Vol. 155, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1988 Pages 351-358

A NEW LUX GENE IN BIOLUMINESCENT BACTERIA CODES FOR A PROTEIN HOMOLOGOUS TO THE BACTERIAL LUCIFERASE SUBUNITS

Robert R. Soly, Joseph A. Mancini, Stefano R. Ferri, Michael Boylan, and Edward A. Meighen

Department of Biochemistry
McGill University
3655 Drummond Street
Montreal, Quebec, Canada H3G 1Y6

Received July 20, 1988

SUMMARY: The nucleotide sequence of a new gene, luxF, located between the luxB and E genes in the bioluminescent system of Photobacterium phosphoreum has been determined. The luxF gene codes for a polypeptide of 231 amino acids which is homologous to the α and β subunits of luciferase coded by the luxA and luxB genes, respectively. The degree of homology of the luxF protein is very high with the β subunit of luciferase (~ 30% identity) with greatest similarity to the Vibrio luxB proteins. The luxF gene appears to have evolved by duplication of the luxB gene followed by deletion of approximately 100 codons just penultimate to the 5'-terminal. The close homology with the luciferase β subunit implicates the luxF protein in a function related to the light-emitting reaction.

Light emission in luminescent bacteria is catalyzed by a heteropolymeric enzyme, luciferase, which consists of two nonidentical subunits, α and β, coded by the luxA and luxB genes, respectively (1-4). The nucleotide sequences of the luxA and B genes from Vibrio harveyi (5,6) and V. fischeri, MJ1 (7) have been determined and the α and β luciferase subunits shown to be homologous with a 29-31% identity in amino acid sequence. In addition to the luxA and B genes, three other genes, luxC,D, and E are needed for synthesis of enzymes involved in producing long chain fatty aldehydes (1,8-10), which along with $FMNH_2$ and O_2 are the substrates in the bioluminescence reaction (11).

By transformation of Escherichia coli with lux DNA, these genes have been shown to be located in close proximity and in the order, C,D,A,B and E in V. harveyi (10) and V. fischeri (8,9). Moreover, these genes have been shown to be part of a single operon in V. fischeri (1) and analyses of the lux polycistronic mRNAs in V. harveyi supports this conclusion (10).

Two additional genes, luxI and R, have also been identified in V. fischeri and are believed to be associated with the cell density-dependent

0006-291X/88 $1.50

regulation of light emission operating in most species of luminescent bacteria (1). LuxI is located immediately upstream of luxC and belongs to the same operon while luxR constitutes an independent operon which is upstream of luxI and transcribed in the opposite direction. Although the organization of the lux structural genes is similar in V. harveyi and V. fischeri, it has recently been shown that V. harveyi lacks genes immediately upstream of luxC where luxI would be expected to be located (12). These results suggest that the luminescent systems in marine bacteria have diverged to some degree.

Recently, the Photobacterium phosphoreum (NCMB844) lux system has been cloned into E. coli and the gene organization determined (13). Protein expression of different regions of the DNA has revealed the presence of a yet uncharacterized gene, luxF, that codes for a 26 kDa polypeptide and appears to be part of the lux system in this bacterium. This gene is located between luxB and luxE and disrupts the structural gene organization common to V. fischeri and V. harveyi. In this paper, we report the complete nucleotide sequence of luxF and demonstrate that the amino acid sequence coded by this gene is homologous with the amino acid sequences of the luciferase subunits.

MATERIALS AND METHODS

P. phosphoreum DNA with the luxF gene was isolated from a SP64 recombinant plasmid containing a 2.5 kbp HindIII fragment from the P. phosphoreum lux system (13). The HindIII insert was isolated, restricted with different restriction endonucleases, and the fragments ligated into the M13mp18 and M13mp19 sequencing vectors (14,15).

The sequencing reactions were performed by the dideoxy chain termination method using the T7 DNA polymerase provided with the Sequenase DNA sequencing kit purchased from United States Biochemicals. [^{35}S]Deoxyadenosine 5'-(α-thio) triphosphate (1400 Ci/mmol) was obtained from New England Nuclear. By separation of the polynucleotide products for up to 16 hours on 6% polyacrylamide gels, nucleotide sequences of up to 600 bases could be determined.

Analysis of data was performed using the programs of DNASIS and PROSIS provided by Hitachi Software Engineering Co., Ltd.

RESULTS

DNA containing the luxF gene of P. phosphoreum and extending upstream into the luxB gene and downstream into the luxE gene was restricted and the fragments ligated in M13 sequencing vectors. The DNA sequence was determined in both directions according to the strategy in Figure 1.

