MOLECULAR AND CELLULAR BIOLOGY, Feb. 1987, p. 725–737
0270-7306/87/020725-13$02.00/0

Vol. 7, No. 2

Firefly Luciferase Gene: Structure and Expression in Mammalian Cells

JEFFREY R. DE WET,[1]† KEITH V. WOOD,[2] MARLENE DELUCA,[2] DONALD R. HELINSKI,[1] AND SURESH SUBRAMANI[1]*

Departments of Biology[1] and Chemistry,[2] University of California, San Diego, La Jolla, California 92093

Received 25 August 1986/Accepted 7 November 1986

The nucleotide sequence of the luciferase gene from the firefly *Photinus pyralis* was determined from the analysis of cDNA and genomic clones. The gene contains six introns, all less than 60 bases in length. The 5′ end of the luciferase mRNA was determined by both S1 nuclease analysis and primer extension. Although the luciferase cDNA clone lacked the six N-terminal codons of the open reading frame, we were able to reconstruct the equivalent of a full-length cDNA using the genomic clone as a source of the missing 5′ sequence. The full-length, intronless luciferase gene was inserted into mammalian expression vectors and introduced into monkey (CV-1) cells in which enzymatically active firefly luciferase was transiently expressed. In addition, cell lines stably expressing firefly luciferase were isolated. Deleting a portion of the 5′-untranslated region of the luciferase gene removed an upstream initiation (AUG) codon and resulted in a twofold increase in the level of luciferase expression. The ability of the full-length luciferase gene to activate cryptic or enhancerless promoters was also greatly reduced or eliminated by this 5′ deletion. Assaying the expression of luciferase provides a rapid and inexpensive method for monitoring promoter activity. Depending on the instrumentation employed to detect luciferase activity, we estimate this assay to be from 30- to 1,000-fold more sensitive than assaying chloramphenicol acetyltransferase expression.

The luciferase isolated from the common North American firefly *Photinus pyralis* is one of the most extensively studied of the enzymes that catalyze light production in bioluminescent organisms (for reviews, see references 8, 35, and 36). *P. pyralis* luciferase has an apparent molecular weight of 62,000 (59) and requires luciferin, ATP, and O_2 as substrates. The structure of firefly luciferin has been determined, and the chemical synthesis of this heterocyclic carboxylic acid has been reported (4, 57). The reactions catalyzed by firefly luciferase are:

$$\text{luciferase} + \text{luciferin} + \text{ATP} \xleftrightarrow{Mg^{2+}} \text{luciferase} \cdot \text{luciferyl-AMP} + PP_i$$

$$\text{luciferase} \cdot \text{luciferyl-AMP} + O_2 \rightarrow \text{luciferase} + \text{oxyluciferin} + \text{AMP} + CO_2 + h\nu$$

The first reaction is the formation of an enzyme-bound luciferyl-adenylate. During the second reaction, the luciferyl-adenylate undergoes an oxidative decarboxylation which results in the production of CO_2, oxyluciferin, AMP, and light. When excess substrates are added to firefly luciferase the reaction produces a flash of light that is proportional to the quantity of luciferase in the reaction mixture (9). The light emission then decays to ≈10% of the peak level within 1 min. This is followed by an extended period of low-level light emission that decays in intensity at a much slower rate. The production of light by firefly luciferase is very efficient; the quantum yield is 0.88 with respect to luciferin (46). The reaction catalyzed by *P. pyralis* luciferase emits yellow-green light at pH 7.5 to 8.5 with the peak emission at 560 nm.

We have reported the isolation and characterization of luciferase cDNA clones from a λgt11 *P. pyralis* lantern cDNA library (10). The cDNA hybridized to a single firefly lantern poly(A)$^+$ RNA species estimated to be 1.95 kilobases (kb) in length. The largest cDNA clone isolated, λLuc23, contained ≈1.8 kb of luciferase cDNA consisting of two *Eco*RI fragments (the terminal *Eco*RI sites were added as synthetic linkers during the construction of the cDNA library). This cDNA was inserted into an *Escherichia coli* expression plasmid and expressed a fusion protein in *E. coli* that exhibited the ATP- and luciferin-dependent light-emitting activity of firefly luciferase.

Sequence analysis of the Luc23 cDNA showed that it did not contain the entire luciferase-coding sequence. However, analysis of genomic clones and mapping of the 5′ end of the luciferase mRNA allowed the identification of the coding sequence and the transcriptional and translational start sites for luciferase. The nucleotide sequence and structural organization of the complete *P. pyralis* luciferase gene were elucidated by comparing the sequences of the genomic and cDNA clones.

Using the firefly luciferase gene as a means of monitoring promoter activity would offer several advantages over the bacterial chloramphenicol acetyltransferase (CAT) gene that is commonly used for this purpose in eucaryotes (19, 20). We wished to determine whether expression of the luciferase gene in mammalian cells was proportional to the strength of the promoter used to direct its transcription to demonstrate that the luciferase gene could be used as a reporter gene to monitor promoter activity. Although the cDNA clone (Luc23) that was used to express luciferase activity in *E. coli* did not contain the complete luciferase-coding region, it was possible to construct a hybrid genomic-cDNA luciferase

* Corresponding author.
† Present address: Department of Biological Sciences, Stanford University, Stanford, CA 94305.

gene that was the equivalent of a full-length cDNA. To explore the potential of this system, we constructed full-length, intronless firefly luciferase genes that included differing amounts of 5′- and 3′-untranslated sequence. The expression properties of these genes were investigated in mammalian cells with several promoters, including the simian virus 40 (SV40) early-region promoter, to direct their transcription. Assays done with cell extracts demonstrated that several of these constructs expressed substantial quantities of active luciferase after transfection into African green monkey kidney (CV-1) cells. Luciferase assays are rapid, inexpensive, very sensitive, and utilize nonradioactive substrates that are readily available from commercial sources. Finally, we demonstrated that luciferin can diffuse across mammalian cytoplasmic membranes, allowing the detection of luciferase in intact cells.

MATERIALS AND METHODS

Enzymes. Restriction endonucleases were purchased from New England BioLabs, Inc. (Beverly, Mass.) and Boehringer Mannheim Biochemicals (Indianapolis, Ind.). T4 DNA polymerase and S1 nuclease were purchased from Bethesda Research Laboratories, Inc. (Gaithersburg, Md.). Avian myeloblastosis virus reverse transcriptase, *E. coli* DNA polymerase I large fragment, calf intestinal alkaline phosphatase (molecular biology grade), and T4 DNA ligase were obtained from Boehringer Mannheim Biochemicals. Enzymes were used under the conditions specified by the suppliers. Luciferase was purified from *P. pyralis* lanterns as described previously (21).

Strains. *E. coli* LE392 (*hsdR514 supE44 supF58 lacY galK2 galT22 metB1 trpR55*) was used as a host for the λ cloning vector EMBL4. TB1 [*ara* Δ(*lac-proAB*) *rpsL* Φ80d*lacZ*Δ*M15 hsdR*$^-$ *hsdM*$^+$] and HB101 [F$^-$ *hsdS20* (r^-_B m^-_B) *recA13 ara14 proA2 lacY1 leu thi galK2 rpsL20* Str^r *xyl5 mtl1 supE44* λ$^-$] were used as hosts for plasmids. 71-18 [Δ(*lac proAB*) *thi supE*/F′ *proAB lacI*^{q}*Z*Δ*M15*] was used as the host for the propagation of M13 bacteriophage.

The *E. coli* plasmids pUC13 (37) and pUC18 and pUC19 (60) were used as vectors for the subcloning of DNA restriction fragments. Restriction fragments were purified from agarose gels by electrophoresis onto Whatman DE-81 paper (11). Plasmid pLuc23-4 consists of the entire luciferase cDNA isolated from a partial *Eco*RI digest of λLuc23 (10) inserted into the *Eco*RI site of pUC19. The insert is oriented in pLuc23-4 so that the 3′ end of the luciferase gene is proximal to the remainder of the pUC19 multicloning site restriction sites (the *Eco*RI site is at one end of the multicloning site region in the pUC series of vectors). Plasmids pLuc23A and pLuc23B were constructed, respectively, by inserting the Luc23A (3′) and Luc23B (5′) *Eco*RI cDNA fragments from λLuc23 into the *Eco*RI site of pUC13. The three *Eco*RI fragments that contained the luciferase gene were isolated from a digest of the genomic *P. pyralis* luciferase clone, λgLuc1. These fragments were inserted into the *Eco*RI site in pUC19 for the 0.7- and 1.0-kb *Eco*RI fragments and into the *Eco*RI site of pUC13 for the 1.8-kb fragment. The resultant plasmids containing the 0.7-, 1.0-, and 1.8-kb *Eco*RI genomic DNA fragments were designated pJD180, pJD181, and pJD182, respectively. pJD183 consists of an ≈4-kb *Pst*I fragment isolated from the genomic clone λgLuc3 inserted into the *Pst*I site of pUC13. This *Pst*I fragment contains the entire luciferase gene and is described in the text.

The full-length, intronless luciferase gene in pJD201 was constructed as follows. pJD180 was digested with *Eco*RI, and the fragment containing exons 1 and 2 and a portion of exon 3 of the luciferase gene was isolated. The *Eco*RI ends were filled in by treatment with *E. coli* DNA polymerase I large fragment in the presence of all four deoxynucleoside triphosphates. *Hin*dIII linkers (pCCAAGCTTGG; New England BioLabs) were ligated to the blunt ends (this regenerates the *Eco*RI sites). The DNA was digested with *Hin*dIII and inserted into the *Hin*dIII site of pUC18. The resultant plasmid, pJD200, contained this fragment oriented with its 3′-most end proximal to the remainder of the polylinker (the *Hin*dIII site in pUC18 is at one end of the polylinker). The cDNA clone pLuc23-4 was digested with *Xba*I and *Bam*HI, and the cDNA fragment was ligated to pJD200 DNA that had been digested with *Xba*I and *Bam*HI to obtain pJD201. pJD202, a 3′ deletion derivative, was constructed by digesting pJD201 with *Hin*dIII and *Ssp*I and inserting the luciferase gene fragment into pUC18 that had been digested with *Hin*dIII and *Hin*cII. pJD204, -205, -206, and -207 are 5′ deletion derivatives of the luciferase constructs carried in pJD201 and 202 and are described in the text. The phosphorylated *Bgl*II linkers (pCAGATCTG) used in the construction of pJD205 and -207 were purchased from New England BioLabs.

The SV40-derived mammalian expression vectors pSV2 and pSV0 have been described previously (54). The structures of the pSV2 derivatives pSV0A, pSV2A, and pSV232A have been described previously (26). The Rous sarcoma virus (RSV) long terminal repeat promoter vector, pRSV, has been described previously (19).

