Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 8447–8451, November 1986
Biochemistry

β-Glucuronidase from *Escherichia coli* as a gene-fusion marker

(DNA sequence/*uidA* gene/reporter gene/enzyme purification)

RICHARD A. JEFFERSON*, SEAN M. BURGESS, AND DAVID HIRSH†

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO 80309

Communicated by William B. Wood, June 23, 1986

ABSTRACT **We have developed a gene-fusion system based on the *Escherichia coli* β-glucuronidase gene (*uidA*). The *uidA* gene has been cloned from *E. coli* K-12 and its entire nucleotide sequence has been determined. β-Glucuronidase has been purified to homogeneity and characterized. The enzyme has a subunit molecular weight of 68,200, is very stable, and is easily and sensitively assayed using commercially available substrates. We have constructed gene fusions of the *E. coli lacZ* promoter and coding region with the coding region of the *uidA* gene that show β-glucuronidase activity under *lac* control. Plasmid vectors have been constructed to facilitate the transfer of the β-glucuronidase coding region to heterologous control regions, using many different restriction endonuclease cleavage sites. There are several biological systems in which *uidA*-encoded β-glucuronidase may be an attractive alternative or complement to previously described gene-fusion markers such as β-galactosidase or chloramphenicol acetyltransferase.**

The use of fusions between a gene of interest and a reporter gene with an easily detectable product offers several advantages for the study of gene expression. The use of a single set of assays to monitor the expression of diverse gene control regions simplifies analysis and often enhances the sensitivity with which measurements of gene activity can be made. Many genes in higher organisms are members of gene families consisting of several related genes whose expression may be independently controlled (1). It is often desirable to study the expression of one member of such a gene family free from the background of the other members of the family. The use of *in vitro*-generated gene fusions and DNA transformation permits such an analysis.

The most frequently used reporter gene is probably the *Escherichia coli lacZ* gene, which encodes a β-galactosidase (2, 3). β-Galactosidase has many features that make it attractive as a gene fusion marker. The gene and gene product are well characterized genetically and biochemically (3). There are sensitive assays for the enzyme that utilize commercially available substrates, including several that allow visualization of enzyme activity *in situ*. β-Galactosidase is not, however, ideal for all systems. There are several intensively studied biological systems in which endogenous β-galactosidase levels are high enough that it is difficult or impossible to detect chimeric β-galactosidase by enzymatic methods. In addition, the enzyme and gene are very large, sometimes making the *in vitro* construction and analysis of gene fusions unwieldy.

Another gene that has been used recently in the analysis of *in vitro*-generated gene fusions encodes a chloramphenicol acetyltransferase (CAT). There is very little endogenous CAT activity in most eukaryotic systems that have been studied, but quantitative enzyme assays are expensive, laborious, and complicated by the presence of endogenous esterases and there are no histochemical methods for analyzing the spatial distribution of enzyme activity in tissues (4).

Because of some of these limitations, we have developed a gene fusion system that uses the *E. coli* β-glucuronidase gene (*uidA*) as the reporter gene. β-Glucuronidase (β-D-glucuronoside glucuronosohydrolase, EC 3.2.1.31) is an acid hydrolase that catalyzes the cleavage of a wide variety of β-glucuronides. Substrates for β-glucuronidase are generally water-soluble, and due to the extensive analysis of mammalian glucuronidases (6), many substrates are commercially available, including substrates for spectrophotometric, fluorometric, and histochemical analyses. This ability to perform histochemical analysis of gene fusions is an important feature for the study of gene expression in metazoans and plants, where spatial discrimination is often essential for assessing the regulation of genes. Methods have been described that allow subcellular localization of glucuronidase activity (reviewed in ref. 7).

The *uidA* gene has been analyzed genetically and was shown by Novel and coworkers (8–11) to be the β-glucuronidase structural gene. Plasmid clones have been obtained that contain the *uidA* locus, and a partial DNA sequence of the *uidA* regulatory region has been published (12, 13).

