that two DNA elements affect transcription from the ilv promoter. One element with sequences between -150 and -50 bp enhances expression of galactokinase from the ilv promoter, while the second (residing more than 375 bp upstream) can limit the extent of the enhancement.

The Effect of Distance and Turns of the DNA helix on the Enhancement Of Expression from ilvGp2 Promoter by Upstream DNA Sequences.

As noted above, the insertion of the 10 bp BamHI linker at -50 bp reduced the expression of galactokinase by 35% (pM0139 compared with pM0173). Dunn et al. (27) observed that for the araBAD operon, both the distance between regulatory sites and turns of the DNA helix were important. To investigate the effect of both distance and helical turns on the effect of the upstream DNA sequences on ilvGp2, the plasmid pM0173 was modified by deletion and insertion (Fig. 4).

To facilitate comparison, the specific activity of the parental plasmid with no insert, (pM0139) and the plasmid with a 10 bp insert (pM0173), are repeated from Figure 3. As described, pM0173 contains the same E. coli DNA sequences as pM0139 except for the insertion of a 10 bp BamHI linker in the HaeIII site at -50 bp. The plasmid pDM197 contains a 22 bp insert that was created by modification of pM0173 (Table 1). The plasmids pM0173 and pDM197 demonstrate the effect of inserting either one or approximately two integral

Figure 4. Effect of insertion of integral and nonintegral turns of DNA helix at -50 bp on expression of *ilvGp2*.

turns of DNA at -50 bp. As the distance between *ilvGp2* and the DNA sequences upstream of -50 bp increases, the expression of galactokinase decreases from 57 units in pM0139 to 37 units in pM0173 to 16 units in pDM197, indicating the importance of distance on the effect of the upstream DNA sequences on *ilvGp2* (Fig. 4). To investigate the role of the helical turns on the interaction of the upstream DNA sequences with *ilvGp2* the plasmids pM0219 and pDM196 were constructed by modification of pM0173 (Table 1). As can be seen in Figure 4, either a 6 or 14 bp insert reduces the specific activity of galactokinase to that observed with pRL137, i.e. deletion of the entire upstream region. Therefore the enhancement of expression from *ilvGp2* by the DNA sequences upstream of -50 bp appears to be influenced by two topological constraints: the distance between *ilvGp2* and the upstream sequences, and their relative angular displacement on the DNA helix.

Transcription of Galactokinase Fusions in vitro.

Bossi and Smith (28) observed that a DNA segment upstream of the *hisR* gene affected expression from the *hisR* promoter. They found that a 3 bp deletion

Figure 5. Autoradiograph of products of transcription *in vitro* of *ilv-galK* fusion plasmids. The plasmids were transcribed as described in the Materials and Methods. The transcription products were fractionated on a 7M urea, 6% polyacrylamide gel. The size of the relevant transcripts are indicated in number of nucleotides.

(*hisR1223*) 70 bp upstream reduced expression from the *hisR* promoter both *in vivo* and *in vitro*. Thus it was concluded that the deletion resulted in a structural alteration of the DNA that reduced its efficiency as a promoter. In an effort to evaluate the mechanism by which the upstream DNA sequences affect transcription from *ilvGp2*, several of the *ilv-galK* fusion plasmids were transcribed *in vitro*. If the observed effect of the DNA sequences upstream of

TABLE 2. Relative Quantities of RNAs Transcribed in vitro from the ilv-galk Fusions

<u>Plasmid</u>	<u>186/108</u>	<u>186/258</u>	<u>Basepairs between the initiating nucleotides for <u>ilvGp1</u> and <u>ilvGp2</u></u>
pRL5	0.96	2.11	72
pRL137	1.31	---	---
pM0171	1.00	---	---
pM0139	1.20	2.10	72
pM0173	1.35	1.55	82
pDM196	1.02	1.27	86
pDM197	1.04	1.02	94

ilvGp2 are the result of either the DNA sequences alone or the interaction of RNA polymerase with ilvGp1, then altering the structure of the DNA upstream of ilvGp2 should affect transcription of ilvGp2 in vitro. One would expect transcription in vitro to mimic the observations concerning galactokinase expression in vivo. That is, the level of expression from ilvGp2 should be greater for plasmids that contain ilvGp1p2 than for plasmids with only ilvGp2.