DNA sequence analysis. DNA fragments to be sequenced were inserted into the single-stranded DNA phage vectors M13mp18 or -19 (60). Plasmids pLuc23A, pLuc23B, and pLuc23-4 were the sources of the cDNA restriction fragments. Genomic luciferase restriction fragments were obtained from pJD180, -181, -182, and -183. M13 cloning, propagation, and dideoxy sequencing were as described previously (37) with the sequencing protocol modified to use deoxyadenosine 5′-[α-^{35}S]thiotriphosphate (3).

Isolation of firefly genomic DNA. *P. pyralis* fireflies were frozen in liquid nitrogen and ground to a powder under liquid nitrogen in a mortar and pestle. The powdered fireflies were suspended in 20 mM Tris hydrochloride (pH 8.0)–10 mM EDTA in a polypropylene tube. Proteinase K was added to a final concentration of 50 μg/ml, and 20% sodium dodecyl sulfate (SDS) was added with gentle stirring to a final concentration of 0.5%. The suspension was incubated at 45°C for 1 h. The lysed material was extracted twice with phenol-chloroform (1:1, vol/vol) and twice with chloroform. The chromosomal DNA was banded in an ethidium bromide-CsCl density equilibrium gradient.

Isolation of *P. pyralis* luciferase genomic clones. *P. pyralis* genomic DNA (80 μg) was partially digested with *Sau*3A (0.2 U of enzyme per μg of DNA) for 20 min. The digestion was stopped by the addition of EDTA to a concentration of 10 mM and was extracted with phenol-chloroform (1:1, vol/vol) followed by an extraction with chloroform. The *Sau*3A fragments were separated by size in a 10 to 40% sucrose velocity gradient (33). DNA fragments ranging from 12 to 23 kb were isolated from the gradient by dripping from the bottom of the tube.

The genomic library was constructed by using the λ vector EMBL4 (17). EMBL4 DNA was prepared by digesting it with *Bam*HI and *Sal*I and precipitating the DNA by the addition of 0.1 volume of 3 M sodium acetate and 1.2 volumes of isopropanol. A 5-μg sample of the prepared vector DNA was ligated to 1.2 μg of the 12- to 23-kb *Sau*3A

genomic DNA fragments. The DNA was packaged in vitro with extracts made from *E. coli* SMR10 (44). Packaged phage were plated on a lawn of *E. coli* LE392 cells without prior amplification of the library and screened by plaque hybridization (1). Filter hybridization conditions are described below.

Filter hybridizations. Nucleic acids bound to nitrocellulose filters (BA85 nitrocellulose; Schleicher & Schuell, Inc., Keene, N.H.) were hybridized with restriction fragments isolated from agarose gels by electrophoresis onto Whatman DE-81 paper (11). The DNA was labeled with [α-^{32}P]dCTP (≈800 Ci/mmol; Amersham Corp., Arlington Heights, Ill.) by nick translation (43) to a specific activity of ≈10^8 cpm/μg. Southern blots of DNA fragments resolved by agarose gel electrophoresis were prepared essentially as described previously (51). Hybridizations of probe to filters were at 37°C in 55% (vol/vol) formamide–5× SSPE (1× SSPE is 0.18 M NaCl, 10 mM sodium phosphate [pH 7.7], 1 mM EDTA)–200 μg of heparin per ml (49)–0.1% SDS containing 10^6 cpm of nick-translated probe per ml. Filters were washed in 0.1× SSPE–0.1% SDS at 37°C, and autoradiography of the filters was performed with Kodak XAR-5 film and Du Pont Cronex Lightning-Plus intensifying screens at −70°C.

5′-end mapping of luciferase mRNA. The isolation of firefly lantern total RNA and poly(A)$^+$ RNA has been described previously (10). S1 nuclease mapping of the 5′ end of the luciferase mRNA was performed essentially as described previously (15) except that a 5′-end labeled DNA fragment was used in the protection experiment (56). The probe was 5′ end labeled at base +87 (*Nar*I site at +84) and extended to base −251 (*Pst*I site at −252). The genomic luciferase clone, pJD183, was the source of the *Nar*I-*Pst*I fragment. The labeled fragment (~10 ng) was hybridized to 100 μg of *P. pyralis* lantern total RNA.

The 5′ end of the RNA was also mapped by a primer extension reaction. The 5′-end-labeled *Pst*I-to-*Nar*I fragment (see above) (0.1 μg) was digested with *Bsm*I. The ≈60-base *Nar*I-to-*Bsm*I fragment was purified by electrophoresis on an 8% polyacrylamide gel. Approximately 10 ng of this fragment was hybridized to 20 μg of *P. pyralis* lantern poly(A)$^+$ RNA. The primer fragment was hybridized to the RNA under the same conditions as used in the S1 nuclease mapping experiment. The primer was extended under the following conditions: 50 mM Tris hydrochloride (pH 8.3), 100 mM KCl, 25 mM β-mercaptoethanol, 250 μM deoxynucleoside triphosphates, 500 U of RNasin per ml, 1,000 U of avian myeloblastosis virus reverse transcriptase per ml in a total volume of 50 μl. The reaction was incubated at 42°C for 30 min after which 4 μl of 0.5 M EDTA was added to stop the reaction.

A sample of the 5′-end-labeled *Pst*I-to-*Nar*I fragment (as prepared above for S1 mapping) was cleaved under the Maxam and Gilbert (34) A > C sequencing (NaOH hydrolysis) reaction conditions to provide size markers for the S1 nuclease and primer extension products. Samples were analyzed by electrophoresis on a 7 M urea–8% polyacrylamide sequencing gel. The gel was fixed in 10% acetic acid–10% methanol for 1 h and then transferred to Whatman 3MM paper and dried. Bands in the gel were detected by autoradiography with Kodak OG-1 X-ray film with a Lanex screen at −70°C.

DNA transfections. Isolated plasmid DNAs were transfected into African green monkey kidney (CV-1) cells as described previously (39). Supercoiled plasmid DNA (10 μg) was used to transfect each 10-cm plate of cells. Each plate contained approximately 10^6 CV-1 cells at the time of transfection. Cells were harvested 48 h after transfection.

Luciferase and CAT assays. Each 10-cm plate of transfected CV-1 cells was washed three times in phosphate-buffered saline without Ca^{2+} or Mg^{2+}, and cells were harvested in 1 ml of extraction buffer (100 mM potassium phosphate [pH 7.8], 1 mM dithiothreitol) by scraping. The cells (≈5 × 10^6) from a single 100-mm dish were pelleted by centrifugation and resuspended in 100 μl of extraction buffer. Cells were lysed by three cycles of freezing on dry ice and thawing at 37°C. Cell debris was pelleted by centrifugation in a microcentrifuge for 5 min at 4°C. A 10-μl sample of extract was added to 350 μl of 25 mM glycylglycine (pH 7.8) containing 5 mM ATP (diluted from a 20 mM ATP [pH 7.5] stock solution) and 15 mM $MgSO_4$ in a small test tube. The tube was placed in an LKB luminometer equipped with a chart recorder, and the reaction was initiated by the injection of 100 μl of 1 mM luciferin (Analytical Luminescence Laboratories). The peak light emission and the time course of the reaction were recorded. Luciferase-containing extracts were stable at 4°C for at least 1 month. Extracts of cells transfected with CAT vectors were prepared as described above. CAT activity was assayed by measuring the acetylation of [^{14}C]chloramphenicol as described previously (20). Total protein in the extracts was determined by Coomassie brilliant blue G250 binding (5) with protein assay reagent (Bio-Rad Laboratories, Richmond, Calif.).

Immunocytochemical detection of luciferase. CV-1 cells were grown on cover slips and transfected with pSV2/L DNA. Forty-eight hours after the transfection, the cells were fixed and permeabilized essentially as described previously (19). Luciferase was detected by the binding of rabbit anti-*P. pyralis* luciferase antibody followed by the binding of fluorescein-conjugated goat anti-rabbit immunoglobulin G (IgG) antibody. Cells were photographed with Nomarski optics and then UV illumination to detect the fluorescent antibody. Immunofluorescent staining and microscopy were performed by Gilbert Keller.

Protein (Western) blot analysis. Protein samples were resolved by electrophoresis in an SDS–7.5% polyacrylamide gel (31). Proteins in the gel were electroblotted onto nitrocellulose in 25 mM Tris–190 mM glycine–20% methanol for 15 h at 80 mA. Rabbit anti-*P. pyralis* luciferase antibody (59) at 5 μg/ml in 10 mM Tris hydrochloride (pH 7.2)–150 mM NaCl–3% gelatin was adsorbed to the filter for 5 h at room temperature. Bound antiluciferase IgG was detected with horseradish peroxidase-conjugated goat anti-rabbit IgG antibody (Bio-Rad) and visualized with 4-chloro-1-naphthol and H_2O_2 essentially as described previously (22).

