Proc. Natl. Acad. Sci. USA
Vol. 78, No. 2, pp. 922–925, February 1981
Biochemistry

Molecular basis of valine resistance in *Escherichia coli* K-12

(DNA sequence/frameshift mutation/transcriptional polarity/isoleucine and valine biosynthesis)

ROBERT P. LAWTHER*, DAVID H. CALHOUN†, CRAIG W. ADAMS‡, CRAIG A. HAUSER‡, JOHN GRAY†, AND G. WESLEY HATFIELD‡

*Department of Biology, University of South Carolina, Columbia, South Carolina 29208; †Department of Microbiology, Mt. Sinai School of Medicine, New York, New York; and ‡Department of Microbiology, College of Medicine, University of California, Irvine, California 92717

Communicated by Maxine F. Singer, November 7, 1980

ABSTRACT **The relationship of valine resistance to the expression of the *ilvGEDA* operon of *Escherichia coli* K-12 has been determined. DNA sequence and *in vivo* protein analyses indicate that in wild-type *E. coli* K-12 there is a frameshift site within the gene (*ilvG*) for valine resistance. The *ilvG*$^+$*2096* (formerly designated *ilv02096*) mutation displaces this frameshift site, resulting in the expression of *ilvG* and the relief of transcriptional polarity on the distal genes of this operon. Thus, the "*ilv0*" mutation, which concomitantly confers valine resistance and increased expression of the *ilvEDA* genes, is, in fact, the "reversion" of a polar site within the first structural gene of the *ilvGEDA* operon.**

Nearly a quarter of a century ago Tatum (1) noted that, in the absence of isoleucine, valine inhibits the growth of *Escherichia coli* K-12. This sensitivity to exogenous valine is due to the feedback inhibition of the acetohydroxy acid synthase (AHAS) isozymes, which catalyze the first step in the biosynthesis of both valine and isoleucine (2). The genomes of wild-type *E. coli* strains contain genes for three AHAS isozymes: *ilvB* (AHAS I; ref. 3), *ilvG* (AHAS II; ref. 4), and *ilvHI* (AHAS III; ref. 5). Although *E. coli* K-12 possesses these three genes, it expresses only the AHAS I and AHAS III activities (4). The AHAS I and AHAS III isozymes are both sensitive to feedback inhibition by valine, whereas the AHAS II isozyme (not expressed in *E. coli* K-12) is resistant to this inhibition. Consequently, exogenous valine inhibits the total AHAS activity in *E. coli* K-12, resulting in a starvation for the other end product of the common biosynthetic pathway, isoleucine.

Several types of valine-resistant derivatives of *E. coli* K-12 can be isolated (6). One class of these valine-resistant strains contains a *cis*-dominant mutation, heretofore designated "*ilv0*" (7), which has been genetically mapped between mutations in *ilvG* and *ilvE* (8, 9). This mutation results in the expression of the normally cryptic *ilvG* gene product AHAS II (hence, the valine-resistant phenotype; ref. 4) and, concomitantly, a 5- to 10-fold increase in the expression of the *ilvEDA* genes. Because of the coordinant derepressing effect of these *ilv0* mutations, it was long thought that they defined a regulatory region for the *ilvEDA* genes (7, 10). However, it has been recently shown that *ilvG* and *ilvEDA* are, in fact, contiguous genes of a single operon, *ilvGEDA*, regulated by a promoter-attenuator region, *ilvL*, preceding *ilvG* (11–14).