An autoradiograph of the products of transcription in vitro is presented in Figure 5. All the plasmids were transcribed under identical conditions at a ratio of approximately 10 RNA polymerase molecules per plasmid molecule with 200 µg of heparin added to prevent reinitiation of transcription. Lanes 1 and 2 of Figure 5 are the control plasmids pRL5 and pK04. As previously described (2,6), transcription of the plasmid pRL5 yields three discrete transcripts of 108, 186 and 258 nt. These correspond to the ColE1 replication RNA (108 nt) and the leader RNAs from the ilv regulatory region (ilvGp2, 186 nt, and ilvGp1, 258 nt). The plasmid pK04 yields the 108 nt ColE1 replication RNA. Transcription of pRL137 and pM0171 in vitro, which are deleted for ilvGp1, yield the expected RNAs of 108 and 186 nt. Transcription of pM0139 yields the same three RNAs as pRL5. The plasmids pM0173, pDM196 and pDM197 contain insertions between ilvGp1 and ilvGp2 of 10 bp, 14 bp and 22 bp, respectively. As a result transcription of these plasmids yields longer transcripts from ilvGp1. All the plasmids that contain ilvGp1, including those which result in reduced expression of galactokinase in vivo (relative to pM0139), are still transcribed from ilvGp1. The data reveal no correlation between the level of expression of

galactokinase in vivo and transcription from ilvGp1 in vitro.

A comparison of the relative quantities of the transcription products of the plasmids in Figure 5 is presented in Table 2. The transcripts were quantitated by densitometric scanning and the 108 nt replication transcript served as an internal standard for comparison of the transcription reactions. Several observations can be made from Table 2. First, no systematic change in the ratio of the 186 nt RNA (transcribed from ilvGp2) to the 108 nt RNA is observed, i.e., the ratio remains relatively constant. Second, the data indicate that the ratio of transcription from ilvGp2:ilvGp1 (186:258 nt RNA) is about 2:1 in both pRL5 and pMO139 when ilvGp1 and ilvGp2 are separated by 72 bp (the wild type juxtaposition). This agrees with published observations (5). Third, as the distance between ilvGp1 and ilvGp2 is increased to 94 bp (pDM197, 22 bp insert) the ratio of the 186 to the 258 nt RNA becomes equimolar. These observations are inconsistent with the enhanced expression from ilvGp2 being due to the presence of the upstream DNA sequences or the interaction of RNA polymerase with ilvGp1. First, since the ratio of the 186:108 nt RNA remains unchanged, the presence of the upstream DNA sequences that include ilvGp1 do not directly of themselves enhance transcription from ilvGp2 (ratio of 186/108 nt RNAs for pRL137 with that for pRL5 and pMO139). Second, the 2:1 molar ratio of the RNAs from ilvGp1 and ilvGp2 indicates that all plasmids cannot have RNA polymerase bound to both of the promoters simultaneously.

DISCUSSION

Transcription of the regulatory region of the ilvGMEDA operon has been examined previously both in vitro (2,5,6) and in vivo (5). Two promoters, ilvGp1, ilvGp2, were identified in vitro, but transcription could be detected only in vivo from the promoter ilvGp2, closest to the structural genes (5). In this study the role of ilvGp1 as an initiation site of transcription in vivo has been further scrutinized. Our analysis of cellular RNA using S1 nuclease does not detect any transcription from ilvGp1 in vivo. Furthermore, the fusion of ilvGp1 to galK also fails to detect transcription in vivo from ilvGp1. The failure to observe RNA initiation from ilvGp1 by either fingerprinting or S1-nuclease could be explained by the 5' terminus of the ilvGp1 transcript being processed too rapidly to be detected. The absence of expression of galK when fused to ilvGp1 argues against this possibility. Thus, three types of analysis [RNA fingerprinting (5); S1 nuclease; and promoter-vector fusion] each fail to detect transcription from ilvGp1. These analyses do not eliminate the possibility that ilvGp1 functions during a specific set of physiological