RESULTS

Sequence analysis of luciferase cDNA. The complete nucleotide sequence of both strands of the *P. pyralis* luciferase cDNA isolated from λLuc23 (10) was determined. The results of the dideoxy sequence analysis of Luc23 are shown in Fig. 1 along with an additional 320 bases of 5′ sequence that were determined from the analysis of luciferase genomic clones. The first nucleotide of the Luc23 cDNA corresponds to base +69 in Fig. 1. This was preceded by an *Eco*RI site with the sequence GAATTCC that was added as a linker during the construction of the cDNA library. A 1,633-base open reading frame in the cDNA extended from the 5′ end of the Luc23 cDNA (base +69) to a termination codon (TAA) at base 1702. The open reading frame was capable of encoding a 544-amino acid peptide. However, the first initi-

```
                                                                CTGCAGAAATAACTAGGTACTAAGCCCGTTTGTGAAAAGTGGCCAAACCCATAAATTTGGCAA
                                                          -252        -240                          -210

TTACAATAAAGAAGCTAAAATTGTGGTCAAACTCACAAACATTTTTATTATATACATTTTAGTAGCTGATGCTTATAAAAGCAATATTTAAATCGTAAACAACAAATAAAATAAAATTTA
      -180                            -150                            -120                            -90

AACGATGTGATTAAGAGCCAAAGGTCCTCTAGAAAAAGGTATTTAAGCAACGGAATTCCTTTGTGTTACATTCTTGAATGTCGCTCGCAGTGACATTAGCATTCCGGTACTGTTGGTAAA
       -60                             -30                              1                               30

met glu asp ala lys asn ile lys lys gly pro ala pro phe tyr pro leu glu asp gly thr ala gly glu gln leu his lys ala met
ATG GAA GAC GCC AAA AAC ATA AAG AAA GGC CCG GCG CCA TTC TAT CCT CTA GAG GAT GGA ACC GCT GGA GAG CAA CTG CAT AAG GCT ATG
        60                              90                              120

lys arg tyr ala leu val pro gly thr ile ala phe thr asp ala his ile glu val asn ile thr tyr ala glu tyr phe glu met ser
AAG AGA TAC GCC CTG GTT CCT GGA ACA ATT GCT TTT ACA GAT GCA CAT ATC GAG GTG AAC ATC ACG TAC GCG GAA TAC TTC GAA ATG TCC
       150                             180                             210

val arg leu ala glu ala met lys arg tyr gly leu asn thr asn his arg ile val val cys ser glu asn ser leu gln phe phe met
GTT CGG TTG GCA GAA GCT ATG AAA CGA TAT GGG CTG AAT ACA AAT CAC AGA ATC GTC GTA TGC AGT GAA AAC TCT CTT CAA TTC TTT ATG
       240                             270                             300

pro val leu gly ala leu phe ile gly val ala val ala pro ala asn asp ile tyr asn glu arg glu leu leu asn ser met asn ile
CCG GTG TTG GGC GCG TTA TTT ATC GGA GTT GCA GTT GCG CCC GCG AAC GAC ATT TAT AAT GAA CGT GAA TTG CTC AAC AGT ATG AAC ATT
       330                             360                             390

ser gln pro thr val val phe val ser lys lys gly leu glu lys ile leu asn val gln lys lys leu pro ile ile gln lys ile ile
TCG CAG CCT ACC GTA GTG TTT GTT TCC AAA AAG GGG TTG CAA AAA ATT TTG AAC GTG CAA AAA AAA TTA CCA ATA ATC CAG AAA ATT ATT
       420                             450                             480

ile met asp ser lys thr asp tyr gln gly phe gln ser met tyr thr phe val thr ser his leu pro pro gly phe asn glu tyr asp
ATC ATG GAT TCT AAA ACG GAT TAC CAG GGA TTT CAG TCG ATG TAC ACG TTC GTC ACA TCT CAT CTA CCT CCC GGT TTT AAT GAA TAC GAT
       510                             540                             570

phe val pro glu ser phe asp arg asp lys thr ile ala leu ile met asn ser ser gly ser thr gly leu pro lys gly val ala leu
TTT GTA CCA GAG TCC TTT GAT CGT GAC AAA ACA ATT GCA CTG ATA ATG AAT TCC TCT GGA TCT ACT GGG TTA CCT AAG GGT GTG GCC CTT
       600                             630                             660

pro his arg thr ala cys val arg phe ser his ala arg asp pro ile phe gly asn gln ile ile pro asp thr ala ile leu ser val
CCG CAT AGA ACT GCC TGC GTC AGA TTC TCG CAT GCC AGA GAT CCT ATT TTT GGC AAT CAA ATC ATT CCG GAT ACT GCG ATT TTA AGT GTT
       690                             720                             750

val pro phe his his gly phe gly met phe thr thr leu gly tyr leu ile cys gly phe arg val val leu met tyr arg phe glu glu
GTT CCA TTC CAT CAC GGT TTT GGA ATG TTT ACT ACA CTC GGA TAT TTG ATA TGT GGA TTT CGA GTC GTC TTA ATG TAT AGA TTT GAA GAA
       780                             810                             840

glu leu phe leu arg ser leu gln asp tyr lys ile gln ser ala leu leu val pro thr leu phe ser phe phe ala lys ser thr leu
GAG CTG TTT TTA CGA TCC CTT CAG GAT TAC AAA ATT CAA AGT GCG TTG CTA GTA CCA ACC CTA TTT TCA TTC TTC GCC AAA AGC ACT CTG
       870                             900                             930

ile asp lys tyr asp leu ser asn leu his glu ile ala ser gly gly ala pro leu ser lys glu val gly glu ala val ala lys arg
ATT GAC AAA TAC GAT TTA TCT AAT TTA CAC GAA ATT GCT TCT GGG GGC GCA CCT CTT TCG AAA GAA GTC GGG GAA GCG GTT GCA AAA CGC
       960                             990                             1020

phe his leu pro gly ile arg gln gly tyr gly leu thr glu thr thr ser ala ile leu ile thr pro glu gly asp asp lys pro gly
TTC CAT CTT CCA GGG ATA CGA CAA GGA TAT GGG CTC ACT GAG ACT ACA TCA GCT ATT CTG ATT ACA CCC GAG GGG GAT GAT AAA CCG GGC
      1050                            1080                            1110

ala val gly lys val val pro phe phe glu ala lys val val asp leu asp thr gly lys thr leu gly val asn gln arg gly glu leu
GCG GTC GGT AAA GTT GTT CCA TTT TTT GAA GCG AAG GTT GTG GAT CTG GAT ACC GGG AAA ACG CTG GGC GTT AAT CAG AGA GGC GAA TTA
      1140                            1170                            1200

cys val arg gly pro met ile met ser gly tyr val asn asn pro glu ala thr asn ala leu ile asp lys asp gly trp leu his ser
TGT GTC AGA GGA CCT ATG ATT ATG TCC GGT TAT GTA AAC AAT CCG GAA GCG ACC AAC GCC TTG ATT GAC AAG GAT GGA TGG CTA CAT TCT
      1230                            1260                            1290

gly asp ile ala tyr trp asp glu asp glu his phe phe ile val asp arg leu lys ser leu ile lys tyr lys gly tyr gln val ala
GGA GAC ATA GCT TAC TGG GAC GAA GAC GAA CAC TTC TTC ATA GTT GAC CGC TTG AAG TCT TTA ATT AAA TAC AAA GGA TAT CAG GTG GCC
      1320                            1350                            1380

pro ala glu leu glu ser ile leu leu gln his pro asn ile phe asp ala gly val ala gly leu pro asp asp asp ala gly glu leu
CCC GCT GAA TTG GAA TCG ATA TTG TTA CAA CAC CCC AAC ATC TTC GAC GCG GGC GTG GCA GGT CTT CCC GAC GAT GAC GCC GGT GAA CTT
      1410                            1440                            1470

pro ala ala val val val leu glu his gly lys thr met thr glu lys glu ile val asp tyr val ala ser gln val thr thr ala lys
CCC GCC GCC GTT GTT GTT TTG GAG CAC GGA AAG ACG ATG ACG GAA AAA GAG ATC GTG GAT TAC GTC GCC AGT CAA GTA ACA ACC GCG AAA
      1500                            1530                            1560

lys leu arg gly gly val val phe val asp glu val pro lys gly leu thr gly lys leu asp ala arg lys ile arg glu ile leu ile
AAG TTG CGC GGA GGA GTT GTG TTT GTG GAC GAA GTA CCG AAA GGT CTT ACC GGA AAA CTC GAC GCA AGA AAA ATC AGA GAG ATC CTC ATA
      1590                            1620                            1650

lys ala lys lys gly gly lys ser lys leu ***
AAG GCC AAG AAG GGC GGA AAG TCC AAA TTG TAA AATGTAACTGTATTCAGCGATGACGAAATTCTTAGCTATTGTAATATTATATGCAAATTGATGAATGGTAATTTTG
      1680                            1710                            1740                            1770

TAATTGTGGGTCACTGTACTATTTTAACGAATAATAAAATCAGGTATAGGTAACTAAAAA
               1800                            1830
```

FIG. 1. Nucleotide sequence of the *P. pyralis* luciferase cDNA and 5′-flanking sequence. The DNA sequence is listed 5′ to 3′, left to right, and is in the same sense as the luciferase mRNA sequence. Sequence from the *Pst*I site at −252 through residue +68 was determined solely from the analysis of genomic luciferase DNA; the remainder of the sequence was determined from the analysis of both luciferase cDNA and genomic DNA. The firefly luciferase cDNA, Luc23, extended from +69 through +1840 and was bounded at both termini by *Eco*RI linkers (5′-GGAATTCC-3′, not shown). The complete sequence of both strands of the cDNA was determined. The transcriptional start site of the major luciferase mRNA is defined as residue +1 (see Fig. 4). The predicted amino acid sequence is shown directly above the only long open reading frame found within the nucleotide sequence. TATA boxes and the TATA-like sequence TATTTAA are underlined in the 5′-flanking region. The presumed polyadenylation signal, AATAAA, is underlined beginning at base 1813. The sequences and locations of the six luciferase gene introns are presented in Fig. 3.

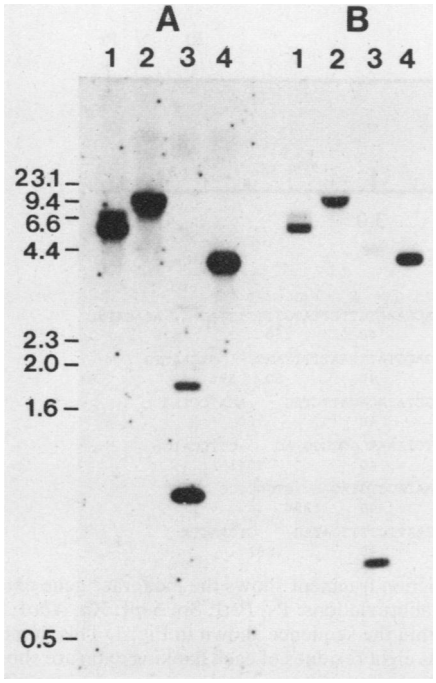


FIG. 2. Southern blot of *P. pyralis* genomic DNA. Restriction digests of *P. pyralis* genomic DNA were analyzed by agarose gel electrophoresis and blot hybridization. The DNA was digested with the following restriction endonucleases: lane 1, *Bgl*II; lane 2, *Hin*dIII; lane 3, *Eco*RI; lane 4, *Pst*I. The blot in panel A was probed with the 3′ *Eco*RI cDNA fragment Luc23A (bases 639 to 1840) labeled with ^{32}P by nick translation. Panel B is a blot of duplicate samples from the same gel as in panel A probed with the 5′ *Eco*RI cDNA fragment Luc23B (bases 69 to 639). The positions and sizes of the λ *Hin*dIII and pBR322 *Hin*fI fragments used as size standards are indicated in kilobases at the left.

ation codon within this reading frame in the cDNA occurred at base +139. Initiation at this site would give rise to a 521-amino acid polypeptide with a predicted molecular weight of ≈57,000 which is less than the apparent molecular weight of 62,000 that has been determined for *P. pyralis* luciferase (59). The discrepancy between the predicted and observed molecular weight for luciferase made it unlikely that Luc23 contained the complete luciferase-coding region. The 131-base 3′-untranslated region ended with a short poly(A) tract followed by an *Eco*RI linker (GGAATTC, not shown). The sequence analysis of other luciferase cDNA clones with longer poly(A) tracts confirmed that this was the site of poly(A) addition. A poly(A) addition signal, AATAAA (6, 38), was located 22 bases upstream from the poly(A) tract.