In order to understand the nature of the *ilv0* site in *E. coli* K-12, we determined the DNA sequence of the *ilvG* gene of wild-type *E. coli* K-12 and of a valine-resistant derivative containing a mutation designated *ilv02096* (9). We find that there is a frameshift site within the *ilvG* gene of wild-type *E. coli* K-12. The *ilv02096* mutation (a two-base insertion event) displaces this site, permitting the expression of the *ilvG* gene product,

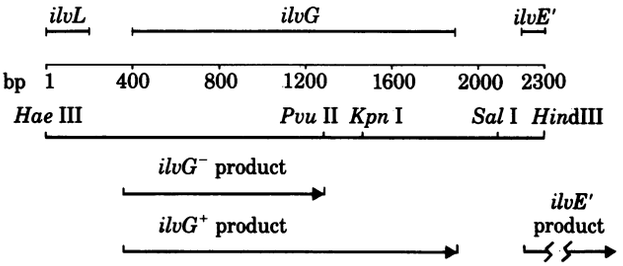

FIG. 1. Schematic representation of *ilvLGE'* region of the *ilvGEDA* operon. The arrows indicate the translation stop sites for the *ilvG*$^-$ and *ilvG*$^+$ gene products according to DNA sequences shown in Figs. 2 and 3. bp, Base pairs.

AHAS II. The concomitant increase in the expression of the distal portion of the *ilvGEDA* operon is therefore presumed to be due to relief of transcriptional polarity.

MATERIALS AND METHODS

Restriction enzymes were obtained from New England BioLabs, were prepared by standard methods, or were a gift of Charles Yanofsky. Plasmids were prepared as described (11). Construction and characterization of plasmid pRL5 have been described (11). pRL5 is a plasmid with a 4.8-kilobase *Hin*dIII fragment from the *ilv* region of wild-type *E. coli* K-12 (*ilvLG*$^-$*E'*) cloned into the single *Hin*dIII site of pBR322. The *ilv02096* and *ilvE2105* mutations were mobilized from strain CU693 (*ilv02096 ilvE2105*; ref. 9) onto λ*dilv58* (8) by homologous recombination to give λ*dilv581*. Plasmids pAH4 and pAH29 were constructed by insertion of the 4.8-kilobase *Hin*dIII restriction fragment from λ*dilv581*, which includes the beginning of the *ilvGEDA* operon (*ilvLGE'*), into the *Hin*dIII restriction site of pBR322. This *Hin*dIII restriction fragment in pAH4 is in the opposite orientation of the analogous *Hin*dIII restriction fragment in pRL5 and pAH29. DNA restriction fragments were prepared and DNA sequence was analyzed as described by Maxam and Gilbert (15). The DNA sequence of both strands of wild-type *E. coli* K-12 DNA was determined from restriction fragments of pRL5 (*ilvG*$^-$) and compared to DNA sequences from pAH4 (*ilvG*$^+$*2096*).

RESULTS

In order to understand the absence of the AHAS II isozyme in wild-type *E. coli* K-12 and its presence in *ilv0*-containing strains,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: AHAS I, II, and III, isozymes of acetohydroxy acid synthase; bp, base pair(s).