MATERIALS AND METHODS

DNA Manipulation. Restriction endonucleases and DNA-modifying enzymes were obtained from New England Biolabs whenever possible and used per the instructions of the supplier. Plasmid DNA preparations were done by the method of Birnboim and Doly (14) as described by Maniatis *et al.* (15). Routine cloning procedures, including ligations and transformation of *E. coli* cells, were performed essentially as described (15). DNA fragments were purified from agarose gels by electrophoresis onto Schleicher & Schuell NA45 DEAE membrane (16) as recommended by the manufacturer. DNA sequences were determined by the dideoxy chain-terminator method of Sanger and Coulson (17), as modified by Biggin *et al.* (18). Oligodeoxynucleotide primers for sequencing and site-directed mutagenesis were synthesized using an Applied Biosystems (Foster City, CA) DNA synthesizer and were purified by preparative polyacrylamide gel electrophoresis. Site-directed mutagenesis was performed on single-stranded DNA obtained from pEMBL-derived plasmids, essentially as described (19). The strain used for routine manipulation of the *uidA* gene was RAJ201, a *recA* derivative of JM83 (20) generated by bacteriophage P1 transduction. Strain PK803 was obtained from P. Kuempel (University of Colorado at Boulder) and contains a deletion of the *manA–uidA* region. Plasmid vectors pUC7, -8, and -9 (20) and pEMBL-9 (21) have been described.

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: bp, base pair(s); kb, kilobase(s).

*Present address: Department of Molecular Genetics, Plant Breeding Institute, Maris Lane, Trumpington, Cambridge, England CB2 2LQ.

†Present address: Synergen, Inc., 1885 33rd Street, Boulder, CO 80301.

Protein Sequencing and Amino Acid Analysis. Sequence analysis was performed by A. Smith (Protein Structure Laboratory, University of California, Davis), using a Beckman 890M spinning-cup sequenator. Amino acid composition was determined by analysis of acid hydrolysates of purified β-glucuronidase on a Beckman 6300 amino acid analyzer.

Protein Analysis. Protein concentrations were determined by the dye-binding method of Bradford (22), using a kit supplied by Bio-Rad Laboratories. $NaDodSO_4$/PAGE was performed using the Laemmli system (23).

β-Glucuronidase Assays. Glucuronidase was assayed in a buffer consisting of 50 mM sodium phosphate (pH 7.0), 10 mM 2-mercaptoethanol, 0.1% Triton X-100, and 1 mM *p*-nitrophenyl β-D-glucuronide. Reactions occurred in 1-ml volumes at 37°C and were terminated by the addition of 0.4 ml of 2.5 M 2-amino-2-methylpropanediol. *p*-Nitrophenol absorbance was measured at 415 nm. Routine testing of bacterial colonies for β-glucuronidase activity was done by transferring bacteria with a toothpick into microtiter wells containing the assay buffer. During the preparation of this paper, the histochemical substrate 5-bromo-4-chloro-3-indolyl β-D-glucuronide (analogous to the β-galactosidase substrate "X-Gal") became commercially available (Research Organics, Cleveland, OH). We found it to be an excellent and sensitive indicator of β-glucuronidase activity *in situ* when included in agar plates at a concentration of 50 μg/ml.

Purification of β-Glucuronidase. β-Glucuronidase was purified by conventional methods from the strain RAJ201 containing the plasmid pRAJ210 (see Fig. 1). Details of the method are available upon request (26).

RESULTS

Subcloning and Sequencing of the *uidA* Gene. The starting point for the subcloning and sequencing of the β-glucuronidase gene was the plasmid pBKuidA (Fig. 1). This plasmid has been shown to complement a deletion of the *uidA–manA* region of the *E. coli* chromosome (R. Bitner and P. Kuempel, personal communication), restoring β-glucuronidase activity when used to transform the deleted strain, PK803. The strategy for the localization of the gene on the insert is shown in Fig. 1. A restriction map of the insert was obtained, and various subclones were generated in the plasmid vector pUC9 and tested for their ability to confer β-glucuronidase activity upon transformation of PK803. The intermediate plasmid pRAJ210 conferred high levels of glucuronidase activity on the deleted strain and was used for the purification of the enzyme. Several overlapping subclones contained within an 800-bp *Eco*RI–*Bam*HI fragment conferred high levels of constitutive β-glucuronidase production only when transformed into a *uidA*$^+$ host strain and showed no effect when transformed into PK803. We surmised that the 800-bp fragment carried the operator region of the *uidA* locus and was possibly titrating repressor to give a constitutively expressing chromosomal *uidA*$^+$ gene. With this information to indicate a probable direction of transcription and a minimum gene size estimate obtained from characterization of the purified enzyme (see below), we generated a series of BAL-31 deletions from the *Xho* I site of pRAJ210. The fragments were gel-purified, ligated into pUC9, and transformed into PK803. The resulting colonies were then assayed for β-glucuronidase activity. The smallest clone obtained that still gave constitutive levels of β-glucuronidase was pRAJ220, which contained a 2.4-kilobase (kb) insert. Subclones of this 2.4-kb fragment were generated in phage vectors M13mp8 and -mp9 and their DNA sequences were determined (Fig. 2).