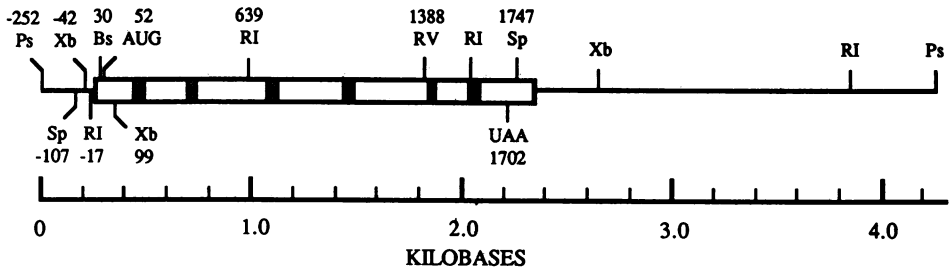
Isolation and analysis of luciferase genomic clones. Since the luciferase cDNA clone (Luc23) appeared to be missing a portion of the luciferase-coding sequence, genomic luciferase clones were isolated to determine the sequence of the 5′ end of the luciferase-coding region. Southern blot analysis of *P. pyralis* genomic DNA was performed to determine the approximate size of the luciferase gene and to determine whether the luciferase gene was a member of a gene family. Blots of genomic DNA digested with different restriction endonucleases were probed separately with the two luciferase cDNA *Eco*RI fragments, Luc23B (bases 69 to 639) and Luc23A (bases 639 to 1840), that had been labeled with ^{32}P by nick translation (Fig. 2). Both Luc23B and Luc23A hybridized to single ≈9-kb *Hin*dIII and ≈4-kb *Pst*I restriction fragments which is consistent with luciferase being a single-copy gene in *P. pyralis*. Both of these probes hybridized strongly to an ≈5-kb *Bgl*II fragment and weakly to an ≈6-kb *Bgl*II fragment. This weakly hybridizing 6-kb fragment could be due either to partial digestion of the DNA with *Bgl*II or to a restriction site polymorphism in the population. Luc23B hybridized to an ≈700-base *Eco*RI fragment, while Luc23A hybridized to ≈1- and ≈1.8-kb *Eco*RI fragments. It was expected that Luc23B and Luc23A would hybridize to different *Eco*RI bands since these cDNA fragments are separated by a naturally occurring *Eco*RI site in the cDNA. The hybridization of Luc23A to two bands resulted from the presence of an *Eco*RI site in the cDNA. The hybridization of Luc23A to two bands resulted from the presence of an *Eco*RI site within an intron in the gene.

A library of partial *Sau*3A digestion products of genomic *P. pyralis* DNA was constructed with the λ cloning vector EMBL4, and the library was screened without prior amplification. Approximately 100,000 clones were screened by plaque hybridization with ^{32}P-labeled Luc23B cDNA as the probe. The probe hybridized to four plaques that were subsequently purified to homogeneity through repeated rounds of screening. These clones were designated λgLuc1, λgLuc2, λgLuc3, and λgLuc4.

DNA from the four genomic luciferase clones was examined by agarose gel electrophoresis and Southern blot analysis (data not shown). Restriction analysis showed that each of the genomic clones was unique but that they all contained several *Eco*RI fragments in common. A Southern blot of the genomic clones was probed with ^{32}P-labeled Luc23 cDNA. All four of the genomic clones contained three *Eco*RI fragments (0.7, 1.0, and 1.8 kb) that hybridized to the luciferase cDNA. The sizes of these restriction fragments corresponded to the three *Eco*RI fragments that were observed in the Southern blot of *P. pyralis* genomic DNA (Fig. 2). A blot of *Pst*I and *Bgl*II digests of λgLuc3 DNA showed that this clone also contained the ≈4-kb *Pst*I and the ≈5-kb *Bgl*II luciferase-homologous restriction fragments that were seen in the genomic DNA Southern blot. The ≈4-kb *Pst*I luciferase fragment was isolated from a digest of λgLuc3 DNA and inserted into the *Pst*I site in pUC13. A clone with the *Pst*I fragment oriented so that the 5′ end of the luciferase gene (the end containing the ≈720-base *Eco*RI fragment) was proximal to the *Hin*dIII site in pUC13 site was designated pJD183. A map of this *Pst*I luciferase genomic DNA fragment is shown in Fig. 3.

Dideoxy sequence analysis was used to determine the nucleotide sequence of the genomic copy of the luciferase gene from the *Pst*I site in the 5′-flanking region of the gene to the site of poly(A) addition. The sequence of the 5′ region of the luciferase gene beginning at the *Pst*I site of the fragment in pJD183 (nucleotide −252) is indicated along with the Luc23 cDNA sequence in Fig. 1. The luciferase gene was found to contain six introns, and the sequences of these introns and their locations within the gene as determined by comparing the genomic sequence with that of the cDNA are shown in Fig. 3. Other than the differences resulting from the presence of the introns, there were no discrepancies between the nucleotide sequence of the Luc23 cDNA and the genomic DNA including both the 3′-untranslated region and the coding region of the luciferase gene.

The initiation (ATG) codon that began the longest open reading frame in the complete luciferase gene sequence was at base +52 (Fig. 1). This provided the coding information for an additional six amino acids upstream of those determined from the sequence of the Luc23 cDNA. In the




```
INTRON 1 (58 bp)  TTGCTTTT   GTGAGTATTTCTGTCTGATTTCTTTCGAGTTAACGAAATGTTCTTATGTTTCTTTAG    ACAGATGC
                       177          10        20        30        40        50         178
INTRON 2 (51 bp)  TAATGAAC   GTAAGCACCCTCGCCATCAGACCAAAGGGAATGACGTATTTAATTTTTAAG    GTGAATTG
                       385          10        20        30        40        50   386
INTRON 3 (48 bp)  CATGCCAG   GTATGTCGTATAACAAGAGATTAAGTAATGTTGCTACACACATTGTAG    AGATCCTA
                       719          10        20        30        40         720
INTRON 4 (49 bp)  GCAAAACG   GTGAGTTAAGCGCATTGCTAGTATTTCAAGGCTCTAAAACGGCGCGTAG    CTTCCATC
                      1040          10        20        30        40          1041
INTRON 5 (43 bp)  GATATCAG   GTAATGAAGATTTTTACATGCACACACGCTACAATACCTGTAG    GTGGCCCC
                      1395          10        20        30        40   1396
INTRON 6 (47 bp)  CGTCGCCA   GTAAATGAATTCGTTTTACGTTACTCGTACTACAATTCTTTTCATAG    GTCAAGTA
                      1561          10        20        30        40         1562
```

FIG. 3. Structure of *P. pyralis* luciferase gene. The map of the genomic *Pst*I restriction fragment shows the luciferase gene exons as open boxes and introns as shaded boxes. Restriction sites are indicated with the following abbreviations: Ps, *Pst*I; Sp, *Ssp*I; Xb, *Xba*I; RI, *Eco*RI; Bs, *Bsm*I; RV, *Eco*RV. The coordinates of the sites correspond to their locations within the sequence shown in Fig. 1. This numbering does not include the introns. The nucleotide sequences of the six luciferase gene introns plus eight residues of each flanking exon are shown directly below the map of the luciferase gene. The coordinates of the terminal bases of the exons are indicated and correspond to their locations in the DNA sequence shown in Fig. 1. The sequences of introns 2, 3, 4, and 5 were determined by sequencing both strands of the DNA. Only one strand of the DNA was sequenced for introns 1 and 6.

absence of any protein-processing events, the long open reading frame potentially encodes a 550-amino acid peptide with a molecular weight of 60,746 which is close to the apparent molecular weight of 62,000 for native *P. pyralis* luciferase.

All six of the introns were very short; the longest (intron 1) was 58 bases in length, and the smallest (intron 5) was only 43 bases in length. Only for intron 1 was it possible to unambiguously locate the intron-exon boundaries by comparing the cDNA and genomic sequences. The boundaries of the remaining five introns could be shifted by one or two bases without affecting the coding sequence that would result from the excision of the intron. In these cases, the boundaries were chosen so that each intron began with the dinucleotide GT and ended with the dinucleotide AG because these are highly conserved bases in the consensus sequences that have been determined for eucaryotic mRNA splice junctions (6, 16, 38).

Mapping the 5′ end of luciferase mRNA. From a comparison of the sequence of the luciferase genomic clone with the sequence of the cDNA it was possible to determine all the endpoints of the seven exons of the luciferase gene except for the start of the first exon, i.e., the point at which transcription of the gene begins. This was achieved by mapping the 5′ end of the luciferase mRNA by primer extension and S1 nuclease analysis (Fig. 4). A long exposure of the gel of the S1 nuclease and primer extension products showed that the longest segment of DNA protected from digestion with S1 nuclease was the same length as the major product of the primer extension reaction. The mRNA start site identified by these two fragments was defined as base +1. Although the majority of the primer extension products terminated at base +1, a shorter exposure of the gel showed that a significant proportion of the primer was extended one base further. The location of the start site of transcription confirms the ATG codon at base +52 as the first in-frame initiation codon and the most likely site for the initiation of translation of luciferase. A potential short reading frame is located upstream from the start of the putative luciferase reading frame. This short reading frame begins with an initiation (ATG) codon at base +9 and terminates at base +48 immediately preceding the presumed luciferase initiation codon.

Expression of luciferase in mammalian cells. We have previously demonstrated that the firefly luciferase cDNA, Luc23, directed the synthesis of active luciferase in *E. coli* when inserted downstream from the λ p_R promoter in an expression plasmid (10). Firefly luciferase was expressed as a fusion protein which lacked the first six amino acids of luciferase and contained eight N-terminal amino acids determined by expression vector and synthetic restriction site sequences. The isolation of the genomic luciferase clones allowed us to reconstruct a full-length, intronless luciferase gene (Fig. 5) as described in Materials and Methods. All the full-length luciferase genes are carried in the plasmid vector pUC18 as *Hin*dIII-*Bam*HI restriction fragments. pJD201, the first of these genomic DNA-cDNA fusions, is composed of genomic sequence from the *Eco*RI site at −17 to the *Xba*I site at base +99 which is located upstream from intron 1. A *Hin*dIII linker has been added immediately upstream from the *Eco*RI site at −17. The remainder of the luciferase construct carried in pJD201 consists of the Luc23 cDNA sequence extending from the *Xba*I site at +99 to the *Eco*RI site (+1840) at the 3′ end of the cDNA followed by the *Eco*RI-to-*Bam*HI portion of the polylinker from pUC19. The full-length luciferase construct in pJD201 is referred to as L in Fig. 5B. The L version of the gene contains the upstream initiation codon located at base +9, the presumed luciferase initiation codon at base +52, and the polyadenylation signal at base +1813. The other intronless luciferase gene constructs shown in Fig. 5B are derivatives of pJD201. pJD202 carries a 3′-deleted luciferase construct designated L−A (L

minus A). L−A lacks the polyadenylation signal and 88 bases of the 3′-untranslated region. pJD204 and pJD206 are 5′ deletion derivatives of pJD201 and pJD202, respectively. These were constructed by cleaving pJD201 and -202 DNAs with *Bsm*I, converting the *Bsm*I termini to blunt ends with T4 DNA polymerase, followed by adding *Hin*dIII linkers to the blunt ends. Digestion with *Hin*dIII and recircularization of the plasmids resulted in the deletion of the 17 bases of 5′-flanking DNA and 30 bases of the 5′-untranslated region including the upstream initiation codon. The luciferase gene carried in pJD204 is referred to as LΔ5′; the luciferase gene carried in pJD206 is referred to as L−AΔ5′. Also constructed, but not shown in Fig. 5, were pJD205 and pJD207. These plasmids were constructed by the insertion of *Bgl*II linkers (pGAGATCTC) at the *Bsm*I sites (+30) of pJD201 and pJD202, respectively. pJD205 and pJD207 are sources of LΔ5′ and L−AΔ5′ luciferase genes bounded by *Bgl*II sites on the 5′ side of the gene.