The nucleotide and predicted amino acid sequence of the luxF gene is shown in Figure 2. The open reading frame corresponding to luxF is 693 nucleotides in length (231 amino acid residues) and codes for a 26.6 kDa polypeptide in excellent agreement with the molecular weight determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (13). Immediately

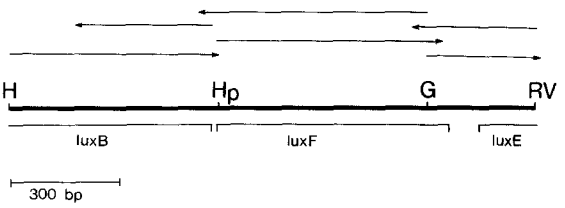


Figure 1. Sequencing strategy for the luxF gene of P. phosphoreum. The map of DNA shows the location of the luxF gene flanked upstream by luxB and downstream by luxE. The arrows above the map represent the sequences obtained from various clones in both directions. Restriction sites shown are HindIII (H), HpaI (Hp), BglII (G), and EcoRV (RV).

```
GTT GTA AAA GTG ATT AAT ATG GTT AAT GAG AAG ATT CAA AAG AAT TTA CCA AGC TCG TAA
Val Val Lys Val Ile Asn Met Val Asn Glu Lys Ile Gln Lys Asn Leu Pro Ser Ser ***

GTGTAAAGGAAGCGGTGTTATT  ATG AAT AAA TGG AAT TAC GGA GTC TTC TTC GTT AAC TTT TAT
                        Met Asn Lys Trp Asn Tyr Gly Val Phe Phe Val Asn Phe Tyr
                          1                                          10

AAT AAA GGC CAA CAA GAG CCA TCA AAA ACG ATG AAT AAT GCA TTA GAA ACA TTA CGT ATT
Asn Lys Gly Gln Gln Glu Pro Ser Lys Thr Met Asn Asn Ala Leu Glu Thr Leu Arg Ile
                      20                                         30

ATT GAT GAA GAT ACA TCT ATT TAT GAT GTG ATT AAT ATT GAT GAC CAC TAT CTT GTA AAG
Ile Asp Glu Asp Thr Ser Ile Tyr Asp Val Ile Asn Ile Asp Asp His Tyr Leu Val Lys
                      40                                         50

AAA GAC AGT GAA GAT AAA AAG CTA GCG TCT TTT ATT ACA CTA GGA GAA AAA CTA TAT GTG
Lys Asp Ser Glu Asp Lys Lys Leu Ala Ser Phe Ile Thr Leu Gly Glu Lys Leu Tyr Val
                      60                                         70

CTT GCT ACC AGT GAA AAC ACA GTT GAT ATT GCA GCG AAA TAT GCA TTA CCG TTA GTT TTC
Leu Ala Thr Ser Glu Asn Thr Val Asp Ile Ala Ala Lys Tyr Ala Leu Pro Leu Val Phe
                      80                                         90

AAA TGG GAT GAT ATA AAT GAG GAA CGA CTT AAA TTG TTG AGT TTT TAT AAT GCA TCC GCA
Lys Trp Asp Asp Ile Asn Glu Glu Arg Leu Lys Leu Leu Ser Phe Tyr Asn Ala Ser Ala
                     100                                        110

AGT AAA TAT AAC AAG AAT ATA GAT TTG GTT CGA CAC CAG CTT ATG TTA CAT GTC AAT GTT
Ser Lys Tyr Asn Lys Asn Ile Asp Leu Val Arg His Gln Leu Met Leu His Val Asn Val
                     120                                        130

AAT GAG GCA GAA ACT GTA GCA AAA GAA GAA CTC AAA TTA TAT ATT GAA AAC TAT GTA GCA
Asn Glu Ala Glu Thr Val Ala Lys Glu Glu Leu Lys Leu Tyr Ile Glu Asn Tyr Val Ala
                     140                                        150

TGT ACA CAG CCT AGT AAT TTT AAT GGC TCG ATT GAT AGT ATT ATT CAG AGT AAC GTG ACA
Cys Thr Gln Pro Ser Asn Phe Asn Gly Ser Ile Asp Ser Ile Ile Gln Ser Asn Val Thr
                     160                                        170

GGG AGT TAT AAA GAC TGT TTG TCA TAT GTA GCG AAT CTT GCT GGT AAA TTT GAT AAT ACT
Gly Ser Tyr Lys Asp Cys Leu Ser Tyr Val Ala Asn Leu Ala Gly Lys Phe Asp Asn Thr
                     180                                        190

GTG GAC TTC TTA CTT TGT TTT GAG TCA ATG CAA GAT CAA AAT AAG AAA AAA TCA GTA ATG
Val Asp Phe Leu Leu Cys Phe Glu Ser Met Gln Asp Gln Asn Lys Lys Lys Ser Val Met
                     200                                        210

ATA GAT CTT AAT AAT CAA GTT ATT AAG TTC CGC CAA GAT AAT AAT CTA ATC TAA TCT ACA
Ile Asp Leu Asn Asn Gln Val Ile Lys Phe Arg Gln Asp Asn Asn Leu Ile ***
                     220                                        230
```