conditions. However, at least under the conditions assayed by Adams et al. (5) and ourselves, the simplest interpretation is that ilvGp1 does not serve as a site for the initiation of transcription in vivo.

Adams and Hatfield (13) observed that the DNA sequences upstream of ilvGp2 (that include ilvGp1) cause approximately an 11 fold increase in the expression of galK from ilvGp2. Our studies confirm their observation and more precisely define the limits of the upstream stimulating element. The data indicate that deletion of DNA sequences between approximately 150 and 100 bp upstream of the site of transcription initiation in vivo reduces expression from ilvGp2. If the plasmid pRL137 (extending to -50 bp) is taken to designate a minimal length for ilvGp2 (based upon both transcription in vitro and the consensus analysis of E. coli promoters, ref. 24-26) then the presence of the upstream DNA sequences can result in a 13 fold enhancement (pCR268 versus pRL137) of expression from ilvGp2. Our analysis of the region between -150 and -50 bp indicates that increasing the distance between ilvGp2 and the upstream DNA sequences reduces the expression of galK fused to the ilvGp2. Also, the insertion analysis indicates that altering the relative helical orientation of the upstream DNA sequences and ilvGp2 to each other is equivalent to deleting the upstream DNA sequences. An unexpected aspect of our results is that DNA sequences between -459 and -375 bp limit the enhancement of expression from ilvGp2. Because of how far these sequences are displaced from the promoter, it is unclear whether these sequences have a direct role in expression from ilvGp2, or how these sequences alter expression from ilvGp2. The observed effect may not reflect a regulatory role for these sequences. The limitation of expression may be due to an alteration of plasmid structure, since transcription in vitro indicates that both ilvGp1 and ilvGp2 are affected by supercoiling (unpublished observation R. P. Lawther).

Transcription in vitro indicates that enhancement of expression from ilvGp2 is not simply due to the presence of the upstream DNA sequences or the interaction of RNA polymerase with ilvGp1. Particularly interesting is the observation that the ratio of transcripts from ilvGp2:ilvGp1 (Table 2, 186/258) is 2:1. This observation may reflect either of two possibilities: (1) RNA polymerase interacts simultaneously with ilvGp1 on only one half the plasmids on which RNA polymerase is interacting with ilvGp2, or (2) on any single plasmid molecule, RNA polymerase can only interact with either ilvGp1 or ilvGp2 and the relative levels of transcripts reflect how effectively each promoter is transcribed by RNA polymerase. At present these alternatives remain unresolved. Although, consistent with the latter hypothesis is the observation

that increasing the distance between ilvGp1 and ilvGp2 results in the ratio of transcripts approaching 1:1. This data seems to imply some form of steric hindrance for the binding of RNA polymerase to both promoters. Two additional observations support the latter hypothesis. First, the sum of the moles of transcripts from both ilvGp1 and ilvGp2 is approximately equal to the number of moles of plasmid present for both pRL5 and pM0139. Second, either increasing or decreasing the ratio of RNA polymerase to plasmid by two fold does not alter the observed transcription pattern. Together these data seem to make it unlikely that RNA polymerase can simultaneously bind to ilvGp1 and ilvGp2 on the same plasmid molecule. If this conclusion is correct, it is unlikely that the interaction of RNA polymerase with ilvGp1 enhances transcription from ilvGp2.

