Cloning and Expression of the *lux*Y Gene from *Vibrio fischeri* Strain Y-1 in *Escherichia coli* and Complete Amino Acid Sequence of the Yellow Fluorescent Protein†

Thomas O. Baldwin,* Mary L. Treat,‡ and S. Colette Daubner

Department of Biochemistry and Biophysics, Texas A&M University, and Texas Agricultural Experiment Station, College Station, Texas 77843

Received December 5, 1989; Revised Manuscript Received February 13, 1990

ABSTRACT: *Vibrio fischeri* strain Y-1 (ATCC 33715) emits light with a λ_{max} of 545 nm rather than the 485-nm emission typical of other strains of *V. fischeri*. The yellow emission is due to the interaction of the enzyme luciferase with a yellow fluorescent protein (YFP). On the basis of the N-terminal amino acid sequence of YFP, a mixed-sequence oligonucleotide probe was synthesized and used to isolate a 1.6-kbp *Hin*dIII fragment containing the first 208 bases of the gene that codes for YFP (*lux*Y). Another synthetic oligonucleotide complementary to bases 167–184 of the YFP coding sequence was used to isolate a second (ca. 1.9 kbp) DNA fragment generated by digestion with both *Eco*RI and *Cla*I that contained the remainder of the *lux*Y gene. The intact *lux*Y gene, which encoded a 22 211-dalton polypeptide composed of 194 amino acid residues, was reconstructed from the two primary clones and is contained within a 765-bp *Ssp*I–*Xho*II fragment. Both strands of the entire *lux*Y coding sequence were determined from the reconstructed gene, while the region surrounding the junction used in the reconstruction was also determined from the original partial clones. As with other genes that have been studied from *V. fischeri*, the *lux*Y gene was unusually AT-rich. The sequence of *lux*Y did not bear any apparent similarity to any of the sequences contained in the current GenBank database. *Escherichia coli* containing a plasmid with the *lux*Y gene expresses a protein that reacts with antibody raised to authentic YFP.

Light emission from bioluminescent bacteria is the result of the action of the enzyme bacterial luciferase, a heterodimeric flavin monooxygenase that catalyzes the reaction:

$$FMNH_2 + O_2 + RCHO \rightarrow FMN + RCOOH + H_2O + h\nu$$

where RCHO and RCOOH are long-chain fatty aldehydes and acids, respectively [see Ziegler and Baldwin (1981) for a review]. Light emission in *Vibrio harveyi* is from the singlet excited state of an enzyme-bound flavin intermediate formed in the luciferase reaction, both in vivo and in vitro (Cline & Hastings, 1972, 1974; Mitchell & Hastings, 1974).

The wavelength maximum for bioluminescence emission from most strains of *Vibrio fischeri* under current study is 485 nm, appearing blue. Intensive study of the regulation of bioluminescence in *V. fischeri* (strains MJ-1 and ATCC 7744) has identified a regulon consisting of two operons (Engebrecht et al., 1983; Engebrecht & Silverman, 1984, 1986; Baldwin et al., 1989). The organization of the regulon is shown in Figure 1. The rightward operon contains *lux*A and *lux*B, the genes that code for the α and β subunits of luciferase, as well as *lux*C, *lux*D, and *lux*E, which encode enzymes involved in the synthesis of the aldehyde substrate of luciferase. The first gene of this operon is *lux*I, which encodes a protein necessary for the synthesis of a small molecule effector or "autoinducer", *N*-(3-oxohexanoyl)homoserine lactone, a compound necessary for the induction of bioluminescence (Nealson et al., 1970; Eberhard et al., 1981; Rosson & Nealson, 1981). The leftward operon consists of one gene, *lux*R, encoding a positive regulatory protein that in the presence of autoinducer stimulates transcription of the rightward operon (Kaplan et al., 1985; Devine et al., 1988).

One strain of *V. fischeri*, designated Y-1 (ATCC 33715), isolated by Ruby and Nealson (1977), was found to display an emission spectrum with a peak at 545 nm if grown at temperatures below 22 °C but a spectrum with the usual 485-nm maximum if grown at higher temperatures. Leisman and Nealson (1982) identified a flavin-containing yellow fluorescent protein (YFP)[1] from *V. fischeri* Y-1. They showed that, in the absence of YFP, Y-1 luciferase supplied with $FMNH_2$ by a coupled assay with NADH–FMN oxidoreductase emits blue light (485 nm), but in the presence of YFP, Y-1 luciferase exhibits a bimodal emission spectrum with the second (new) peak at 545 nm. They also found that the effect of temperature on the system in vitro is similar to that in vivo: at temperatures above 20 °C, YFP does not effect the same bimodal emission that it does at lower temperatures (Leisman, 1981). Presumably *V. fischeri* Y-1 employs a mode of regulation quite similar to those of *V. fischeri* MJ-1 and ATCC 7744, since its growth and luminescence respond to the *V. fischeri* autoinducer (Ruby & Nealson, 1977), yet *V. fischeri* Y-1 expresses an additional protein, YFP, the product of a gene that we designate *lux*Y. Little is known at this time concerning the proximity of *lux*Y to the *V. fischeri* Y-1 *lux* regulon.

Photobacterium phosphoreum and *Photobacterium leiognathi* are other examples of bioluminescent marine bacteria that emit light in vivo at a different λ_{max} (470 nm) than that

† This work was supported in part by grants from the National Science Foundation (NSF DMB 87-16262) and the Robert A. Welch Foundation (A865).

‡ Current address: University of Tennessee Biomedical School, Biology Division, Oak Ridge National Laboratory, P.O. Box 2009, Oak Ridge, TN 37831-8077.

[1] Abbreviations: YFP, yellow fluorescent protein; $FMNH_2$, reduced flavin mononucleotide.