Manipulation of the *uidA* Gene for Vector Construction. The plasmid pRAJ220 contains the promoter and operator of the *E. coli uidA* locus, as well as additional out-of-frame ATG codons that would reduce the efficiency of proper translational initiation in eukaryotic systems (24). It was necessary to remove this DNA to facilitate using the structural gene as a reporter module in gene-fusion experiments. This was done by cloning and manipulating the 5′ region of the gene separately from the 3′ region and then rejoining the two parts as a *lacZ-uidA* fusion that showed β-glucuronidase activity under *lac* control. The resulting plasmid was further modified by progressive subcloning, linker additions, and site-directed mutagenesis to generate a set of useful gene-module vectors. The details of these manipulations are in the legend to Fig. 3.

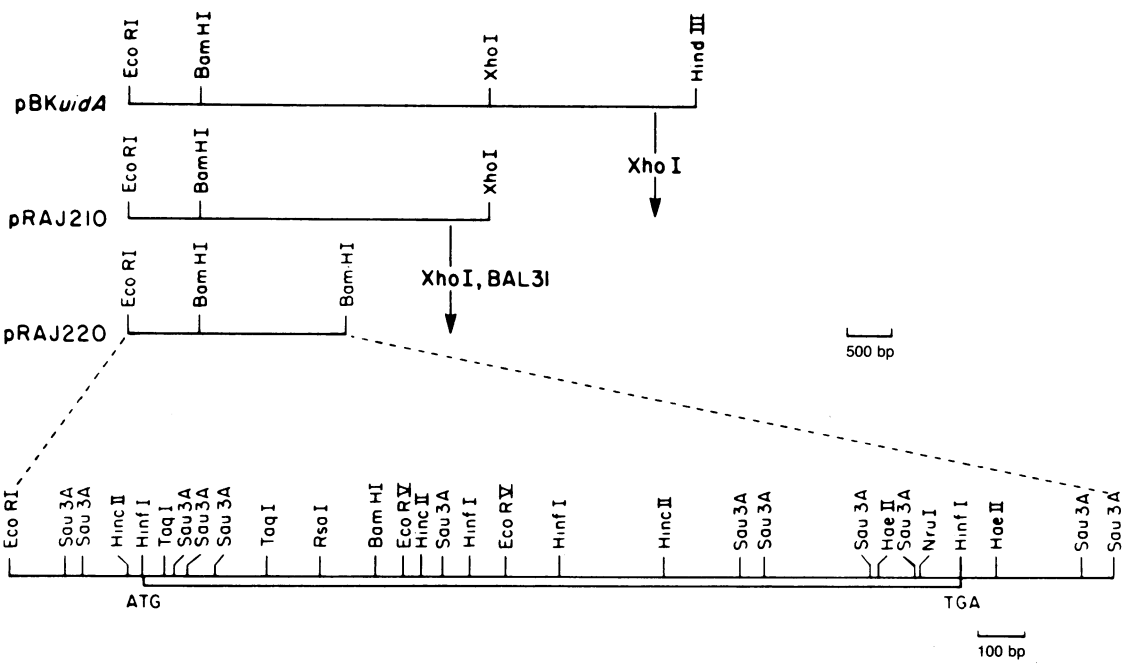

FIG. 1. Subcloning and strategy for determining the nucleotide sequence of the *uidA* gene. pBKuidA was generated by cloning into pBR325. pRAJ210 and pRAJ220 were generated in pUC9, with the orientation of the *uidA* gene opposite to that of the *lacZ* gene in the vector. Sequence was determined from both strands for all of the region indicated except nucleotides 1–125. Orientation of the coding region is from left to right. bp, Base pairs.