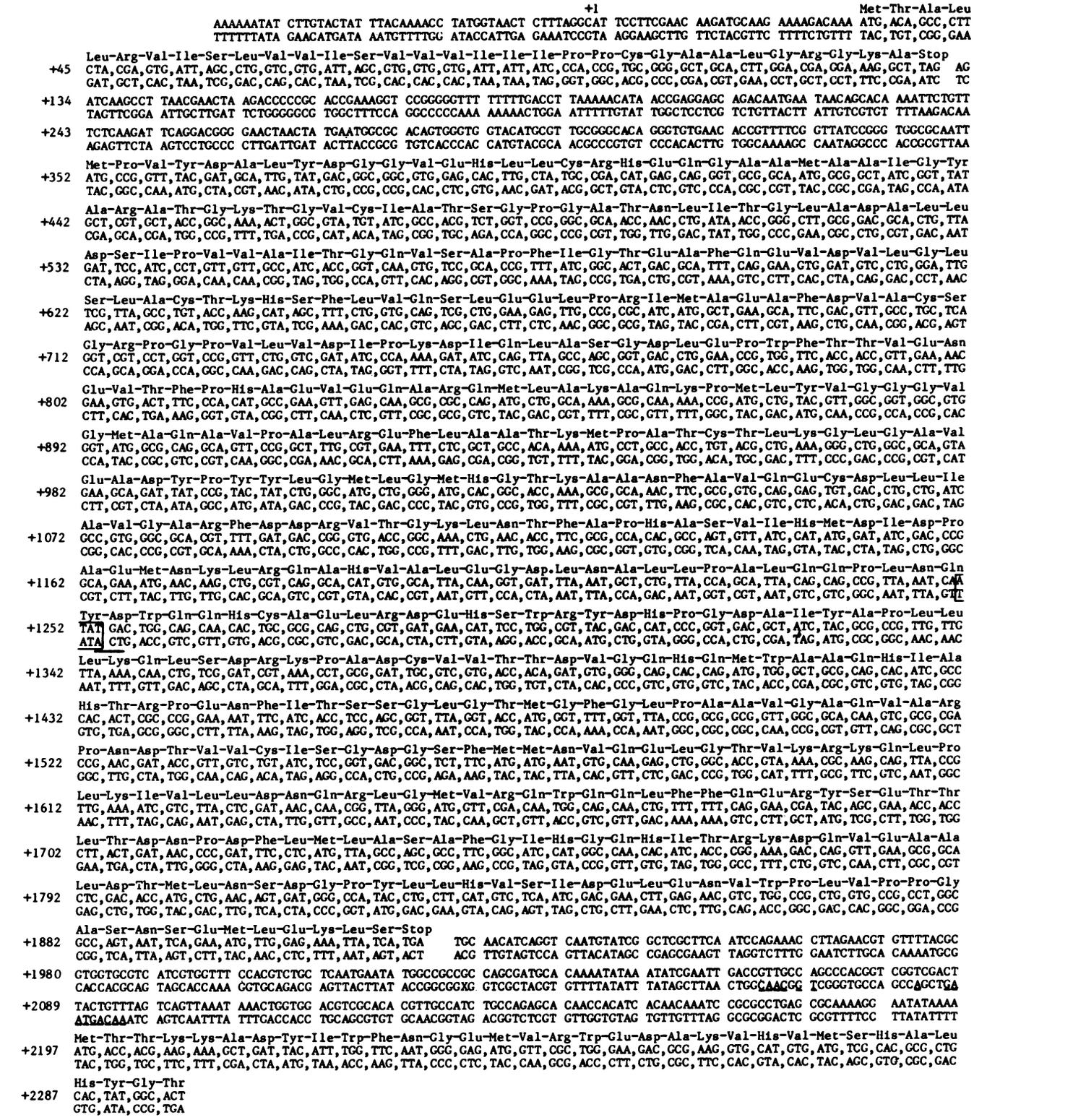

FIG. 2. Nucleotide sequence of the *ilvLGE'* portion of the *ilvGEDA* operon. The sequence was determined from sequence analysis of both strands of wild-type *E. coli* K-12 DNA from pRL5 (*ilvG*$^-$), which is as shown except that the wild-type sequence is missing two nucleotide pairs (either $\frac{AT}{TA}$ or $\frac{TA}{AT}$) in the region from bp +1251 to +1254. The regions of *ilvG*$^+$ (*ilv02096*) DNA sequenced from pAH4 were from bp +1070 to +1355 and +1455 to +2214. These sequences are identical to those shown. bp +1 denotes the point of transcription initiation (12); bp +33 to +128 encodes the leader polypeptide (12); bp +352 to +1914 is the proposed coding region for the *ilvG* gene; bp +2197 is the beginning of the *ilvE* gene. The box around bp +1251 to +1254 denotes the site of the *ilv02096* mutation. bp +1254 to +1256 denote the stop codon for the truncated *ilvG* polypeptide of wild-type *E. coli* K-12. The base pairs underlined between bp +2063 and +2095 represent the proposed internal promoter based on consensus promoter sequences (16).