0006-2960/90/0429-5509$02.50/0

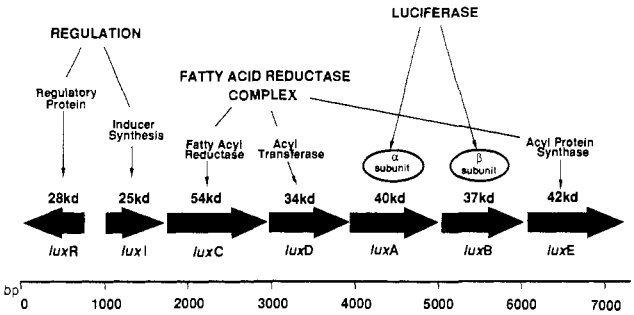

FIGURE 1: Physical map of the *lux* regulon of *V. fischeri* ATCC 7744.

emitted by the luciferase purified from the organism ($\lambda_{max} \sim$ 495 nm) (Lee, 1982). These species produce an accessory protein ("lumazine protein"), slightly different between the two species, that has a molecular weight of about 22000 (Gast & Lee, 1978). The lumazine proteins, which contain the prosthetic group 6,7-dimethylribityllumazine (Koka & Lee, 1979), cause a shift in the bioluminescence emission spectra of *P. phosphoreum*, *P. leiognathi*, *V. fischeri*, and *V. harveyi* luciferases in vitro (Lee, 1982). Similarly, we have found that YFP causes a shift in the λ_{max} of emission by luciferases purified from *P. phosphoreum*, *P. leiognathi*, and *V. fischeri* ATCC 7744. Although it does not affect the emission wavelength of *V. harveyi* luciferase, it does enhance the intensity of light emitted by that enzyme as well as the luciferases from *P. phosphoreum* and *P. leiognathi*, apparently by increasing the rate of the luciferase-catalyzed reaction (Daubner & Baldwin, 1989).

Active YFP holoprotein has been purified to homogeneity in our laboratory (Daubner et al., 1987); reconstitutable apoprotein has also been purified (Macheroux et al., 1987). The protein was found to be a homodimer of subunits of approximately M_r 22 300 (determined by polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate). It contains FMN as a noncovalently bound prosthetic group. After determination of the N-terminal sequence of the protein, we designed an oligonucleotide probe to isolate *lux*Y from genomic DNA of *V. fischeri* Y-1. We report here the use of the oligonucleotide probe to clone the *lux*Y gene and demonstrate expression of the *lux*Y gene in *Escherichia coli*. We present the sequence of *lux*Y and the encoded YFP and comment on the base composition of the gene. We also compare the amino acid composition of YFP to that of the lumazine protein.

Experimental Procedures

Materials. Restriction endonucleases, T4 DNA ligase, and T4 polynucleotide kinase were purchased from Boehringer Mannheim and New England Biolabs. M13 universal and pBR322 *Hin*dIII counterclockwise primers were purchased from New England Biolabs. Nitrocellulose was purchased from Micro Filtration Systems, and [γ-^{32}P]dATP and [α-^{32}P]dATP were obtained from New England Nuclear. Electrophoresis chemicals were purchased from ICN Biomedicals. Sequenase version 2.0 and the Sequenase DNA sequencing reagent kit were from United States Biochemical Corp. Custom oligonucleotides were synthesized with an Applied Biosystems Model 380B DNA synthesizer.

Bacterial Strains and Plasmids. *E. coli* strain TB1 (Baldwin et al., 1984) was used for the genomic DNA cloning experiments, and strain LE392 (Maniatis et al., 1982) was used for expression of the YFP-encoding plasmid. Plasmid pBR322 that had been digested with *Hin*dIII and treated with calf intestine phosphatase was purchased from New England Biolabs. Plasmid pGEM7Zf(+) was purchased from Promega Corp. Plasmid pUC19 was obtained from Bethesda Research Laboratories. *E. coli* cells were grown in LB medium containing 100 μg/mL carbenicillin at 37 °C. *V. fischeri* Y-1 cells were grown in sodium chloride complete medium at 18 °C (Nealson, 1978).

Chart I: Sequence of the Mixed Oligonucleotide Used To Probe the *Hin*dIII-Digest-Generated Clone Bank of Y-1 Genomic DNA in *E. coli*

5'-ATG	TTT	AAA	GGA	ATA	GT-3'	
	C	G	G	T		MT-1
				C		
5'-ATG	TTT	AAA	GGT	ATA	GT-3'	
	C	G	C	T		MT-2
				C		

Construction of Recombinant Plasmids. *V. fischeri* Y-1 genomic DNA from cells grown into late log phase was isolated by lysozyme–SDS lysis of the cells, phenol extraction, and ethanol precipitation and was purified by CsCl ultracentrifugation (Sato & Miura, 1963). The Y-1 genomic DNA was digested with *Hin*dIII for 3 h at 37 °C and then phenol-extracted; 10 μg of the digested Y-1 genomic DNA was mixed with 1 μg of pBR322 DNA previously cut with *Hin*dIII. The ligation reaction was allowed to proceed for 16 h at 16 °C with T4 DNA ligase. Half of this mixture was added to competent TB1 cells and incubated on ice for 110 min, then heated to 45 °C for 3 min, and diluted into 10 mL of LB media. After 60 min at 37 °C with aeration, 50 μL of the culture was spread onto each of 120 agar plates containing 100 μg/mL carbenicillin. The plates were incubated for 18 h at 37 °C. Approximately 2300 individual colonies were picked onto gridded agar plates supplemented with carbenicillin. Colonies were transferred to nitrocellulose filter circles; duplicate circles were made from each plate. The colonies on the filter circles were lysed and the DNA fixed to the filters (Maniatis et al., 1982). The filters were then probed with the ^{32}P-labeled mixed oligonucleotide shown in Chart I. Labeling of the oligonucleotide was accomplished by using T4 polynucleotide kinase. Hybridization and washing were carried out as described below under Southern Blot Analysis. Plasmid preparations were made (Holmes & Quigley, 1981) from 12 isolated colonies that appeared to hybridize with the probe. One plasmid (designated pMT23.1) hybridized with the probe in Southern blots (Southern, 1975). The partial nucleotide sequence of plasmid pMT23.1 was determined by the "dideoxy" chain termination method (Sanger et al., 1977) using MT-2 and pBR322 *Hin*dIII counterclockwise primer (New England Biolabs no. 1205) as primers. Data from these reactions demonstrated that the 5′ end of *lux*Y resided between the *Ssp*I site and the *Eco*RI site (see Figure 2). Plasmid pMT23.1 (1.0 μg) was digested with *Hin*dIII for 1 h at 37 °C and then with *Ssp*I and *Sau*3AI for 1 h at 37 °C; a small fragment (252 bases long) was cloned into 0.1 μg of pGEM7Zf(+) previously cut with *Sma*I and *Hin*dIII. The resulting plasmid was designated pMT23.2 (Figure 2).