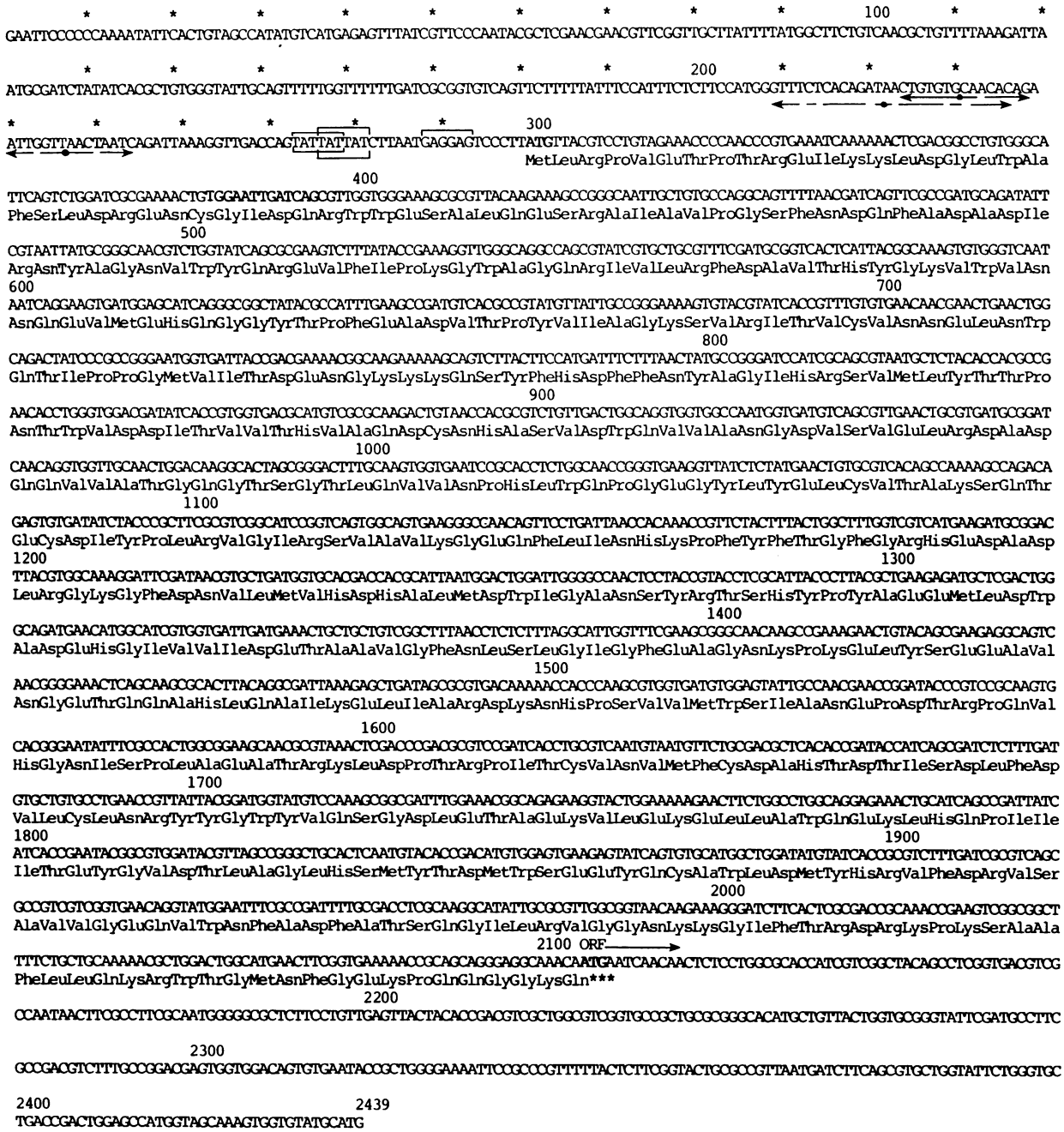

FIG. 2. DNA sequence of the 2439-bp insert of pRAJ220, containing the β-glucuronidase gene. The arrows before the coding sequence indicate regions of dyad symmetry that could be recognition sequences for effector molecules. The overlined region is the putative Shine–Dalgarno (ribosome binding site) sequence for the *uidA* gene; brackets indicate two possible Pribnow boxes. All of the palindromic regions fall within the smallest subcloned region (from the *Sau*3A site at 166 to the *Hin*fI site at 291) that gave constitutive genomic expression of *uidA* when present in high copy in *trans*, consistent with their proposed function as repressor binding sites. The terminator codon at 2106 overlaps with an ATG that may be the initiator codon of a second open reading frame, as indicated (see *Discussion*).