The *Hin*dIII-to-*Bam*HI luciferase fragment from pJD201 was inserted into pSV2 (Fig. 5) that had been prepared by digestion with *Hin*dIII and *Bgl*II. The resultant plasmid, pSV2/L, was then used to transfect CV-1 cells. Transcription of sequences inserted into pSV2 is controlled by the SV40 early promoter that is located just upstream from the *Hin*dIII site. Extracts of the cells were made and assayed for luciferase activity in a luminometer as described in Materials and Methods. The luciferase in 10 μl of extract (1/10th of a plate, ≈5 × 10^5 cells) showed an average peak emission of 5.6 × 10^4 light units/50 μg of protein (10 μl of extract usually contained ≈50 μg of protein). A peak emission of 1,000 light units is produced by 34 pg of purified *P. pyralis* luciferase when assayed in the presence of excess substrates. Assays of extracts of CV-1 cells or CV-1 cells transfected with pSV2 DNA showed a light emission of 1 to 2 light units which is essentially the background noise of the luminometer.

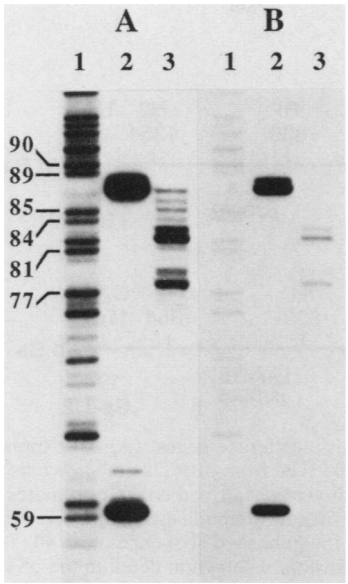


FIG. 4. Mapping the 5′ end of luciferase mRNA. The 5′ end of the *P. pyralis* luciferase mRNA was located by primer extension and S1 nuclease protection as described in the text. Restriction fragments used in the mapping experiments were obtained from the genomic *Pst*I clone (pJD183) and were 5′ end labled with ^{32}P at base +87 after cleavage with *Nar*I. Lane 1, Marker fragments produced by cleavage of the 5′ *Nar*I-to-*Pst*I fragment at A residues (34) (numbers on left are in bases). Lane 2, A 59-base primer extending from the *Nar*I site to the *Bsm*I site located at +30 was hybridized to *P. pyralis* lantern poly(A)$^+$ RNA and extended with avian myeloblastosis virus reverse transcriptase. Lane 3, The 331-base *Nar*I-to-*Pst*I fragment was hybridized to total lantern RNA and digested with S1 nuclease. The samples were analyzed by electrophoresis in a 7 M urea–8% polyacrylamide sequencing gel followed by autoradiography. Panel B is a shorter exposure of the same gel as in panel A.

The minimum amount of purified *P. pyralis* luciferase that we could reliably detect in our luminometer was 0.34 pg which had a peak emission of 10 light units. This is equivalent to ≈3 × 10^6 molecules of luciferase. We compared this with the detection limits of the CAT assay. The extract from a single plate of CV-1 cells transfected with pSV2ACAT contained a total of 2.1 U (nanomoles of chloramphenicol acetylated per minute) of CAT activity. The specific activity of CAT is 153,000 nmol/min per mg of protein (47), and the molecular weight of CAT is 25,668 (48). Therefore, the extract contained ≈3.2 × 10^{11} molecules of CAT. Decreasing amounts of cell extract were incubated with [^{14}C]chloramphenicol for 10 to 15 h instead of the standard 1 h. We found that 1/3,000th of the extract from a 10-cm plate of CV-1 cells transfected with pSV2ACAT produced two times as much acetylated chloramphenicol as an equal amount of extract from CV-1 cells alone. Since transfection of a 10-cm plate of CV-1 cells with pSV2ACAT resulted in the production of ≈3.2 × 10^{11} molecules of CAT, at least 1 × 10^8 molecules of CAT were required to produce sufficient acetylated product to allow reliable detection by autoradiography and scintillation counting. The luciferase assay, therefore, is at least 30 times more sensitive than the CAT assay on a per mole of enzyme basis with the conditions and instrumentation employed. It should be noted that luminometers more sensitive than the one used in this study are commercially available (e.g., Monolight 2001; Analytical Luminescence Laboratory). Use of these instruments could make the luciferase assay about 1,000-fold more sensitive than the CAT assay.

To examine the percentage of CV-1 cells that were transiently expressing luciferase, we transfected cells grown on cover slips with pSV2/L DNA. The cells were fixed and permeabilized 48 h after the transfection, and luciferase was detected by immunocytochemical staining (Fig. 6). Approximately 5% of the cells bound rabbit anti-*P. pyralis* luciferase antibody as visualized by the binding of fluorescein-conjugated goat anti-rabbit IgG antibody. In many of the cells, the fluorescence was seen to have a distinct punctate appearance as if much of the luciferase was localized in aggregates or perhaps to some subcellular membrane structures. The pattern and intensity of fluorescence observed was highly variable from cell to cell; it varied from relatively few spots to a generalized fluorescence of the entire cytoplasm. Protein blot analysis of extracts of CV-1 cells transiently expressing firefly luciferase showed that these cells synthesized a protein that comigrated with native *P. pyralis* luciferase on SDS-polyacrylamide gels and was recognized by anti-*P. pyralis* luciferase antibody (Fig. 7).

The eucaryotic expression vector pSV2 and several of its derivatives were used to further characterize the transient expression properties of the luciferase gene constructs (Fig. 5). The results are shown in Table 1. The data from a similar series of transient expression experiments with the CAT gene obtained from pSV2CAT (20) are included for comparison. Each plate of CV-1 cells was transfected with 10 μg of plasmid DNA, cells were harvested 48 h after the transfec-