Genomic DNA from *V. fischeri* Y-1 (10 μg), digested with *Eco*RI and *Cla*I, was mixed with 1 μg of pGEM7Zf(+) previously digested with the same enzymes. After addition of T4 DNA ligase, the mixture was incubated at 16 °C for 6 h. Half of the mixture was used to transform competent TB1 cells as described above. About 4900 colonies were picked and transferred to nitrocellulose filters. After cell lysis and DNA fixation, the filters were probed with the oligonucleotide shown in Chart II (designated *lux*Y4B). Hybridization and washing were carried out as before. Six colonies were isolated

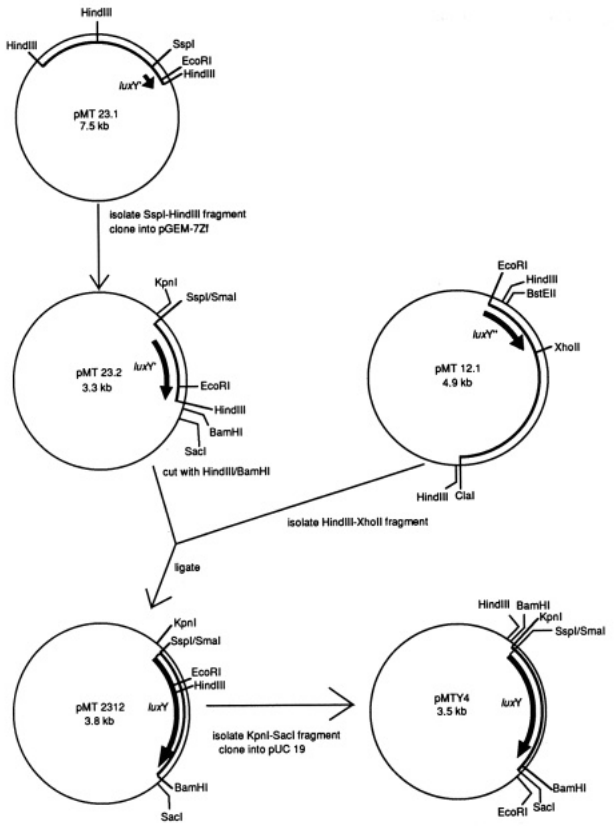

FIGURE 2: Construction of plasmids used in the cloning of the *lux*Y gene of *V. fischeri* strain Y-1. Details are given under Experimental Procedures.

Chart II: Sequence of the Oligonucleotide Used To Probe the *Eco*RI–*Cla*I-Digest-Generated Clone Bank of Y-1 Genomic DNA in *E. coli*

```
5'-CAA  ACA  TTG  TCT  GGT  TTG-3'        luxY4B
```

that hybridized with the probe under these conditions. Plasmid preparations from four of the colonies were made and partial nucleotide sequences determined by using *lux*Y4B as a primer. The nucleotide sequence from one of the plasmids, designated pMT12.1, allowed extension of the sequence of *lux*Y through the 3′ end of the coding region.

Intact *lux*Y was reconstructed from pMT12.1 and pMT23.2. Plasmid pMT12.1 was digested with *Xho*II, *Hin*dIII, and *Taq*I. The digest was mixed with pMT23.2 digested with *Hin*dIII and *Bam*HI and the mixture treated with T4 DNA ligase, producing the plasmid pMT2312 (see Figure 2). This plasmid was digested with *Kpn*I and *Sac*I and the *lux*Y insert transferred into pUC19 previously digested with the same enzymes, yielding plasmid pMTY4 (Figure 2).

Southern Blot Analysis. Southern transfers and blots were performed as previously described with minor modifications (Maniatis et al., 1982; Southern, 1975). Oligonucleotide probes were end-labeled by using T4 polynucleotide kinase and [γ-^{32}P]dATP. Probing and washing were carried out with the buffers suggested by Maniatis et al. (1982), but hybridization was carried out at 47 °C for 1 h, then 45 °C for 2 h, and 42 °C for the remaining 12–14 h. Washes were carried out as follows: four washes in 6× SSC and 0.1% SDS for 20 min each at room temperature, followed by two washes with the same buffer at 32 °C for 20 min each, and then two washes with the same buffer at 37 °C for 10 min each.

Western Blot Analysis. Cultures of *E. coli* (LE392) were grown in 5-mL aliquots at 18 °C for 48 h. The cells were pelleted by centrifugation and rinsed with 75 μL of 10% trichloroacetic acid. Pellets were resuspended in 100 μL of 125 mM Tris-HCl, pH 8.8, 2.5% SDS, and 2.5% 2-mercaptoethanol. The samples were subjected to Western blot analysis as described by Daubner et al. (1986); antibodies for probing the blots were prepared in rabbits injected with YFP by the methods of Crowle (1973).