Purification and Properties of β-Glucuronidase. β-Glucuronidase activity in *E. coli* is induced by a variety of β-glucuronides; methyl glucuronide is among the most effective (25). To determine the size and properties of the enzyme and to verify that the enzyme produced by the clone pRAJ210 was in fact the product of the *uidA* locus, we purified the protein from the overproducing strain and compared the purified product with the enzyme induced from the single genomic locus by methyl glucuronide.

Aliquots of supernatants from induced and uninduced cultures of *E. coli* C600 were analyzed by $NaDodSO_4$/PAGE and compared with aliquots of the purified β-glucuronidase (Fig. 4). The induced culture of C600 shows only a single band difference relative to the uninduced culture. The new band comigrates with the purified β-glucuronidase, indicating that the enzyme purified from the overproducing plasmid strain has the same subunit molecular weight as the wild-type enzyme.

The purified enzyme was analyzed for amino acid composition and subjected to 11 cycles of Edman degradation to determine the amino-terminal sequence of amino acids. The amino acid composition agrees with the predicted composition derived from the DNA sequence, and the determined amino acid sequence agrees with the predicted sequence, identifying the site of translational initiation and indicating that the mature enzyme is not processed at the amino terminus (26).

E. coli β-glucuronidase is a very stable enzyme, with a broad pH optimum (pH 5.0–7.5); it is half as active at pH 4.3

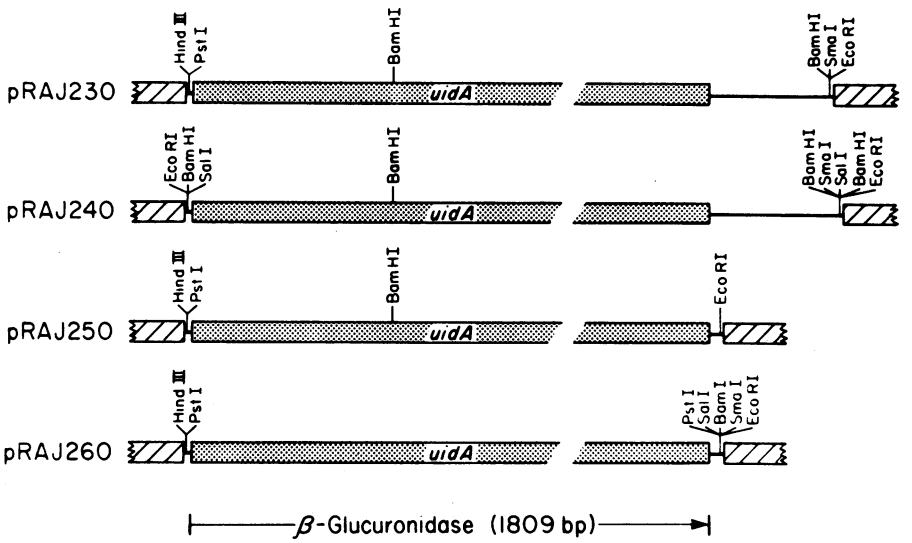

FIG. 3. β-Glucuronidase gene-module vectors. pRAJ220 (see Fig. 1) was digested with *Hin*fI, which cleaves between the Shine–Dalgarno sequence and the initiator ATG, the single-stranded tails were filled in and digested with *Bam*HI, and the resulting 515-bp fragment was gel-purified and cloned into pUC9 that had been cut with *Hin*cII and *Bam*HI. This plasmid was digested with *Bam*HI, and the 3′ region of the *uidA* gene carried on a 1.6-kb *Bam*HI fragment from pRAJ220 was ligated into it. The resulting plasmid, pRAJ230, showed isopropyl β-D-thiogalactoside inducible β-glucuronidase activity when transformed into *E. coli* JM103. pRAJ230 was further modified by the addition of *Sal* I linkers to generate pRAJ240, an in-frame *lacZ-uidA* fusion in pUC7. pRAJ230 was digested with *Aat* II, which cuts 45 bp 3′ of the *uidA* translational terminator, the ends were filled and digested with *Pst* I, and the resulting 1860-bp fragment was gel-purified and cloned into pEMBL9 cut with *Pst* I and *Sma* I. The resulting plasmid, pRAJ250, is an in-frame *lacZ-uidA* fusion. We eliminated the *Bam*HI site that occurs within the coding region at nucleotide 807 by oligonucleotide-directed mutagenesis of single-stranded DNA prepared from pRAJ250, changing the *Bam*HI site from GGATCC to GAATCC, with no change in the predicted amino acid sequence. The clone resulting from the mutagenesis, pRAJ255, shows normal β-glucuronidase activity and lacks the *Bam*HI site. This plasmid was further modified by the addition of a *Pst* I linker to the 3′ end and cloned into pEMBL9 cut with *Pst* I, to generate pRAJ260.