Our deletion and insertion analyses indicate that the DNA sequences that reside upstream of ilvGp2 enhance expression from this promoter in vivo. The mechanism by which these sequences influence expression from the promoter remains to be elucidated. Two obvious questions arise from this observation. First, does the enhanced expression from the ilv promoter require a protein? Possibly, the DNA sequences (between -150 and -50 bp) modify the structure of the DNA in a manner that increases the affinity of RNA polymerase for the promoter. Alternatively these DNA sequences could be the binding site for a protein(s) that activates transcription from the ilv promoter. Second, if a protein is required, what is the identity of that protein? For instance, do the previous observations concerning transcription in vitro by RNA polymerase from ilvGp1 identify an upstream site at which polymerase binds in vivo to activate transcription from ilvGp2? Alternatively, is the DNA structure characterized in vitro as ilvGp1 fortuitous and merely overlap the binding site for another protein(s)?

Our analysis of transcription in vitro is not consistent with the enhancement of expression being simply due to either the upstream DNA sequences alone or their interacting with RNA polymerase. Deletion or displacement of the upstream sequences has no effect upon transcription from ilvGp2 in vitro. Furthermore, our results clearly indicate that ilvGp1 does not appear to serve as a site for the initiation of transcription in vivo. Nevertheless, ilvGp1 functions effectively as a promoter in vitro. These results differ from those observed for ilvEp, a weak internal promoter in the ilvGMEDA operon (29,30). Expression from ilvEp in vivo can be easily measured, yet it is a very poor promoter in vitro. In fact, when transcribed in vitro, ilvEp is an order of magnitude less effective than ilvGp1 (J.M. Lopes personal communication). We

propose that the effect of the upstream DNA sequence (between -150 and -100 bp) on expression from ilvGp2 is mediated by a protein, but that protein is probably not RNA polymerase. By this hypothesis, ilvGp1 is not available for interaction with RNA polymerase *in vivo* because the hypothetical protein(s) binds to DNA sequences that overlap ilvGp1.

Several additional observations are consistent with the above interpretation. Insertion of the 10 bp BamHI linker at the DraI site (-98 bp, Fig. 1) results in the upstream DNA sequences having no effect on the expression from ilvGp2 (i.e., the level of expression is identical to that observed with pRL137). This linker insertion experiment is also consistent with a protein interacting with the upstream DNA region since such an insert would disrupt the proposed protein binding site. The insertion of either 6 or 14 bp at -50 bp results in the loss of any enhancement of transcription from the ilv promoter (Fig. 4). Similar effects of the insertion of less than integral turns of the DNA helix have been observed for several systems known to be regulated by proteins (27,31). Both the araC protein (27) and the λ repressor (31) interact cooperatively at separate DNA sites. The insertion of less than integral turns of the DNA helix between these sites results in the loss of cooperativity between these sites. Also, our initial observations suggest that random mutagenesis of M152 with either Tn5 or Tn10 yields mutant strains that allow expression of galK from ilvGp1 (D. Shrontz and R.P. Lawther, unpublished observations). These mutant strains were obtained by randomly mutagenizing M152 with Tn5 or Tn10 and subsequently transfecting with pM0164.

Consistent with the role of the DNA sequences between -150 and -100 bp is the comparison of the E. coli K-12 DNA sequence to the corresponding region from Salmonella typhimurium (Fig. 6). The DNA sequences between -47 and +20 bp are nearly identical (32). A second region of homology exists between -126 bp and -90 bp, while an upstream element equivalent to ilvGp1 does not exist in S. typhimurium (32). However, galK fusion plasmids that contain the ilvGMEDA regulatory region from S. typhimurium show a pattern of enhanced expression similar to that observed for the ilv regulatory region from E. coli K-12 (R. P. Lawther, unpublished observations). An interesting aspect of this region is the high percentage of AT base pairs conserved between E. coli and S. typhimurium (approximate 80% AT from -130 to -90 bp). Such AT rich regions have been implicated in enhancing E. coli K-12 promoter activity (33-35), although similar sequences upstream of the leu operon appear not to enhance expression from the leu promoter (36). That these sequences do not serve as general enhancer is indicated by the observation that insertion of the DNA