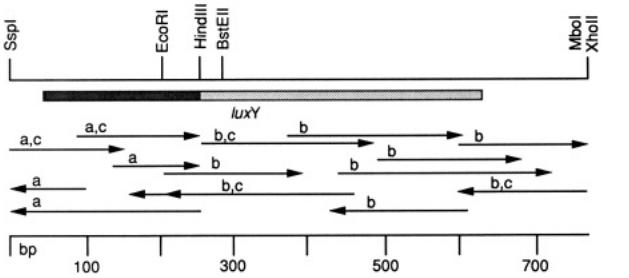

FIGURE 3: Strategy for determination of the sequence of *V. fischeri* Y-1 *lux*Y. Arrows indicate direction and length of the sequence determined. Notations above the arrows identify the plasmid that was used for that region's sequence determination: a refers to plasmid pMT23.1; b, plasmid pMT 12.1; and c, plasmid pMTY4. The bar labeled *lux*Y is shaded differentially to show the relative sizes of the inserts in pMT23.1 and pMT12.1. The dark shading represents the portion of the sequence contained in pMT23.1; the light shading, the sequence in pMT12.1.

DNA Sequencing and Sequence Analysis. The nucleotide sequence of both strands of plasmid DNA was determined by using modifications (Tabor & Richardson, 1987) of the dideoxynucleotide chain termination method (Sanger et al., 1977). DNA sequence data were analyzed by using software from Intelligenetics, Inc., on a Digital VAX Model 750 computer and DNA Inspector IIE from Textco, West Lebanon, NH, on a Macintosh II.

Results and Discussion

The construction of the plasmids used in the cloning of *lux*Y is depicted in Figure 2. We had previously determined the N-terminal amino acid sequence of YFP, Met-Phe-Lys-Gly-Ile-Val-Glu-Gly-Ile-Gly-Ile-Ile-Glu-Lys-Ile (Daubner et al., 1987). This amino acid sequence was scanned for the corresponding DNA sequence giving the longest possible least degenerate oligonucleotide mixture. Two mixtures of oligonucleotides, all 17-mers, were synthesized, the mixtures differing only in the bases at position 12 (Chart I). These mixtures were used separately to probe Southern blots containing restriction enzyme digests (*Hin*dIII, *Bam*HI, *Eco*RI, *Sal*I, and *Pst*I, used separately) of *V. fischeri* Y-1 genomic DNA. Only one mixture, MT-2, hybridized to fragments in the Southern blot and was therefore used to probe the *Hin*dIII library of Y-1 genomic DNA generated by using the vector pBR322 in *E. coli*. A plasmid containing a *Hin*dIII restriction fragment from the Y-1 genomic DNA library that hybridized with MT-2 (plasmid pMT23.1) was isolated and the nucleotide sequence of a portion of the insert determined. Sequence determinations performed on pMT23.1 with MT-2 as primer yielded information extending from base 105 of *lux*Y to base 200, the location of the *Hin*dIII site used in the cloning (see Figure 3). The next primer used for sequencing consisted of the sequence complementary to bases 117–131. This experiment yielded the sequence of the complementary strand extending beyond the 5′ end of the *lux*Y coding region. The first 45 bases of the *lux*Y coding sequence encoded an amino acid sequence that exactly matched the N-terminal amino acid sequence of YFP determined by Edman degradation. The first 17 bases of the coding region matched one of the sequences present in the hybridization probe mixture MT-2 (see Chart I). A putative ribosome binding site was found in the 5′ untranslated region. Since the sequence from the start of *lux*Y

```
                                                    50
AATATTTTTA  TTAATTCATT  AGAAAAATGA  GAGGAAGGAT TATT ATG TTT AAA GGT ATA GTA GAA GGT ATA
                                                    met-phe-lys-gly-ile-val-glu-gly-ile

                        100
GGA ATC ATT GAA AAA ATT GAT ATA TAT ACT GAC CTA GAT AAG TAT GCA ATT CGA TTT CCT GAA AAT
gly-ile-ile-glu-lys-ile-asp-ile-tyr-thr-asp-leu-asp-lys-tyr-ala-ile-arg-phe-pro-glu-asn

          150                                                       200
ATG TTG AAT GGA ATT AAA AAG GAG TCG TCA ATA ATG TTT AAC GGA TGC TTC TTA ACG GTA ACT AGC
met-leu-asn-gly-ile-lys-lys-glu-ser-ser-ile-met-phe-asn-gly-cys-phe-leu-thr-val-thr-ser

                                                250
GTG AAT TCA AAC ATT GTC TGG TTT GAT ATA TTT GAA AAA GAA GCA CGT AAG CTT GAT ACT TTT CGG
val-asn-ser-asn-ile-val-trp-phe-asp-ile-phe-glu-lys-glu-ala-arg-lys-leu-asp-thr-phe-arg

                    300
GAA TAT AAG GTA GGT GAC CGA GTA AAT TTA GGA ACA TTC CCA AAA TTT GGC GCT GCA TCT GGT GGG
glu-tyr-lys-val-gly-asp-arg-val-asn-leu-gly-thr-phe-pro-lys-phe-gly-ala-ala-ser-gly-gly

          350                                                       400
CAT ATA TTA TCA GCA AGG ATT TCA TGT GTA GCA AGT ATT ATT GAA ATA ATA GAA AAT GAG GAT TAT
his-ile-leu-ser-ala-arg-ile-ser-cys-val-ala-ser-ile-ile-glu-ile-ile-glu-asn-glu-asp-tyr

                                                450
CAA CAA ATG TGG ATT CAA ATT CCT GAA AAT TTT ACA GAG TTT CTT ATT GAT AAA GAC TAT ATT GCT
gln-gln-met-trp-ile-gln-ile-pro-glu-asn-phe-thr-glu-phe-leu-ile-asp-lys-asp-tyr-ile-ala

                    500
GTG GAT GGT ATT AGC TTA ACT ATT GAC ACT ATA AAA AAC AAC CAA TTT TTC ATT AGT TTA CCC TTA
val-asp-gly-ile-ser-leu-thr-ile-asp-thr-ile-lys-asn-asn-gln-phe-phe-ile-ser-leu-pro-leu

          550
AAA ATA GCA CAA AAT ACA AAT ATG AAA TGG CGA AAA AAA GGT GAT AAG GTA AAT GTT GAG TTA TCA
lys-ile-ala-gln-asn-thr-asn-met-lys-trp-arg-lys-lys-gly-asp-lys-val-asn-val-glu-leu-ser

                                                650
AAC AAA ATT AAT GCT AAC CAG TGT TGG TAA T TTACTGAGGA TAGTAAAAAT GAACTGTTTA AAATAATATT T
asn-lys-ile-asn-ala-asn-gln-cys-trp-STOP

                    700
AAATTTTTA  TTTATAATAC  AGAGTCAGTT  GTTGTAAATA  GTCTGAGTGG  TAAATAAGTT  CTACCATTAA  TTAA

   750                 •771
ATATTA  TCCATATTAA  ATAAAGGATC  T
```