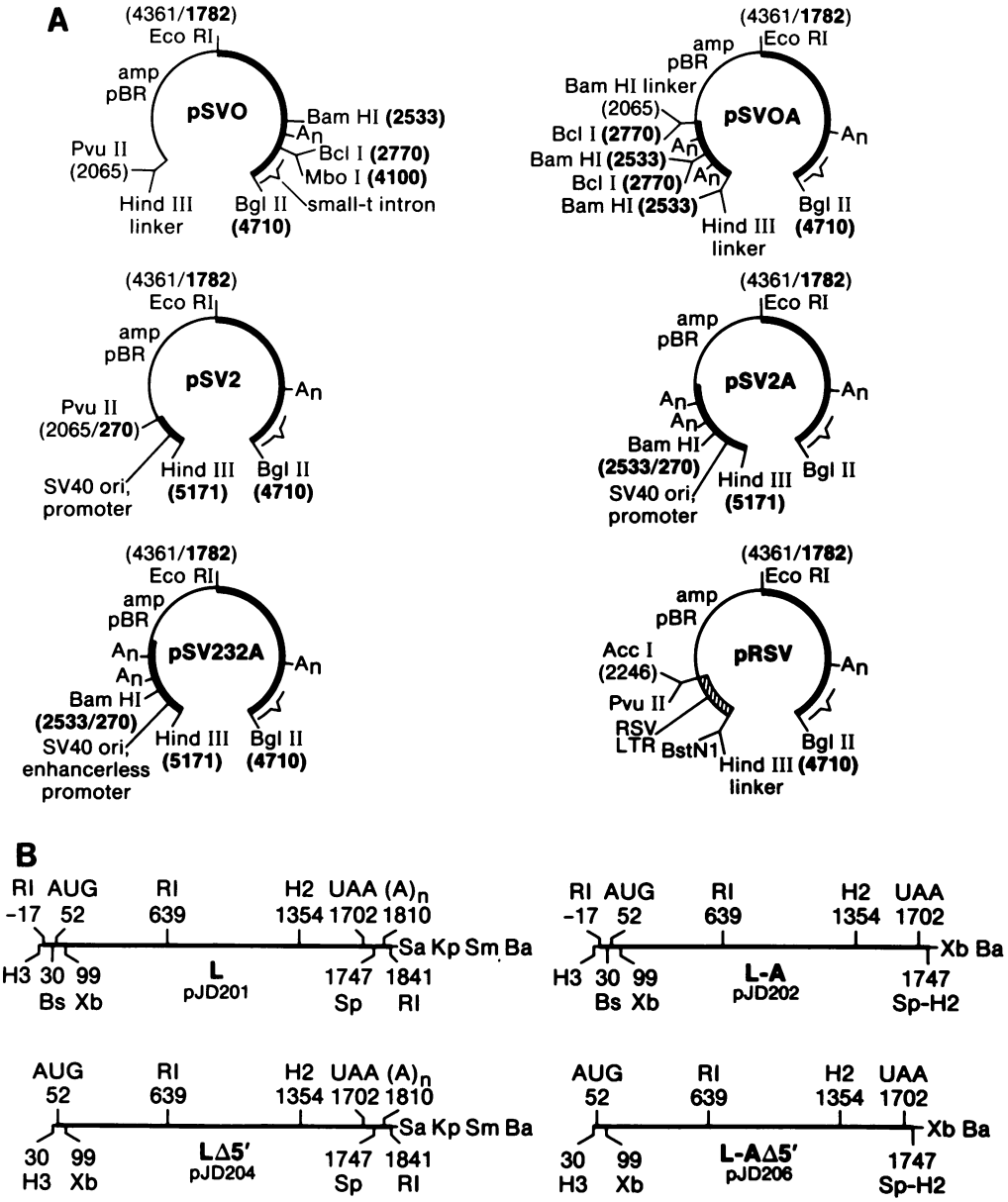


FIG. 5. Structures of the mammalian expression vectors and the full-length, intronless luciferase genes. (A) All expression vectors are derivatives of pSV2, an SV40 early-region promoter vector (54). Portions of the vectors derived from pBR322 are shown as a thin line, those from SV40 are shown as a heavy line, and those from the RSV long terminal repeat are shown as a striped box. Coordinates of the endpoints of the pBR322 segments are indicated and correspond to the location of the restriction sites in the published sequence of pBR322 (42, 55). Coordinates indicated in bold type correspond to the locations of the restriction sites in the published sequence of SV40 (7). All the vectors contain the SV40 small-t-antigen intron and an SV40 polyadenylation signal labeled A_n (structure shown in detail in the pSV0 map). Vectors whose names end with the letter A all contain two copies of the SV40 polyadenylation signal (shown in detail in the pSV0A map). pRSV contains an RSV long terminal repeat promoter in place of the SV40 early-region promoter (19). The *Acc*I and *Pvu*II sites shown at the junction between the pBR322 and RSV portions of the vector were destroyed during the construction of pRSV. ori, Origin. (B) The four full-length, intronless luciferase constructs, designated L, L−A, LΔ5′, and L−AΔ5′, are carried as *Hin*dIII-*Bam*HI fragments in the plasmid vector pUC18. The names of the corresponding plasmids are also indicated below each map. The coordinates of the indicated sites are taken from the sequence shown in Fig. 1. Restriction sites are indicated with the following abbreviations: H3, *Hin*dIII; RI, *Eco*RI; Bs, *Bsm*I; Xb, *Xba*I; H2, *Hin*cII; Sp, *Ssp*I; Sa, *Sac*I; Kp, *Kpn*I; Sm, *Sma*I; Ba, *Bam*HI. Sp-H2 indicates the product of ligating blunt *Ssp*I and *Hin*cII termini. Not shown are pJD205 and pJD207. pJD205 and -207 were constructed by inserting a *Bgl*II linker at the *Bsm*I sites (+30) of pJD201 and -202, respectively. pJD205 and -207 are a source of the LΔ5′ and L−AΔ5′ versions of the luciferase gene with a 5′-terminal *Bgl*II site in place of the *Hin*dIII site as shown for pJD204 and -206.

tion, and the enzyme activity per 50 μg of protein in the cell extract was determined as described in Materials and Methods. The levels of enzyme expressed by the various vector-gene combinations were normalized relative to the levels of enzyme activity expressed by pSV2/L or pSV2CAT, both defined as 100%.

pSV2 and its derivatives contain the SV40 small-t-antigen splice site and an SV40 polyadenylation signal located

downstream from the site into which foreign sequences are inserted (Fig. 5). The various vector-luciferase gene combinations were transfected into CV-1 cells, and the level of luciferase transient expression was determined and compared with the expression of pSV2/L (Table 1). Surprisingly, pSV0/L and pSV0/L−A transiently expressed substantial amounts of luciferase activity in CV-1 cells. pSV0 lacks the SV40 early promoter and enhancer region (Fig. 5) and is expected to lack transcriptional activity. It has been shown that pSV0 expresses CAT at a level that is less than 0.05% of the level expressed by pSV2CAT (20).

The relatively high level of luciferase expression in the absence of an active SV40 promoter could have resulted from the presence of an active promoter within the luciferase DNA segments or the presence of an enhancer sequence that activated cryptic promoters in the vector. Cryptic promoters are known to be present in the pBR322 region of the pSV vectors. These promoters require the presence of an enhancer for activity, and as much as 14% of the transcription from pSV2 has been found to originate in the pBR322 portion of the vector (32; S. Subramani, unpublished observations). To test these possibilities, we used pSV0A as the vector in transient expression experiments. pSV0A contains two copies of an SV40 polyadenylation signal located between the pBR322 segment of the vector and the 5′ end of the inserted gene (Fig. 5). If the luciferase fragments contain an endogenous promoter activity that is responsible for the luciferase expression observed in cells transfected with the pSV0-luciferase constructs, then there should be no difference between the expression of luciferase from pSV0 and pSV0A. If, however, a region of the luciferase gene is activating cryptic promoters in the pBR322 portion of the vector, then the level of luciferase expression from pSV0A/L should be much less than that observed for pSV0/L because the polyadenylation signals should truncate transcripts originating in pBR322 before they reach the luciferase gene. The pSV0A-luciferase constructs expressed no detectable luciferase activity in CV-1 cells. On the basis of this result it is reasonable to conclude that the luciferase transcripts from pSV0-luciferase constructs are originating from cryptic promoters in the pBR322 portion of the vector and that the

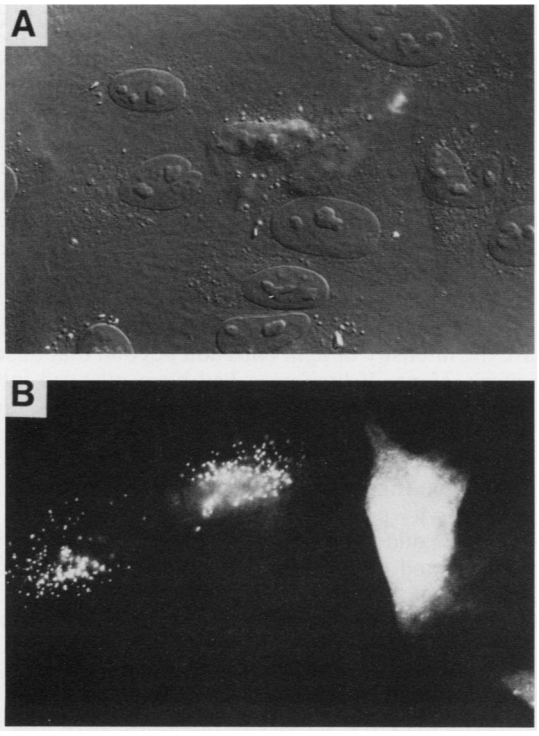


FIG. 6. Immunocytochemical detection of firefly luciferase in CV-1 cells. CV-1 cells grown on cover slips were transfected with pSV2/L DNA. The cells were fixed and permeabilized 48 h after the transfection. The cells were treated with rabbit anti-*P. pyralis* luciferase antibody followed by fluorescein-conjugated goat anti-rabbit IgG antibody. (A) Photomicrograph of the cells taken with Nomarski interference optics. (B) Corresponding photograph of the same field taken under UV illumination.

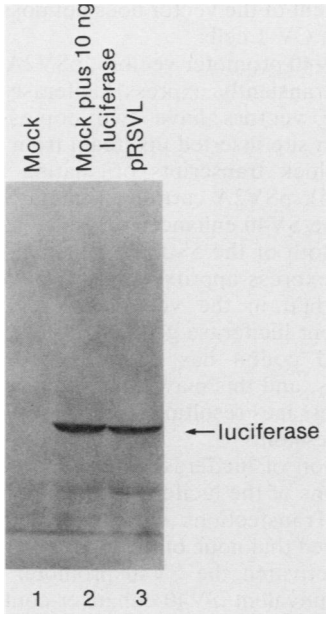


FIG. 7. Western blot analysis of cell extracts. CV-1 cells were transfected with 10 μg of plasmid DNA per 10-cm plate and harvested 48 h later. Cell extracts (150 μg of protein per lane, one-fourth to one-third of a 10-cm plate of cells) were analyzed by SDS-polyacrylamide gel electrophoresis. Proteins were electroblotted from the gel onto a nitrocellulose filter, and luciferase was detected by the binding of rabbit anti-*P. pyralis* antibody as described in Materials and Methods. The cell extracts were as follow: lane 1, CV-1 cells transfected with pSV2ACAT; lane 2, CV-1 cells transfected with pRSVCAT plus 10 ng of purified *P. pyralis* luciferase; lane 3, CV-1 cells transfected with pRSV/L.

TABLE 1. Relative levels of transient expression of the luciferase and CAT genes in CV-1 cells[a]

Vector	Gene expressed				
	L	L−A	LΔ5′	L−AΔ5′	CAT
pSV2	100	33.6			100
pSV0	14.2	8.8	3.8	5	<0.5[b]
pSV0A	0	0	0	0	<0.5
pSV2A	73		156	134	129
pSV232A	8.2		1	2.6	2.9
pRSV	250				300[c]

[a] Levels of luciferase expression were normalized relative to pSV2/L, defined as 100%. Levels of CAT expression were normalized relative to pSV2CAT, defined as 100%. Each value is the average of the results of at least four independent transfection experiments. In parallel transfections of duplicate plates of cells, the absolute number of light units produced by a given luciferase expression vector varied by less than ±15%.

[b] From Gorman et al. (20).

[c] From Gorman et al. (19).

luciferase segment of the vector does not possess a promoter that is active in CV-1 cells.

Two other SV40 promoter vectors, pSV2A and pSV232A, were used to transiently express luciferase in CV-1 cells. Both of these vectors have two copies of the SV40 polyadenylation site inserted upstream from the SV40 early promoter to block transcripts originating in pBR322 sequences (Fig. 5). pSV2A carries an intact SV40 early promoter, while the SV40 enhancer sequence has been deleted in pSV232A. Both of the 5′-deleted luciferase genes, LΔ5′ and L−AΔ5′, express approximately twofold more luciferase activity when in the vector pSV2A than does the equivalent parent luciferase gene construct, pSV2A/L. The upstream ATG codon has been deleted from the Δ5′ luciferase genes, and this may result in more efficient translation of the messages resulting in the observed high level of luciferase expression.

The expression of luciferase from the vector pSV0 suggests that regions of the luciferase gene are able to enhance transcription. Transfections with the enhancerless vector pSV232A showed that none of the luciferase gene constructs substantially activated the SV40 promoter (compare with pSV2A, the equivalent SV40-enhancer-containing vector). Since pSV232A/L expresses more luciferase than either pSV232A/LΔ5′ or pSV232A/L−AΔ5′, the L version of the luciferase gene may weakly activate the SV40 promoter.

The transient expression of an additional luciferase expression plasmid, pRSV/L, was also examined. pRSV carries a promoter from the RSV long terminal repeat. The RSV promoter is more active in CV-1 cells, and pRSVCAT has been shown to express threefold-higher levels of CAT than does pSV2CAT (19). As expected, pRSV/L expressed greater amounts of luciferase activity in CV-1 cells than pSV2/L.