Figure 6. DNA sequence comparison of -160 bp to 20 bp of the *ilvGMEDA* operons of *E. coli* K-12 and *S. typhimurium*. The upper strand is the *E. coli* K-12 DNA sequence and the lower strand indicates the nucleotides that are not identical in *S. typhimurium*.

sequence upstream of *ilvGp2* (459 to -50 bp) does not enhance expression from *ptet*. Thus, the enhancement of expression by the upstream DNA sequences may be specific for the *ilvGMEDA* operon.

Our observations indicating that full expression from *ilvGp2* requires the presence of upstream DNA sequences is similar to that described for the *tyrT* (12), *ilvIH* (9) *nrd* (37) and *glnA2* (38) promoters. For each of these promoters some protein is believed to interact with the upstream DNA sequences to stimulate transcription from these promoters. The identity of the protein(s) presumed to bind to the DNA sequences upstream of the *ilv* promoter is unestablished. Recently, a number of genes have been implicated as possible regulatory effectors of *ilv* gene expression. Observations by Friden et al. (39) and Friedman et al. (40) indicate that the integration host factor could be involved in *ilvGMEDA* operon expression. Johnson and Sommerville (41,42) have isolated mutations in two genes, *ileR* and *ilvR*. The *ileR* gene product is presumably a negative regulatory protein that regulates both the *thr* and *ilvGMEDA* operons, while the *ilvR* gene product is presumably a positive regulatory protein for the *ilvGMEDA* operon. Whether any of these are involved in the enhancement of transcription by the upstream DNA sequences indicated here is currently unestablished but is being pursued by biochemical and genetic analysis. The availability of techniques (43,44) that allow the detection of the *in vivo* binding of proteins to specific DNA segments makes possible the examination of the role of protein factors in *ilvGp2* expression, i.e., whether RNAP (43) or other proteins bind to the DNA sequences upstream of *ilvGp2*.

ACKNOWLEDGEMENTS

We thank E. Aubrey Thompson and Richard Musso for assistance with this manuscript. Also, we are grateful to Carol Moline and Bruce Taillon for technical assistance and Debra Williams for secretarial services. This work was supported in part by grant GM28021 from the National Institutes of Health.