FIGURE 4: Complete nucleotide sequence of the *lux*Y gene from *V. fischeri* Y-1 and the encoded amino acid sequence of the yellow fluorescent protein. A possible ribosome binding site is underlined. The *Eco*RI and *Hin*dIII sites are indicated by italic type. Numbering of the bases is from the first base of the *Ssp*I site.

to the *Hin*dIII site was too short (208 bases) to encode the entire protein, an overlapping fragment was cloned. An oligonucleotide corresponding to bases 167–184 of *lux*Y (bases 211–228 in Figure 4) was synthesized (*lux*Y4B, Chart II). A second Y-1 genomic DNA library, generated by digestion with *Eco*RI and *Cla*I, was prepared for the isolation of the remainder of the gene. The oligonucleotide *lux*Y4B hybridized with a fragment in the *Eco*RI–*Cla*I-generated Y-1 genomic DNA library. Sequence analysis revealed that the *lux*Y coding region in this second fragment overlapped the coding region in the previously isolated *Hin*dIII fragment; the two were recombined by using the *Hin*dIII site they had in common, giving plasmid pMT2312. Plasmid pMT2312 was treated with *Kpn*I and *Sac*I and the *lux*Y gene transferred into pUC19, yielding plasmid pMTY4.

The strategy by which the gene was sequenced is presented in Figure 3. Since the *lux*Y gene was isolated in two parts (with the *Eco*RI–*Hin*dIII fragment in common) and reassembled, primers that allowed sequencing across the point where our two clones were recombined were used twice, once for sequencing in pMT23.2 or pMT12.1 and once for sequencing in pMTY4. Also, the 3′ and 5′ ends were sequenced twice, first in pMT23.2 or pMT12.1 and then again in pMTY4, to assure that no modifications had occurred during cloning. The complete nucleotide sequence of *lux*Y and the encoded amino acid sequence of YFP are given in Figure 4. The sequence shown is of a 771-base-pair region extending from the *Ssp*I site upstream of the coding region through the *Hin*dIII site (position 252) that was used in reconstruction of *lux*Y to the *Xho*II downstream of the YFP coding region. This fragment had a single long open reading frame encoding a 22 211-dalton polypeptide composed of 194 amino acid residues. The amino acid sequence of YFP was compared to other protein sequences contained in the GenBank database. No complete or partial similarities were found to any protein in the current contents of the database. The distribution of hydrophobic and hydrophilic residues in the encoded amino acid sequence is suggestive of a globular protein. The only striking feature of the amino acid sequence is the high level of isoleucine in the protein. The high level of isoleucine may result from the same factors that render the *lux*Y gene high in content of A + T residues (see discussion below).

The amino acid compositions of lumazine protein (Small et al., 1980) and of YFP predicted from our DNA sequence are compared in Table I. Although both proteins bind heterocyclic fluorescent molecules as prosthetic groups, interact with luciferases from various species of bioluminescent bacteria, and have a monomer molecular weight of about 22 000, they are very different in amino acid composition. In particular, very striking differences between the lumazine protein and YFP are seen in content of valine, isoleucine, tryptophan, and phenylalanine. From these data we conclude that the

Table I: Comparison of Amino Acid Compositions of Lumazine Protein of *P. phosphoreum* and Yellow Fluorescent Protein of *V. fischeri* Strain Y-1

amino acid	mol/22 000 g of lumazine protein[a]	mol/mol of YFP subunit
alanine	4	9
arginine	3	6
aspartate/asparagine	28	
asparagine		16
aspartate		12
cysteine	2	3
glutamate/glutamine	17	
glutamate		13
glutamine		6
glycine	17	12
histidine	2	1
isoleucine	20	27
leucine	14	11
lysine	12	17
methionine	1	5
phenylalanine	5	13
proline	4	4
serine	10	11
threonine	13	9
tryptophan		4
tyrosine	2	5
valine	22	10

[a] Small et al. (1980).

similarities in the functions and molecular weights of lumazine protein and YFP are probably coincidental.

The base composition of the coding region of *lux*Y is remarkably rich in adenine and thymine (71% AT). This is also a feature of other genes that have been studied from *V. fischeri* (ATCC 7744): *lux*R and *lux*I, 69% AT, and *lux*A and *lux*B, 67% AT (Devine et al., 1988; Foran & Brown, 1988; Baldwin et al., 1989). In comparison, the *lux*A and *lux*B genes in *V. harveyi* contain 55% adenine and thymine (Johnston et al., 1986).

The heavy bias toward use of AT is mirrored in the codon usage of *V. fischeri*. Table II presents the complete codon usage for the *V. fischeri* ATCC 7744 genes *lux*CDABE, plus *lux*R and *lux*I (Baldwin et al., 1989), *V. harveyi lux*AB (Johnston et al., 1986), *V. fischeri* strain Y-1 *lux*Y, and the genes of the *E. coli str* operon (Post & Nomura, 1980). The *str* operon is included to enable comparison to a codon usage that is based on the abundance of isoaccepting tRNA species in *E. coli*. For some amino acids that can be encoded by two codons, *V. fischeri* displays a strong bias toward the codon with an A or T in the third position. For instance, in encoding phenylalanine, the *V. fischeri lux* region used TTT 94 times and TTC only 14 times; similar bias is seen in the coding for lysine, asparagine, aspartate, tyrosine, glutamate, and glutamine (Table II). *V. harveyi* shows the same preference only when coding for glutamate. For some amino acids that can be encoded by more than two codons, such as threonine, valine, alanine, and isoleucine, the genes of *V. fischeri* heavily favor the use of codons with A and T in the third position over use of codons with G and C. Neither the *E. coli str* operon nor *V. harveyi* appears to favor either G and C or A and T in that position.