The preceding luciferase expression experiments all utilized intronless luciferase genes. We also tested the *P. pyralis* genomic luciferase clone for its ability to express active luciferase in mammalian cells. pJD183, the pUC13 clone of the *Pst*I genomic luciferase fragment, was introduced into CV-1 cells. pJD183 failed to express detectable amounts of luciferase. pJD203 was, therefore, constructed to provide a *Hin*dIII-*Bam*HI restriction fragment containing the transcribed region of the luciferase gene. The luciferase gene in pJD203 began with a *Hin*dIII linker added upstream from the *Eco*RI site at −17 and continued to the *Xba*I site that is located ≈300 bases beyond the polyadenylation signal. The *Bam*HI site was provided by the polylinker of the vector, pUC18. The 5′ end of the luciferase gene in pJD203 was identical to the 5′ end of the full-length, intronless luciferase gene L carried in pJD201; however, all six introns were present. The *Hin*dIII-*Bam*HI fragment from pJD203 was inserted into pSV2, and CV-1 cells were transfected with the resultant plasmid. Again, no luciferase activity could be detected in the cell extracts. Thus, it would appear that the luciferase gene with all six introns present cannot be transcribed or processed in CV-1 cells to produce a mature, functional mRNA. This failure of the genomic luciferase clone to express luciferase in CV-1 cells may result from the size of the introns. Weiringa et al. (58) have shown that proper splicing of the rabbit β-globin large intron in HeLa cells required no specific internal sequences other than those at the splice junctions. However, if the intron was less than 80 to 90 bases in length splicing did not occur or was aberrant. None of the six luciferase gene introns meets this minimal size requirement.

We also examined the stable expression of firefly luciferase in CV-1 cells. A *Pvu*II-*Bam*HI fragment containing the SV40 early promoter and the neomycin resistance gene was isolated from pSV2neo (52). The *Pvu*II site was then converted to a *Bam*HI site with synthetic *Bam*HI linkers. The resultant 2.68-kb *Bam*HI fragment was inserted into the unique *Bam*HI site that is present in the pSV2 portion of pSV2/L to produce pSV2/L/SVneo. CV-1 cells were transfected with pSV2/L/SVneo DNA by calcium phosphate coprecipitation of the DNA. Three days after the transfection, the cells were split 1:20 and maintained in medium containing the antibiotic G418 at 400 μg/ml. After 3 weeks of selection there were hundreds of G418-resistant clones per plate. The clones from one plate were pooled, allowed to grow to confluency on a 10-cm plate, and assayed for luciferase. Cell extracts from this plate had 166% luciferase activity relative to CV-1 cells transiently expressing luciferase after transfection with pSV2/L DNA. We have also detected stable luciferase mRNAs quite easily in Northern blots of RNA obtained from cell lines expressing luciferase (data not shown).

To test the stability of luciferase expression in the cell line, we passaged the cells from the pooled G418-resistant clones for 25 generations in the presence or absence of G418. The cells maintained in the presence of G418 had 48% luciferase activity, while those grown in the absence of G418 had 29% activity (relative to pSV2/L transient expression). Luciferase was expressed in a reasonably stable fashion in these cells; however, the multiclonal origin of this line makes it difficult to arrive at any firm conclusions as to the rate of loss of the ability of the cells to express luciferase.

The addition of luciferin (0.1 mM) and magnesium-ATP (5 mM) to cells stably expressing luciferase did not produce sufficient light to be visible to the naked eye or to fog X-ray film (Kodak OG-1). However, the light produced upon the entry of luciferin into intact cells stably expressing luciferase could be detected with a luminometer (Fig. 8). The addition of luciferin to cells expressing luciferase clearly resulted in the production of light. Immediately before the luciferin was added to the cells, a sample of the buffer was removed and assayed for luciferase. The buffer contained no detectable luciferase activity; this indicated that the cells were still intact and had not released luciferase into the surrounding buffer. Furthermore, the addition of ATP, Mg^{2+}, and luciferin to a vial of intact cells expressing luciferase did not increase the intensity of light emission beyond that observed for luciferin alone. If luciferase had been released into the buffer during the course of the experiment, the addition of ATP and Mg^{2+} should have resulted in an increase in light emission. Cells in a duplicate vial were lysed by the addition of Triton X-100 (final concentration, 1%) to allow the total luciferase activity present in the cells to be measured. ATP and Mg^{2+} were added to the buffer, the vial was placed in the luminometer, and the reaction was initiated by the addition of luciferin. Peak light emission was ≈7.5-fold higher for the lysed cells as compared with the intact cells. The increase in light emission seen for the cells that had been lysed with Triton X-100 was not due entirely to the increased accessibility of luciferin to the luciferase; Triton X-100 has been shown to stimulate the activity of luciferase two- to fivefold (30).

DISCUSSION

DNA sequence analysis of the *P. pyralis* luciferase cDNA and genomic clones allowed us to determine the structure of this gene and to deduce the amino acid sequence of the

enzyme. The luciferase gene is a single-copy gene in the firefly *P. pyralis*. The gene is composed of seven exons separated by six very short introns ranging from 43 to 58 bases in length. No canonical TATA box was found in the usual position of 20 to 30 bases upstream from the putative transcriptional start site (6, 38). This sequence has been found at the proper spacing from the capsite in several insect genes including four cuticle protein genes (50), two actin genes (45), three glue polypeptide genes (18), and two yolk protein genes (23) from *Drosophila melanogaster* and eight chorion genes from the moth *Antheraea polyphemus* (25). The TATA boxes in the luciferase gene began at base −141 (TATATA) and base −116 (TATAAA), both of which are much further upstream from the site of initiation of transcription than the usual 20 to 30 bases. However, the sequence TATTTAA occurs beginning at base −30, and this sequence has been identified as the probable TATA-like sequence that is utilized in the initiation of transcription from the *D. melanogaster* alcohol dehydrogenase adult promoter (2). The alcohol dehydrogenase gene has a second larval promoter that apparently utilizes a more typical TATA sequence. The *D. melanogaster* alcohol dehydrogenase gene is expressed in different tissues of the insect in its larval and adult stages. The luciferase gene in the firefly *Photinus pennsylvanica* is similarly expressed in different tissues in larvae and adults since the adult lantern is not derived from the larval light organ (53). It may be that different transcriptional signals are utilized in controlling the expression of the luciferase gene during development.

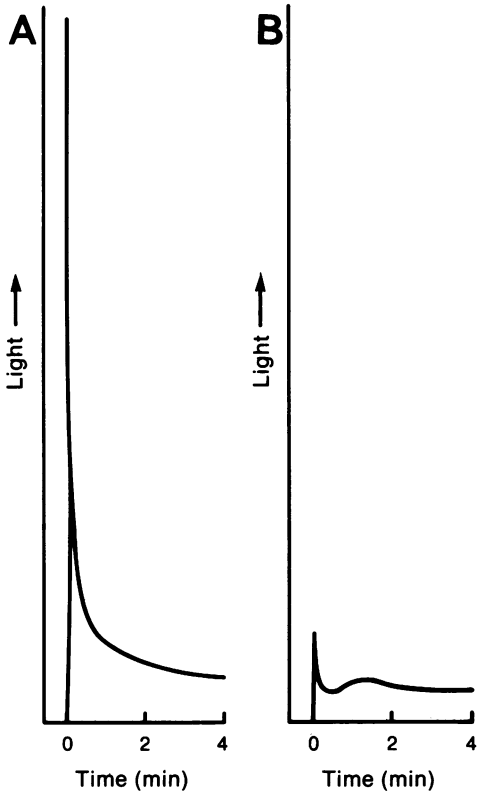


FIG. 8. Detection of luciferase activity in intact cells. CV-1 cells stably expressing luciferase were grown on cover slips in sterile scintillation vials. The cells on cover slips were washed once with buffer (25 mM glycylglycine [pH 7.8], 150 mM NaCl), and 300 μl of the same buffer was then added to each vial. After 10 min, a sample of the buffer was removed and assayed for luciferase to test for the lysis of cells. No detectable luciferase activity was released from the cells into the buffer. (A) The cells were lysed by the addition of Triton X-100 to a final concentration of 1%. ATP and $MgCl_2$ were added to final concentrations of 3.3 and 6.6 mM, respectively. The vial was placed in a luminometer (Monolight 401; Analytical Luminescence Laboratory), 200 μl of 1 mM luciferin was injected, and the time course of light emission was recorded. (B) The vial containing intact cells in buffer was placed in the luminometer, 200 μl of 1 mM luciferin was injected, and the time course of light emission was recorded.

In an earlier paper we described the expression of firefly luciferase activity in *E. coli* cells and suggested that this gene could be useful as an indicator of promoter activity in eucaryotes (10). The cloned *lux* (luciferase) operon from the marine bacterium *Vibrio fischeri* has been used as an indicator of promoter activity in *E. coli* (14). Bacterial luciferase is a two-subunit enzyme, both subunits of which are required for activity (12, 13). Since expression of bacterial luciferase requires the expression of two cistrons (*luxA* and *luxB*), the bacterial luciferase operon is less suitable for use in eucaryotes than the single-subunit firefly luciferase which is encoded by a single gene.

Although we did not isolate a full-length *P. pyralis* luciferase cDNA clone, we were able to construct its equivalent using the genomic luciferase clone as a source of the 5′ sequence that was lacking from the cDNA. This full-length, intronless luciferase gene and deletion derivatives lacking portions of the 5′- and 3′-untranslated regions of the transcription unit were placed under the control of viral promoters and expressed in monkey cells. We found that ≈5% of the CV-1 cells that were transfected with a luciferase expression vector (pSV2/L) transiently expressed the gene. An average transfection of CV-1 cells with pSV2/L yielded 560,000 light units of luciferase activity per 5×10^6 cells. Each expressing cell (5% of the total cells) on the average, therefore, produced 2.2 light units of luciferase activity. This corresponds to 0.075 pg or 7×10^5 to 8×10^5 molecules of luciferase per cell (1,000 light units is emitted by 34 pg of *P. pyralis* luciferase, molecular weight 60,746). We found it difficult to determine accurately this quantity of luciferase in a single cell because the luminometer used in this study produced 1 to 2 light units of background noise. However, with the use of more sensitive, commercially available luminometers, we have recently been able to detect the luciferase activity from a single cell expressing the enzyme from pSV2/L (unpublished data).

Comparison of the levels of luciferase and CAT activities expressed from identical vectors showed that while the L version of the luciferase gene had no endogenous promoter activity, it was most likely activating cryptic promoters within the pBR322 segment of the vector. It was also able to activate the enhancerless SV40 early promoter in the vector pSV232A, although at a much lower efficiency than the SV40 enhancer. Deletion of the 5′-most 47 bases (luciferase constructs LΔ5′ and L−AΔ5′) greatly reduced or eliminated the ability of the luciferase gene to activate transcription. While it appeared that the 5′-flanking and -untranslated region of the luciferase gene activated transcription, we did not demonstrate that this resulted from the presence of a true enhancer element that functions regardless of orientation or location with respect to the promoter. The level of luciferase expression from the LΔ5′ and L−AΔ5′ segments most closely paralleled the level of CAT expression from similar vectors. Since CAT levels have been shown to be proportional to mRNA levels (19), we expect that of the luciferase constructs we tested, the Δ5′ segments will function most reliably as reporter genes.