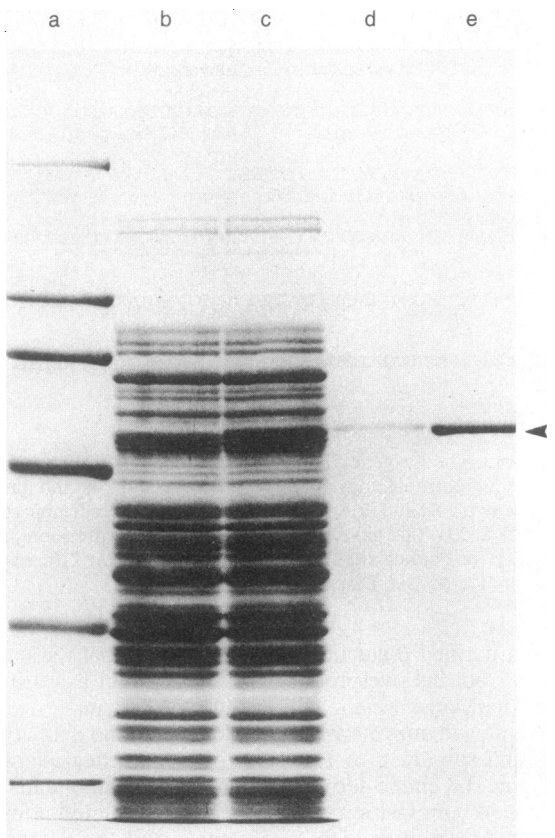

FIG. 4. NaDodSO$_4$/7.5% PAGE analysis of β-glucuronidase. Lanes: a, molecular weight standards; b, extract from uninduced *E. coli* C600; c, extract from C600 induced for β-glucuronidase with methyl glucuronide; d, 0.3 μg of purified β-glucuronidase (calculated to contain the same activity as the induced extract); e, 3.0 μg of purified β-glucuronidase.

and pH 8.5 as at its neutral optimum, and it is resistant to thermal inactivation at 50°C (26).

DISCUSSION

Molecular Analysis of the *uid* Locus. We have determined the complete nucleotide sequence of the *E. coli uidA* gene, encoding β-glucuronidase. The coding region of the gene is 1809 bp long, giving a predicted subunit molecular weight for the enzyme of 68,200, in agreement with the experimentally determined value of about 73,000. The translational initiation site was verified by direct amino acid sequence analysis of the purified enzyme.

Genetic analysis of the *uidA* locus has shown three distinct controlling mechanisms, two repressors and a cAMP-dependent factor, presumably the catabolite activator protein CAP (11). The DNA sequence that we have determined includes three striking regions of dyad symmetry that could be the binding sites for the two repressors and CAP. One of the sequences matches well with the consensus sequence for CAP binding and is located at the same distance from the putative transcriptional initiation point as the CAP binding site of the *lac* promoter. It is interesting that the putative CAP binding site overlaps one of the other palindromic sequences, suggesting a possible antagonistic effect of CAP and one or both repressors.

Our sequence analysis indicates the presence of a second open reading frame of at least 340 bp, whose initiator codon overlaps the translational terminator of the *uidA* gene. This open reading frame is translationally active (26). Although a specific glucuronide permease has been described biochemically (25), the level of genetic analysis performed on the *uid* locus would not have distinguished a mutation that eliminated glucuronidase function from a mutation that eliminated transport of the substrate (8, 9). All mutations that specifically eliminated the ability to grow on a glucuronide mapped to the *uidA* region of the *E. coli* map, indicating that if there is a gene responsible for the transport of glucuronides, it is

tightly linked to *uidA*. By analogy to the *lac* operon, we propose that the coupled open reading frame may encode a permease that facilitates the uptake of β-glucuronides.