the DNA sequence of the 2-kilobase *ilvG* region between *ilvL* and the beginning of the *ilvE* gene was determined (Fig. 1). The DNA sequence of the *ilvG*$^+$ (*ilv02096*) gene presented in Fig. 2 differs from the wild-type *E. coli* K-12 sequence by the

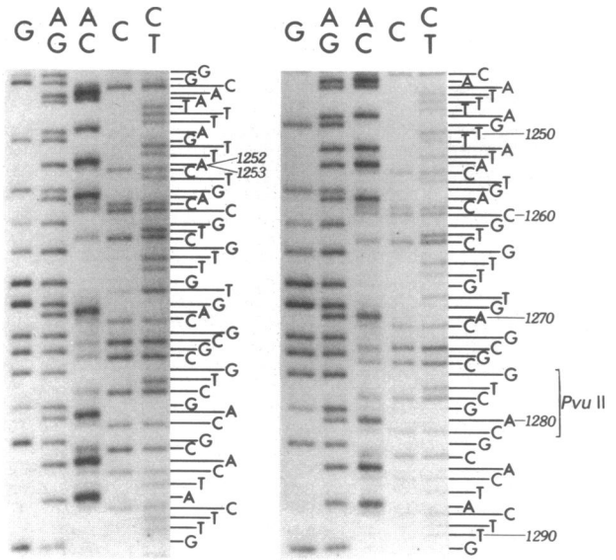

FIG. 3. Autoradiograph of 8% polyacrylamide sequencing gel of Maxam and Gilbert of the *ilvG* gene in the region of the *ilv02096* mutation in wild-type *E. coli* K-12 (*Left*) and the *ilv02096* mutant (*Right*). The sequence is for the lower strand shown in Fig. 2. Two nucleotides absent in the wild-type sequence are indicated (bp 1252 and 1253). The single *Pvu* II endonuclease restriction site in *ilvG* (Fig. 1) is identified.

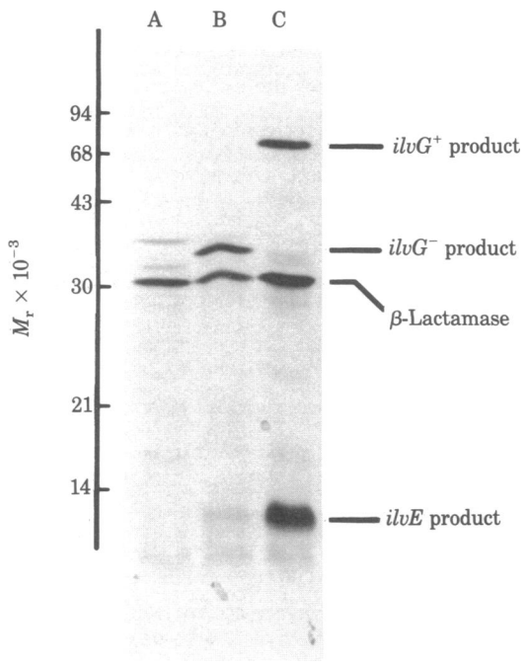

FIG. 4. Autoradiograph of *in vivo* [^{35}S]methionine-labeled proteins encoded by plasmids pBR322 (lane A), pRL5 containing the *ilvG*⁻ gene from wild-type *E. coli* K-12 (lane B), and pAH 29 containing the *ilvG*⁺ gene from the *ilv02096* mutant (lane C). Samples were prepared and electrophoresis was performed on a NaDodSO$_4$/polyacrylamide (12.5%) gel by the methods of Sancar *et al.* (17).