YFP can be expressed in *E. coli* as shown in Figure 5. Western blot analysis showed that *E. coli* transformed with pMTY4 contained a protein that reacted with antibody raised to authentic YFP purified from *V. fischeri* Y-1. A standard of purified YFP, which reacted strongly with the antibody, and a sample of lysate from *E. coli* transformed with pUC19, which does not react with the antibody, are shown on the same blot.

Table II: Codon Usage for *V. fischeri* Y-1 *luxY*, the *V. fischeri lux* Regulon, *V. harveyi luxA* and *luxB*, and the *E. coli str* Operon[a]

	A	*luxY*	*Vf*	*Vh*	*Ec*		T	*luxY*	*Vf*	*Vh*	*Ec*		C	*luxY*	*Vf*	*Vh*	*Ec*		G	*luxY*	*Vf*	*Vh*	*Ec*
A	AAA (Lys)	12	**125**	(22)	124*		ATA (Ile)	10	**46**	(0)	21*		ACA (Thr)	3	**54**	(7)	19*		AGA (Arg)	0	**30**	(4)	10*
	AAT (Asn)	10	**125**	(20)	29*		ATT (Ile)	16	**118**	(14)	45*		ACT (Thr)	5	**48**	(12)	60*		AGT (Ser)	2	**45**	(7)	18*
	AAC (Asn)	6	**27**	(19)	73*		ATC (Ile)	1	**17**	(17)	80*		ACC (Thr)	0	**14**	(8)	63*		AGC (Ser)	2	**15**	(4)	32*
	AAG (Lys)	5	**32**	(13)	62*		ATG (Met)	5	**59**	(18)			ACG (Thr)	1	**15**	(9)	21*		AGG (Arg)	1	**7**	(1)	6*
T	TAA (stop)						TTA (Leu)	7	**101**	(7)	27*		TCA (Ser)	5	**41**	(9)	21*		TGA (stop)				
	TAT (Tyr)	5	**82**	(9)	23*		TTT (Phe)	10	**94**	(24)	48*		TCT (Ser)	1	**40**	(12)	51*		TGT (Cys)	2	**2**	(12)	10*
	TAC (Tyr)	0	**15**	(17)	52*		TTC (Phe)	3	**14**	(16)	34*		TCC (Ser)	0	**8**	(2)	53*		TGC (Cys)	1	**1**	(2)	18*
	TAG (stop)						TTG (Leu)	1	**21**	(17)	20*		TCG (Ser)	1	**13**	(5)	31*		TGG (Trp)	4	**30**	(8)	29*
C	CAA (Gln)	5	**52**	(14)	42*		CTA (Leu)	1	**19**	(10)	18*		CCA (Pro)	1	**33**	(11)	27*		CGA (Arg)	3	**22**	(8)	15*
	CAT (His)	1	**1**	(11)	15*		CTT (Leu)	2	**29**	(6)	26*		CCT (Pro)	2	**35**	(6)	17*		CGT (Arg)	1	**25**	(8)	89*
	CAC (His)	0	**8**	(12)	31*		CTC (Leu)	0	**11**	(5)	36*		CCC (Pro)	1	**5**	(0)	19*		CGC (Arg)	0	**7**	(5)	59*
	CAG (Gln)	1	**16**	(13)	71*		CTG (Leu)	0	**8**	(7)	117*		CCG (Pro)	0	**6**	(3)	58*		CGG (Arg)	1	**4**	(0)	14*
G	GAA (Glu)	9	**108**	(38)	92*		GTA (Val)	6	**47**	(7)	68*		GCA (Ala)	6	**43**	(16)	71*		GGA (Gly)	4	**39**	(4)	14*
	GAT (Asp)	8	**124**	(26)	61*		GTT (Val)	1	**74**	(15)	90*		GCT (Ala)	3	**47**	(11)	108*		GGT (Gly)	6	**48**	(22)	101*
	GAC (Asp)	4	**32**	(24)	65*		GTC (Val)	1	**20**	(11)	40*		GCC (Ala)	0	**9**	(11)	51*		GGC (Gly)	1	**7**	(12)	72*
	GAG (Glu)	4	**43**	(12)	58*		GTG (Val)	2	**14**	(14)	61*		GCG (Ala)	0	**26**	(20)	79*		GGG (Gly)	1	**21**	(5)	21*

[a] Plain text numbers refer to the codon usage for *luxY* from *V. fischeri* Y-1, numbers in boldface type refer to codon usage in the *V. fischeri* regulon [*lux*CDABE, *lux*R, and *lux*I (Devine et al., 1988; Foran & Brown, 1988; Baldwin et al., 1989)], numbers in parentheses refer to the codon usage for *V. harveyi lux*A and *lux*B (Johnston et al., 1986), and numbers marked with an asterisk refer to codon usage in then *E. coli str* operon (Post & Nomura, 1980).

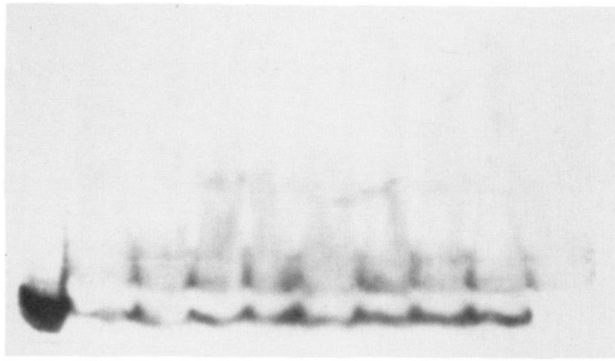

FIGURE 5: Western blot analysis of a polyacrylamide gel, showing antibody recognition of purified YFP and YFP produced in *E. coli* carrying the plasmid pMTY4. Slot a contains 1.6 μg of pure YFP. Slots b–i contain total cellular proteins from 5-mL cultures of *E. coli* carrying separate isolates of the plasmid pMTY4. Slot j is the control exhibiting proteins produced by *E. coli* carrying the plasmid pUC19. Cultures displayed in slots b–i produced a protein that had the same electrophoretic mobility as authentic YFP and reacted with antibody produced by using authentic YFP as antigen. This protein was not present in cells that did not contain the pMTY4 plasmid, leading us to conclude that the plasmid pMTY4 encodes production of authentic YFP in *E. coli*.