REFERENCES

1. Wek, R.C., Hauser, C.A. and Hatfield, G.W. (1985) *Nucl. Acids Res.* **13**, 3995-4010.
2. Lawther, R.P. and Hatfield, G.W. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1862-1866.
3. Nargang, F.E., Subrahmanyam, C.S. and Umbarger, H.E. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1823-1827.
4. Berg, C.M., Shaw, K.J., Vender, J. and Borucka-Mankiewicz, M. (1979) *Genetics* **93**, 309-319.
5. Adams, C.W., Rosenberg, M. and Hatfield, G.W. (1985) *J. Biol. Chem.* **160**, 8538-8544.
6. Lawther, R.P., Nichols, B., Zurawski, G. and Hatfield, G.W. (1979) *Nucl. Acids Res.* **7**, 2289-2301.
7. Bachmann, B.J. and Low, K.B. (1980) *Microbiol. Rev.* **44**, 1-56.
8. Travers, A.A., Lamond, A.I., Mace, H.A.F. and Berman, M.L. (1983) *Cell* **35**, 265-273.
9. Haughn, G.W., Squires, C.H., DeFelice, M., Largo, C.T. and Calvo, J.M. (1985) *J. Bacteriol.* **163**, 186-198.
10. Altman, S. and Smith, J.D. (1971) *Nature* **233**, 35-39.
11. Kupper, H., Contreras, R., Landy, A. and Khovano, H.G. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4754-4758.
12. Lamond, A.I. and Travers, A. (1983) *Nature (London)* **305**, 248-250.
13. Adams, C.W. and Hatfield, G.W. (1984) *J. Biol. Chem.* **254**, 7399-7403.
14. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynekes, H.L., Boyer, H.L., Crosa, J.H. and Falkow, S. (1977) *Gene* **2**, 95-113.
15. McKenney, K.H., Shimatake, H., Court, D., Schmeissner, U., and Rosenberg, M. (1981) in *Gene Amplification and Analysis*, Chirkkijian, J. and Papos, T. Eds. Vol. II pp. 383-415, Elsevier-North Holland Publishing Co., New York.
16. Vieira, J. and Messing, J. (1982) *Gene* **19**, 259-268.
17. Miller, J.H. (1972) in *Experiments in Molecular Genetics*, Cold Spring Harbor Laboratory.
18. Driver, R.P. and Lawther, R.P. (1985) *J. Bacteriol.* **147**, 452-462.
19. Davis, R.W., Botstein, D. and Roth, J.R. (1980) in *Advanced Bacteriol. Genetics*, Cold Spring Harbor Laboratory.
20. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in *Molecular Cloning*, Cold Spring Harbor Laboratory.
21. Maxam, A. and Gilbert, W. (1980) in *Methods of Enzymology* **65**, 499-560.
22. Jones, H.M., Brajkovich, C.M. and Gunsalus, R.P. (1983) *J. Bacteriol.* **155**, 1279-1287.
23. Miller, K.G. and Sollner-Webb, B. (1981) *Cell* **27**, 165-174.
24. Rosenberg, M. and Court, D. (1979) *Ann. Rev. Genet.* **13**, 319-353.
25. Sienbenlist, U., Simpson, R.B. and Gilbert, W. (1980) *Cell* **20**, 269-281.
26. Hawley, D.K. and McClure, W.R. (1983) *Nucl. Acids Res.* **11**, 2237-2255.
27. Dunn, T.M., Hahn, S., Ogden, S. and Schleif, R. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5017-5020.
28. Bossi, L. and Smith, D.M. (1984) *Cell* **39**, 643-652.
29. Wek, R.C. and Hatfield, G.W. (1986) *Nucl. Acids Res.* **14**, 2763-2777.

30. Lopes, M.J. and Lawther, R.P. (1986) Nucl. Acids Res. 14, 2779-2798.
31. Hochschild, A. and Ptashne, M. (1986) Cell 44, 681-687.
32. Taillon, M.P., Gotto, D.A. and Lawther, R.P. (1981) Nucl. Acids Res. 9, 3419-3432.
33. Horn, G.T. and Wells, R.D. (1981) J. Biol. Chem. 256, 2003-2009.
34. Nakamura, K. and Inouye, M. (1979) Cell 18, 1109-1117.
35. Nishi, T. and Itoh, S. (1986) Gene 44, 29-36.
36. Haughn, G.W., Wessler, S.R., Gemmill, R.M. and Calvo, J.M. (1986) J. Bacteriol. 166, 1113-1117.
37. Tuggle, K.C. and Fuchs, J.A. (1986) EMBO J. 5, 1077-1085.
38. Reitzer, L.J. and Magasanik, B. (1986) Cell 45, 785-792.
39. Friden, P., Voelkel, K., Sternglanz, R. and Freundlich, M. (1984) J. Mol. Biol. 172, 573-579.
40. Friedman, D.I., Olson, E.J., Carver, D. and Gellert, M. (1984) J. Bacteriol. 157, 484-489.
41. Johnson, D.I. and Sommerville, R.L. (1983) J. Bacteriol. 155, 549-555.
42. Johnson, D.I. and Sommerville, R.L. (1984) Mol. Gen. Genet. 195, 70-76.
43. Gilmour, D.S. and Lis, J.T. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 4275-4279.
44. Nick, H. and Gilbert, W. (1985) Nature (London) 313, 795-798.