presence of two additional nucleotides, either A-T or T-A, in the region from base pairs (bp) 1251 to 1254. That is, in mutant *ilv02096* (*ilvG*⁺), the DNA sequence from bp 1250 to 1255 is A-A-T-A-T-G whereas the wild-type (*ilvG*⁻) sequence is A-A-T-G. An autoradiograph comparing the DNA sequence of the segment of the *ilvG* gene that contains the *ilv02096* mutation (from pAH4) with the DNA sequence of the same segment of this gene from wild-type *E. coli* K-12 (from pRL5) is shown in Fig. 3. This DNA sequence is the lower strand presented in Fig. 2. The single *Pvu* II site between *ilvL* and *ilvE* is indicated (Figs. 1 and 3). From these data it appears that, in wild-type *E. coli* K-12, translation of *ilvG* is terminated by the UGA codon (bp 1254–1256, Fig. 2), which is in the translation frame of the proximal portion of *ilvG*. This would result in the formation of a 300-amino-acid protein that terminates prior to the *Pvu* II site (Fig. 1, *ilvG*⁻ product). A mutation to *ilvG*⁺ (presumably either a single-base deletion or a two-base insertion as in *ilv02096*) must change the reading frame and allow the formation of the 521-amino-acid AHAS II protein (Fig. 1, *ilvG*⁺ product).

This expectation is confirmed by analysis of the protein products of plasmids pBR322, pRL5, and pAH29 in the "maxi-cell" strain described by Sancar *et al.* (17). This *E. coli* strain contains *recA* and *uvrA* mutations, which, upon irradiation, result in extensive degradation of chromosomal DNA. Plasmids that have not received a direct UV hit continue to replicate, resulting in an enrichment of the plasmid and, hence, of plasmid products. Thus, addition of a radiolabeled amino acid after plasmid enrichment permits specific labeling of the plasmid-encoded proteins. Lanes A, B, and C in Fig. 4 present, respectively, the protein products of pBR322, pRL5, and pAH29. As can be seen, there is a protein of $M_r \approx 35{,}000$ in lane B (*ilvG*⁻) that is not present in lane A or C. Likewise, there is a protein of $M_r \approx 70{,}000$ present in lane C (*ilvG*⁺, *ilv02096*) that is not present in lane A or B. The protein of M_r 11,000 is the truncated *ilvE'* gene product formed by the fusion of the first 243 bp of this gene with the 15 bp following the *Hin*dIII site of pBR322 (18). These data are consistent with our analysis of the DNA sequence presented in Figs. 2 and 3.

DISCUSSION

Our knowledge of the organization and regulation of the genes required for the biosynthesis of isoleucine, leucine, and valine has grown dramatically. The *ilv0* locus was initially believed to be a regulatory site between *ilvA* and *ilvC* that affected the expression of the *ilvADE* genes (7, 10). Subsequently, it became apparent that the valine-resistant phenotype of *ilv0* is due to the expression of the valine-resistant AHAS II, the product of *ilvG* (4), and that the *ilv0* locus is located prior to *ilvE* (i.e., *ilv0EDA*; ref. 19). Recent work indicated that *ilv0* is located between *ilvG* and *ilvE* (i.e., *ilvG0EDA*; refs. 8 and 9). It is now clear from the sequences presented in Figs. 2 and 3 that the site of *ilv0* is actually within *ilvG* and that the *ilv0* mutation eliminates a site of polarity within the structural gene for AHAS II. This change results in both the production of AHAS II and an increase in the expression of the distal portions of the *ilvGEDA* operon. Consequently, the *ilv0* designation should no longer be used because it is not a regulatory locus and because wild-type *E. coli* K-12 is *ilvG*⁻ and valine-resistant derivatives that "restore" AHAS II activity are *ilvG*⁺.

It seems most likely that the start of the *ilvG* gene is the AUG codon at bp 352–354 and not the AUG codon at bp 271–273 as previously suggested (12). Our revised estimate is based upon evidence from the complete DNA sequence of this gene (Fig. 2) and the size of the protein products of the plasmids containing either *ilvG*⁻ or *ilvG*⁺ genes (Figs. 1 and 4). Examination of the DNA sequence prior to the AUG codon at bp 271–273 indicates the absence of a region complementary to the 3′ terminus of 16S ribosomal RNA, which is believed to be necessary for initiation of translation (20). Both the AUG codon at bp 352–354 and the AUG codon at bp 424–426 are preceded by the sequence G-G-T-G at bp 340–343 and 415–418, respectively. These sequences are a reasonable match for the 3′ ter-

minus of 16S ribosomal RNA (20). The former of these two AUG codons has been chosen because of its relative proximity to the attenuator and the sizes of the protein products of the plasmids. Absolute assignment of the start of this gene must, of course, await determination of the amino-terminal sequence of the AHAS II protein.