The *lux*Y gene does not appear to reside within the *lux* regulon of *V. fischeri* Y-1. The sequence determined upstream of *lux*Y extended over 150 base pairs, well beyond the *Ssp*I site limit of Figure 4, while the sequence determined downstream extended over 210 base pairs, about 100 base pairs beyond the *Xho*I site limit of Figure 4, without an open reading frame. In contrast, the genes within the *lux* regulon of *V. fischeri* ATCC7744 are closely spaced. There is no apparent homology between any of the sequence reported here and any of the *V. fischeri* ATCC 7744 regulon, which has been completely sequenced (Baldwin et al., 1989). There were no obvious promoter elements upstream of the *lux*Y coding region and no fusions have been constructed to elucidate a *lux*Y promoter. The location of the *lux*Y gene relative to the *lux* regulon of *V. fischeri* Y-1 therefore remains unknown, as does the mechanism that regulates expression of this apparently integral component of the bioluminescence system of *V. fischeri* Y-1.

Summary

We have cloned and sequenced the *lux*Y gene of *V. fischeri* Y-1, which encodes the yellow fluorescent protein, a protein of M_r 22211. *E. coli* cells carrying the *lux*Y gene on a plasmid produce a protein that reacts with antibody to authentic YFP and has the same electrophoretic mobility in sodium dodecyl sulfate gels as YFP. The amino acid composition of YFP is quite different from that of the lumazine protein of *P. phosphoreum*. Furthermore, the amino acid sequence of YFP does not show any similarity to other proteins sequenced to date. The AT-enriched nature of the *lux*Y gene and the *V. fischeri* ATCC 7744 *lux* regulon seems to lead to a codon usage quite different from that of *V. harveyi*.

Acknowledgments

We thank Lawrence Chlumsky for his assistance with certain aspects of the cloning work and Drs. Miriam Ziegler and Paul Fitzpatrick for critical reading of the manuscript.

References

Baldwin, T. O., Berends, T., Bunch, T. A., Holzman, T. F., Rausch, S. K., Shamansky, L., Treat, M. L., & Ziegler, M. M. (1984) *Biochemistry 23*, 3663–3667.

Baldwin, T. O., Devine, J. H., Heckel, R. C., Lin, J.-W., & Shadel, G. S. (1989) *J. Biolumin. Chemilumin. 4*, 326–341.

Cline, T. W., & Hastings, J. W. (1972) *Biochemistry 11*, 3359–3369.

Cline, T. W., & Hastings, J. W. (1974) *J. Biol. Chem. 249*, 4668–4669.

Crowle, A. J. (1973) *Immunodiffusion*, Academic Press, New York.

Daubner, S. C., & Baldwin, T. O. (1989) *Biochem. Biophys. Res. Commun. 161*, 1191–1198.

Daubner, S. C., Young, M., Sammons, R. D., Courtney, L. F., & Benkovic, S. J. (1986) *Biochemistry 25*, 2951–2957.

Daubner, S. C., Astorga, A. M., Leisman, G. B., & Baldwin, T. O. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8912–8916.

Devine, J. H., Countryman, C., & Baldwin, T. O. (1988) *Biochemistry 27*, 837–842.

Eberhard, A., Burlingame, A., Eberhard, C., Kenyon, G. L., Nealson, K. H., & Oppenheimer, N. J. (1981) *Biochemistry 20*, 2444–2449.

Engebrecht, J., & Silverman, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4154–4158.

Engebrecht, J., & Silverman, M. (1986) *Genet. Eng. 8*, 31–44.

Engebrecht, J., Nealson, K., & Silverman, M. (1983) *Cell (Cambridge, Mass.) 32*, 773–781.

Foran, D. R., & Brown, W. M. (1988) *Nucleic Acids Res. 16*, 777.

Gast, R., & Lee, J. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 833–837.

Holmes, D. S., & Quigley, M. (1981) *Anal. Biochem. 114*, 193.

Johnston, T. C., Thompson, R. B., & Baldwin, T. O. (1986) *J. Biol. Chem. 261*, 4805–4811.

Kaplan, H. B., Eberhard, A., Widrig, C., & Greenberg, E. P. (1985) *J. Labelled Compd. Radiopharm. 22*, 387–395.

Koka, P., & Lee, J. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3068–3072.

Lee, J. (1982) *Photochem. Photobiol. 36*, 689–697.

Leisman, G. B. (1981) Ph.D. Thesis, University of California, San Diego.

Leisman, G., & Nealson, K. H. (1982) in *Flavins and Flavoproteins* (Massey, V., & Williams, C. H., Eds.) pp 383–386, Elsevier, Amsterdam.

Macheroux, P., Schmidt, K. U., Steinerstauch, P., Ghisla, S., Colepicolo, P., Buntic, R., & Hastings, J. W. (1987) *Biochem. Biophys. Res. Commun. 146*, 101–106.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mitchell, G. W., & Hastings, J. W. (1974) *J. Biol. Chem. 244*, 2572–2576.

Nealson, K. H. (1978) *Methods Enzymol. 17*, 153–166.

Nealson, K. H., Platt, T., & Hastings, J. W. (1970) *J. Bacteriol. 104*, 313–322.

Post, L. E., & Nomura, M. (1980) *J. Biol. Chem. 255*, 4660–4666.

Rosson, R. A., & Nealson, K. H. (1981) *Arch. Microbiol. 129*, 299–304.