The same 5′ deletion that greatly reduces the transcriptional activation effects of the luciferase constructs also removes an upstream, out-of-frame initiation codon that was present in the 5′-untranslated region of the luciferase gene. The luciferase genes with an intact 5′-untranslated region were expressed quite well, and the Δ5′ luciferase constructs showed only a modest twofold increase in their level of expression when compared with the parent gene. The utilization of a downstream ATG codon as a site for the initiation of protein synthesis can be greatly reduced by the presence of an upstream ATG codon (24, 29). However, experiments have shown that eucaryotic ribosomes appear to reinitiate protein synthesis efficiently if the potential upstream reading frame is terminated near a second ATG codon (24, 29, 40, 41). This may explain why the upstream reading frame in the 5′-untranslated region of the luciferase gene did not greatly inhibit translation. Furthermore, the upstream ATG codon occurs within an unfavorable context for it to be an efficient site for the initiation of protein synthesis (27, 28).

We used the neomycin resistance gene to allow the selection of CV-1 cells that stably expressed firefly luciferase and demonstrated that luciferin could enter these cells. This allowed the detection of luciferase activity in a population of intact cells. The addition of luciferin to intact cells resulted in a lower level of light emission than that yielded by a lysate of an equal number of cells assayed in the presence of excess substrates. By manipulating the buffer conditions, however, it may be possible to increase the rate of entry of luciferin into cells thus increasing the sensitivity of the in vivo luciferase assay. The simplicity and sensitivity of the luciferase assay together with the prospects for assaying luciferase expression in intact cells make the firefly luciferase gene a significant addition to the array of genes that are currently used to monitor promoter activity.

ACKNOWLEDGMENTS

We thank Gilbert Keller for performing the immunofluorescent staining and photomicrography of cells expressing firefly luciferase.

This work was supported by Public Health Service grants GM31253 and CA01062 from the National Institutes of Health, to S.S., grant PCM 77-66573 from the National Science Foundation to D.H., and grant DMB 8305446 from the National Science Foundation to M.D.

LITERATURE CITED

1. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
2. **Benyajati, C., N. Spoerel, H. Haymerle, and M. Ashburner.** 1983. The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5′ end in different developmental stages. Cell **33**:125–133.
3. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
4. **Bowie, L. J.** 1978. Synthesis of firefly luciferin and structural analogs. Methods Enzymol. **57**:15–28.
5. **Bradford, M.** 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
6. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
7. **Buchman, A. R., L. Burnett, and P. Berg.** 1981. The SV40 nucleotide sequence, p. 799–845. *In* J. Tooze (ed.), DNA tumor viruses. Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y.
8. **DeLuca, M., and W. D. McElroy.** 1978. Purification and properties of firefly luciferase. Methods Enzymol. **57**:3–15.
9. **DeLuca, M., and W. D. McElroy.** 1984. Two kinetically distinguishable ATP sites in firefly luciferase. Biochem. Biophys. Res. Commun. **123**:764–770.
10. **de Wet, J. R., K. V. Wood, D. R. Helinski, and M. DeLuca.** 1985. Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **82**:7870–7873.
11. **Dretzen, G., M. Bellard, P. Sassone-Corsi, and P. Chambon.** 1981. A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. Anal. Biochem. **112**:295–298.
12. **Engebrecht, J., K. Nealson, and M. Silverman.** 1983. Bacterial bioluminescence: isolation and genetic analysis of functions from *Vibrio fischeri*. Cell **32**:773–781.
13. **Engebrecht, J., and M. Silverman.** 1984. Identification of genes and gene products necessary for bacterial bioluminescence. Proc. Natl. Acad. Sci. USA **81**:4154–4158.
14. **Engebrecht, J., M. Simon, and M. Silverman.** 1985. Measuring gene expression with light. Science **227**:1345–1347.
15. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
16. **Flint, S. J.** 1984. Processing of mRNA precursors in eukaryotic cells, p. 151–179. *In* D. Apirion (ed.), Processing of RNA. CRC Press, Inc., Boca Raton, Fla.
17. **Frischauf, A.-M., H. Lehrach, A. Poustka, and N. Murray.** 1983. Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**:827–842.
18. **Garfinkel, M. D., R. E. Pruitt, and E. M. Meyerowitz.** 1983. DNA sequences, gene regulation and modular protein evolution in the *Drosophila* 68C glue gene cluster. J. Mol. Biol. **168**:765–789.
19. **Gorman, C. M., G. T. Merlino, M. C. Willingham, I. Pastan, and B. H. Howard.** 1982. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. Proc. Natl. Acad. Sci. USA **79**:6777–6781.
20. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
21. **Green, A. A., and W. D. McElroy.** 1956. Crystalline firefly luciferase. Biochim. Biophys. Acta **20**:170–176.
22. **Hawkes, R., E. Niday, and J. Gordon.** 1982. A dot-immunobinding assay for monoclonal and other antibodies. Anal. Biochem. **119**:142–147.
23. **Hung, M.-C., and P. C. Wensink.** 1983. Sequence and structure conservation in yolk proteins and their genes. J. Mol. Biol. **164**:481–492.
24. **Johansen, H., D. Schümperli, and M. Rosenberg.** 1984. Affecting gene expression by altering the length and sequence of the 5′ leader. Proc. Natl. Acad. Sci. USA **81**:7698–7702.
25. **Jones, C. W., and F. C. Kafatos.** 1980. Structure, organization and evolution of developmentally regulated chorion genes in a silkmoth. Cell **22**:855–867.
26. **Kadesch, T., and P. Berg.** 1986. Effects of the position of the simian virus 40 enhancer on expression of multiple transcription units in a single plasmid. Mol. Cell. Biol. **6**:2593–2601.
27. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**: 1–45.
28. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
29. **Kozak, M.** 1984. Selection of initiation sites by eucaryotic ribosomes: effect of inserting AUG triplets upstream from the coding sequences for preproinsulin. Nucleic Acids Res. **12**: 3873–3893.
30. **Kricka, L. J., and M. DeLuca.** 1982. Effect of solvents on the catalytic activity of firefly luciferase. Arch. Biochem. Biophys. **217**:674–681.
31. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London)

227:680–685.
32. **Langer, K.-D., V. Wenger, and W. Doerfler.** 1986. Trans effect of the E1 region of adenoviruses on the expression of a prokaryotic gene in mammalian cells: resistance to 5'-CCGG-3' methylation. Proc. Natl. Acad. Sci. USA **83**:1598–1602.
33. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual, p. 284–285. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
35. **McElroy, W. D., and M. DeLuca.** 1978. Chemistry of firefly luminescence, p. 109–127. *In* P. J. Herring (ed.), Bioluminescence in action. Academic Press, Inc. (London), Ltd., London.
36. **McElroy, W. D., and M. DeLuca.** 1985. Firefly luminescence, p. 387–399. *In* J. G. Burr (ed.), Chemi- and bioluminescence. Marcel Dekker, Inc., New York.
37. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
38. **Nevins, J. R.** 1983. The pathway of eukaryotic mRNA formation. Annu. Rev. Biochem. **52**:441–466.
39. **Parker, B. A., and G. R. Stark.** 1979. Regulation of simian virus 40 transcription: sensitive analysis of the RNA species present early in infections by virus or viral DNA. J. Virol. **31**:360–369.
40. **Peabody, D. S., and P. Berg.** 1986. Termination-reinitiation occurs in the translation of mammalian cell mRNAs. Mol. Cell. Biol. **6**:2695–2703.
41. **Peabody, D. S., S. Subramani, and P. Berg.** 1986. The effect of upstream reading frames on translation efficiency in SV40 recombinants. Mol. Cell. Biol. **6**:2704–2711.
42. **Peden, K. W. C.** 1983. Revised sequence of the tetracycline resistance gene of pBR322. Gene **22**:277–280.
43. **Rigby, P. W. J., M. Dieckman, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick-translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
44. **Rosenberg, S. M., M. M. Stahl, I. Kobayashi, and F. W. Stahl.** 1985. Improved *in vitro* packaging of coliphage lambda DNA: a one strain system free from endogenous phage. Gene **38**: 165–175.
45. **Sanchez, F., S. L. Tobin, U. Rdest, E. Zulauf, and B. J. McCarthy.** 1983. Two *Drosophila* actin genes in detail: gene structure, protein structure and transcription during development. J. Mol. Biol. **163**:533–551.
46. **Seliger, H. H., and W. D. McElroy.** 1960. Spectral emission and quantum yield of firefly bioluminescence. Arch. Biochem. Biophys. **88**:136–141.
47. **Shaw, W. V., and R. F. Brodsky.** 1968. Characterization of chloramphenicol acetyltransferase from chloramphenicol-resistant *Staphylococcus aureus*. J. Bacteriol. **95**:28–36.
48. **Shaw, W. V., L. C. Packman, B. D. Burleigh, A. Dell, H. R. Morris, and B. S. Hartley.** 1979. Primary structure of a chloramphenicol acetyltransferase specified by R plasmids. Nature (London) **282**:870–872.
49. **Singh, L., and K. W. Jones.** 1984. The use of heparin as a simple cost-effective means of controlling background in nucleic acid hybridization procedures. Nucleic Acids Res. **12**:5627–5638.
50. **Snyder, M., M. Hunkapiller, D. Yuen, D. Silvert, J. Fristrom, and N. Davidson.** 1982. Cuticle protein genes of *Drosophila*: structure, organization and evolution of four clustered genes. Cell **29**:1027–1040.
51. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
52. **Southern, P. J., and P. Berg.** 1982. Transformation of cells to antibiotic resistance with a bacterial gene under the control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**: 327–341.
53. **Strause, L. G., M. DeLuca, and J. F. Case.** 1979. Biochemical and morphological changes accompanying light organ development in the firefly *Photuris pennsylvanica*. J. Insect Physiol. **25**:339–347.
54. **Subramani, S., and P. J. Southern.** 1983. Analysis of gene expression using simian virus 40 vectors. Anal. Biochem. **135**:1–15.
55. **Sutcliffe, J. G.** 1979. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43**:77–90.
56. **Weaver, R. F., and C. Weissmann.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5' termini of 15S β-globin precursor and mature 10S β-globin mRNA have identical map coordinates. Nucleic Acids Res. **6**:1175–1193.
57. **White, E. H., F. McCapra, G. F. Field, and W. D. McElroy.** 1963. The structure and synthesis of firefly luciferin. J. Am. Chem. Soc. **85**:337–343.
58. **Wieringa, B., E. Hofer, and C. Weissmann.** 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β-globin intron. Cell **37**:915–925.
59. **Wood, K. V., J. R. de Wet, N. Dewji, and M. DeLuca.** 1985. Synthesis of active firefly luciferase by *in vitro* translation of RNA obtained from adult lanterns. Biochem. Biophys. Res. Commun. **124**:592–596.
60. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.