Published results on the selection of valine-sensitive (*ilvG*$^-$) derivatives of *ilv0* strains (9) and the restriction endonuclease mapping of the *ilv0* site (2) are not entirely consistent with the results reported here. Two classes of *ilvG* mutations isolated either by chemical mutagenesis or by insertion of bacteriophage Mu have been reported (9). These *ilvG*$^-$ strains differ in the level of expression of distal portions of the operon. In the chemically mutagenized strains, this may reflect the difference between *ilvG*$^-$ missense mutations (*ilvG*$^-$, with high expression of *ilvEDA*) and the polarity of nonsense mutations (*ilvG*$^-$, with low expression of *ilvEDA*). Because insertion of bacteriophage Mu is presumably polar (21), there should be no expression of *ilvEDA*. However, in the distal portion of *ilvG* there is an RNA polymerase binding site (bp 2063–2095; ref. 11) which may serve as a promoter, allowing a low expression of *ilvEDA*. This is similar to the low-level internal promoter proposed in *Salmonella typhimurium* (22). This second promoter may explain why *ilvG*$^-$::Mu strains could exhibit a low expression of the *ilvEDA* genes but does not explain *ilvG*$^-$::Mu strains with high expression (9). Subrahmanyam *et al.* (23) have concluded from restriction endonuclease mapping that the *ilv0* site is between the *Kpn* I and *Sal* I restriction sites (Fig. 1). This conclusion is based upon exchanging different portions of wild-type and *ilv0*-containing plasmids for one another. A careful resequencing of the DNA between the *Kpn* I and *Sal* I restriction sites of pAH4 (*ilvG*$^+$, *ilv02096*) indicates that the DNA from this region is identical to that of wild type (unpublished observations).

The existence of a site of polarity within the *ilvG* gene of wild-type *E. coli* K-12 can explain the large increase in expression of the distal *ilvEDA* genes in *ilvG*$^+$ strains or in strains containing an altered rho factor (24–26). Comparison of the intensity of the "*ilvE* product" band in lanes B and C of Fig. 4 demonstrates the increased expression of the *ilvE* gene when it is distal to an *ilvG*$^+$ gene. The polarity-relieving effect of altered rho factor has been interpreted as indicating an attenuator prior to these genes (24–27). Indeed, the *ilvGEDA* operon does have an attenuator prior to *ilvG*, which has been characterized by DNA sequence and *in vitro* transcription analyses (12, 13). However, because the characteristic phenotype of *rho* mutations is the suppression of polarity (28), it seems most likely that the strong effect of an altered rho factor on the expression of the *ilvEDA* genes in wild-type *E. coli* K-12 is due to polarity suppression within the *ilvG* gene. An interesting aspect of these results is that the mutation of *ilvG*$^-$ to *ilvG*$^+$*2096* is a two-base pair insertion. The mechanism of this insertion event and the genetic flexibility of this region are unknown.

Note Added in Proof. Henceforth, in compliance with the rules of genetic nomenclature (29), the *ilv* promoter-leader-attenuator region (*ilvL* in this publication) shall be designated $ilvG_{p,e,a}$. The *ilvG* gene in wild-type *E. coli* K12 shall be designated *ilvG*$^+$(IlvG$^-$) and the mutations previously designated *ilv0* (e.g., *ilv0 2096*) shall be designated *ilvG*$^-$2096(IlvG$^+$).