***uidA* as a Gene-Fusion Marker.** We have constructed plasmid vectors in which the *uidA* structural gene has been separated from its promoter/operator and Shine–Dalgarno (ribosome binding site) region and placed within a variety of convenient restriction sites. The gene on these restriction fragments contains all of the β-glucuronidase coding information, including the initiator codon; there are no ATGs upstream of the initiator. These vectors allow the routine transfer of the β-glucuronidase structural gene to the control of heterologous sequences, thereby facilitating the study of chimeric gene expression in other systems.

The *uidA*-encoded β-glucuronidase is functional with several combinations of up to 20 amino acids derived from the *lacZ* gene and/or polylinker sequences. Translational fusions to β-glucuronidase have also been used successfully in transformation experiments in the nematode *Caenorhabditis elegans*, and in *Nicotiana tabacum*, giving enzyme activity with many different combinations of amino-terminal structures (refs. 5 and 26; T. Kavanagh, R.A.J., and M. Bevan, unpublished data).

There are several systems currently amenable to DNA transformation in which the study of gene fusions using β-glucuronidase as the reporter enzyme may be advantageous. Very little, if any, β-glucuronidase activity has been detected in most higher plants, including tobacco (*Nicotiana tabacum*) and potato (*Solanum tuberosum*) (unpublished data). Fusions of β-glucuronidase to several plant genes have recently been used to monitor tissue-specific gene activity in transformed tobacco plants (R.A.J., T. Kavanagh and M. Bevan, unpublished data). There is no detectable β-glucuronidase activity in the slime mold *Dictyostelium discoidum* (R. Firtel, personal communication) or the yeast *Saccharomyces cerevisiae* (unpublished data). Extracts from *Drosophila melanogaster* have shown no β-glucuronidase activity under conditions that show β-galactosidase levels several hundredfold over background (26).

We thank Peter Kuempel and Rex Bitner for the gift of pBKuidA; Chris Link for enthusiasm; and Fran Storfer, Chris Fields, and Mike Krause for reading the manuscript. This work was supported by Grant GM26515 from the National Institutes of Health.

1. Cox, G. N. & Hirsh, D. (1985) *Mol. Cell. Biol.* **5**, 363–372.
2. Lis, J. T., Simon, J. A. & Sutton, C. A. (1984) *Cell* **35**, 403–410.
3. Beckwith, J. R. & Zipser, D., eds. (1970) *The Lactose Operon* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
4. Gorman, C. M., Moffatt, L. F. & Howard, B. H. (1982) *Mol. Cell. Biol.* **2**, 1044–1051.
5. Jefferson, R. A., Klass, M. & Hirsh, D. (1986) *J. Mol. Biol.*, in press.
6. Paigen, K. (1979) *Annu. Rev. Genet.* **13**, 417–466.
7. Pearse, A. G. E. (1972) *Histochemistry: Theoretical and Applied* (Churchill Livingstone, Edinburgh), 3rd Ed., Vol. 2, pp. 808–840.
8. Novel, G. & Novel, M. (1973) *Mol. Gen. Genet.* **120**, 319–335.
9. Novel, G., Didier-Fichet, M. L. & Stoeber, F. (1974) *J. Bacteriol.* **120**, 89–95.
10. Novel, M. & Novel, G. (1976) *J. Bacteriol.* **127**, 406–417.
11. Novel, M. & Novel, G. (1976) *J. Bacteriol.* **127**, 418–432.
12. Blanco, C., Ritzenthaler, P. & Mata-Gilsinger, M. (1982) *J. Bacteriol.* **149**, 587–594.
13. Blanco, C., Ritzenthaler, P. & Mata-Gilsinger, M. (1985) *Mol. Gen. Genet.* **199**, 101–105.
14. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
15. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
16. Dretzen, G., Bellare, M., Sassone-Corsi, P. & Chambon, P. (1981) *Anal. Biochem.* **112**, 295–298.
17. Sanger, F. & Coulson, A. R. (1975) *J. Mol. Biol.* **94**, 441–448.
18. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
19. Zoller, M. J. & Smith, M. (1982) *Nucleic Acids Res.* **10**, 6487–6500.
20. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
21. Dente, L., Cesarini, G. & Cortese, R. (1983) *Nucleic Acids Res.* **11**, 1645–1655.
22. Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–254.
23. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
24. Kozak, M. (1983) *Microbiol. Rev.* **47**, 1–45.
25. Stoeber, F. (1961) These de Docteur des Sciences (University of Paris, Paris, France).
26. Jefferson, R. A. (1985) Dissertation (University of Colorado, Boulder, CO).