Ruby, E. G., & Nealson, K. H. (1977) *Science 196*, 432–434.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Sato & Miura (1963) *Biochim. Biophys. Acta 72*, 619–629.
Small, E. D., Koka, P., & Lee, J. (1980) *J. Biol. Chem. 255*, 8804–8810.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
Tabor, S., & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4767–4771.
Ziegler, M. M., & Baldwin, T. O. (1981) *Curr. Top. Bioenerg. 12*, 65–113.

Identification of the NADH-Binding Subunit of NADH–Ubiquinone Oxidoreductase of *Paracoccus denitrificans*†

Takao Yagi* and Tri M. Dinh
Division of Biochemistry, Department of Molecular and Experimental Medicine, Research Institute of Scripps Clinic, La Jolla, California 92037

Received November 27, 1989; Revised Manuscript Received February 16, 1990

ABSTRACT: The NADH dehydrogenase complex isolated from *Paracoccus denitrificans* is composed of approximately 10 unlike polypeptides and contains noncovalently bound FMN, non-heme iron, and acid-labile sulfide [Yagi, T. (1986) *Arch. Biochem. Biophys. 250*, 302–311]. When the *Paracoccus* NADH dehydrogenase complex was irradiated by UV light in the presence of [*adenylate*-^{32}P]NAD, radioactivity was incorporated exclusively into one of three polypeptides of M_r ~50 000. Similar results were obtained when [*adenylate*-^{32}P]NADH was used. The labeling of the M_r 50 000 polypeptide was diminished when UV irradiation of the enzyme with [*adenylate*-^{32}P]NAD was performed in the presence of NADH, but not in the presence of NADP(H). The labeled polypeptide was isolated by preparative sodium dodecyl sulfate gel electrophoresis and was shown to cross-react with antiserum to the NADH-binding subunit (M_r = 51 000) of bovine NADH–ubiquinone oxidoreductase. Its amino acid composition was also very similar to that of the bovine NADH-binding subunit. These chemical and immunological results indicate that the M_r 50 000 polypeptide is an NADH-binding subunit of the *Paracoccus* NADH dehydrogenase complex.

Mitochondrial NADH–ubiquinone oxidoreductase (complex I)[1] bears coupling site 1 of oxidative phosphorylation and is composed of more than 25 unlike polypeptides (Hatefi, 1985; Yagi, 1989; Ragan, 1987). This complexity has hampered progress in the study of many aspects of the structure and function of mitochondrial complex I (Hatefi, 1985; Yagi, 1989; Ragan, 1987). In an effort to find a simpler system with which to clarify the structure and function of NADH–Q oxidoreductase, we have purified the NADH–Q oxidoreductase from *Paracoccus denitrificans* membranes (Yagi, 1986) which bear coupling site 1 (Stouthamer, 1980) and exhibit similar EPR signals to the mammalian complex I (Albracht et al., 1980; Meinhardt et al., 1987). The NADH dehydrogenase complex purified from *Paracoccus* membranes is composed of 10 unlike polypeptides and contains noncovalently bound FMN and multiple iron–sulfur clusters (Yagi, 1986), suggesting that the *Paracoccus* system is structurally simpler than its mammalian counterpart, as has been shown also for other energy-transducing enzyme complexes of the respiratory chain [cytochrome oxidase (Ludwig & Schatz, 1980; Haltia et al., 1988) and ubiquinol–cytochrome *c* oxidoreductase (Yang & Trumpower, 1986)]. Recently, we (Yagi et al., 1988) have isolated the NADH–Q oxidoreductase of *Thermus thermophilus* HB-8 which bears coupling site 1 (Meinhardt et al., 1990). The *Thermus* NADH–Q oxidoreductase, which is partially sensitive to rotenone, is also composed of 10 unlike polypeptides, although EPR signals of the *Thermus* NADH–Q oxidoreductase segment are different from those of mammalian complex I (Meinhardt et al., 1990).

In the case of bovine complex I, some of the subunits have been characterized (Hatefi, 1985; Yagi, 1987, 1989; Ragan, 1987; Yagi & Hatefi, 1988; Fearnley et al., 1989). For example, Chen and Guillory (1981) have shown that the M_r 51 000 polypeptide of bovine complex I is the NADH-binding subunit, using a tritiated photoaffinity NAD analogue. Recently, it was shown by our laboratory (Yagi, 1987; Yagi & Hatefi, 1988) that DCCD inhibits the NADH–Q oxidoreductases of various organisms bearing energy coupling site 1, and the DCCD-binding subunit of bovine complex I was assigned to the mitochondrial ND-1 gene product. Studies of Earley et al. (1987) have suggested that the rotenone-binding subunit of complex I is also the ND-1 gene product. However, nothing is known about the characteristics of the subunits of the NADH–Q oxidoreductases of bacteria bearing coupling site 1. Therefore, it was of interest to elucidate the function of various polypeptides in *Paracoccus* NADH dehydrogenase complex.

This paper describes the identification of the NADH-binding subunit of the NADH dehydrogenase complex from *Paracoccus denitrificans*. UV irradiation of the NADH dehydrogenase complex with [^{32}P]NAD or [^{32}P]NADH resulted in incorporation of radioactivity into a single polypeptide of M_r 50 000. The M_r 50 000 protein band was protected from [^{32}P]NAD labeling in the presence of NADH, but not in the presence of NADP(H). In addition, antiserum to the NADH-binding subunit of bovine complex I cross-reacted with

† This work was supported by U.S. Public Health Service Grant GM33712 to T.Y. This is Publication 5977-BCR from the Research Institute of Scripps Clinic, La Jolla, CA.
* To whom correspondence should be addressed.

[1] Abbreviations: complex I, NADH–quinone oxidoreductase bearing energy coupling site; Q, quinone; EPR, electron paramagnetic resonance; DCCD, *N,N'*-dicyclohexylcarbodiimide; SDS, sodium dodecyl sulfate; IEF, isoelectric focusing; FP, IP, and HP, flavoprotein, iron–sulfur protein, and hydrophobic protein fractions of complex I, respectively; K_i, inhibition constant.

0006-2960/90/0429-5515$02.50/0