We thank Mike Stallcup and Bert Ely for critically reading this manuscript. We are grateful to Marie Pollack and Twyla Miner for their expert technical assistance. This work was supported by grants from the National Institutes of Health [GM-24330 (to G.W.H.), GM-23182 (to D.H.C.), GM-28021 (to R.P.L.), and 5-507-RR07160 (to R.P.L.)].

1. Tatum, E. L. (1946) *Cold Spring Harbor Symp. Quant. Biol.* **11**, 278–284.
2. DeFelice, M., Levinthal, M., Iaccarino, M. & Guardiola, J. (1979) *Microbiol. Rev.* **43**, 42–58.
3. Newman, T. C. & Levinthal, M. (1980) *Genetics* **96**, 59–77.
4. Favre, R., Wiater, A., Puppo, S., Iaccarino, M., Noelle, R. & Freundlich, M. (1976) *Mol. Gen. Genet.* **143**, 243–252.
5. DeFelice, M., Guardiola, J., Esposito, B. & Iaccarino, M. (1974) *J. Bacteriol.* **120**, 1068–1077.
6. Glover, S. W. (1962) *Genet. Res.* **3**, 448–460.
7. Ramakrishnan, T. & Adelberg, E. A. (1965) *J. Bacteriol.* **89**, 654–660.
8. Baez, M., Patin, D. W. & Calhoun, D. H. (1979) *Mol. Gen. Genet.* **169**, 289–297.
9. Smith, J. M., Smith, F. J. & Umbarger, H. E. (1979) *Mol. Gen. Genet.* **169**, 299–314.
10. Ramakrishnan, T. & Adelberg, E. H. (1965) *J. Bacteriol.* **89**, 661–664.
11. Lawther, R. P., Nichols, B., Zurawski, G. & Hatfield, G. W. (1979) *Nucleic Acids Res.* **7**, 2289–2301.
12. Lawther, R. P. & Hatfield, G. W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1862–1866.
13. Nargang, F. E., Subrahmanyam, C. S. & Umbarger, H. E. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1823–1827.
14. Gayda, D. J., Leathers, T. D., Noti, J. D., Smith, F. J., Smith, J. M., Subrahmanyam, C. S. & Umbarger, H. E. (1980) *J. Bacteriol.* **142**, 556–567.
15. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
16. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353.
17. Sancar, A., Hack, A. M. & Rupp, W. D. (1979) *J. Bacteriol.* **137**, 692–693.
18. Sutcliffe, J. G. (1978) *Cold Spring Harbor Symp. Quant. Biol.* **43**, 77–90.
19. Cohen, B. N. & Jones, E. W. (1976) *Genetics* **83**, 201–225.
20. Steitz, J. A. (1979) in *Biological Regulation and Development*, ed. Goldberger, R. F. (Plenum, New York), Vol. 1, pp. 349–399.
21. Howe, M. M. & Bade, E. G. (1975) *Science* **190**, 624–632.
22. Shaw, K. J., Berg, C. M. & Sobol, T. J. (1980) *J. Bacteriol.* **141**, 1258–1263.
23. Subrahmanyam, C. S., McCorkle, G. M. & Umbarger, H. E. (1980) *J. Bacteriol.* **142**, 547–555.
24. Wasmuth, J. J. & Umbarger, H. E. (1973) *J. Bacteriol.* **116**, 548–561.
25. Smith, J. M., Smolin, D. E. & Umbarger, H. E. (1976) *Mol. Gen. Genet.* **148**, 111–124.
26. Lawther, R. P. & Hatfield, G. W. (1978) *Mol. Gen. Genet.* **167**, 227–234.
27. Bertrand, K., Korn, L., Lee, F., Platt, T., Squires, C. L., Squires, C. & Yanofsky, C. (1975) *Science* **189**, 22–26.
28. Richardson, J. P., Grimley, C. & Lowery, C. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1725–1729.
29. Bachmann, B. & Brooks Low, K. (1980) *Microbiol. Rev.* **44**, 1–56.