Figure 2. Complete nucleotide and predicted amino acid sequence of the luxF gene of P. phosphoreum. The numbers refer to the amino acid positions of luxF. The end of the luxB gene is shown upstream. (The coding region of luxE begins 84 nucleotides downstream from the end of the luxF coding sequence, data not shown).

upstream of the ATG initiation codon is a clearly recognizable Shine-Delgarno sequence which is part of a 25 bp intergenomic region between the luxB and luxF genes. Another open reading frame is found 84 bp downstream of the luxF coding region and corresponds to the luxE gene (data not given).

Comparison of the amino acid sequence coded by luxF to the known amino acid sequences of the luciferase subunits (α and β) encoded by the luxA and luxB genes from V. fischeri MJ1 and V. harveyi shows that these proteins are homologous (Figure 3). In particular, the luxF gene product has a very high

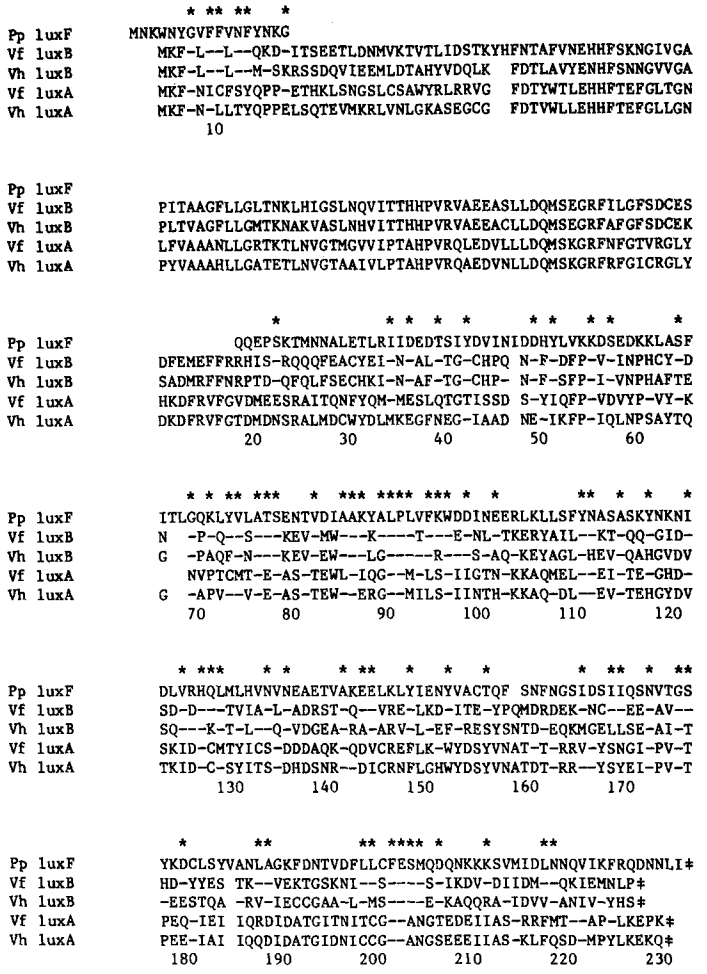


Figure 3. Comparison of the luxF amino acid sequence with the luxA and luxB proteins. The numbered positions refer to the luxF sequence. Blank spaces were introduced in the sequences as shown to permit alignment showing maximum homology, including the separation of the luxF protein into two fragments between amino acids 17 and 18. Twenty-eight residues were removed from the luxA proteins (amino acids 259 to 286 in Vf luxA) to facilitate alignment with the other sequences. Residues in the luxA and B proteins which are identical to the corresponding amino acid of luxF are shown as a horizontal bar (—). Residues identical between luxF and Vf luxB are further highlighted by the asterisk (*). Abbreviations: Vf, V. fischeri MJ1; Vh, V. harveyi; Pp, P. phosphoreum NCMB844; #, termination codons. Amino acid sequences of Vf luxA and luxB are from (7), Vh luxA from (5) and Vh luxB from (6).

homology to the β subunit of luciferase encoded by luxB. Except for a short region at the amino terminal, the majority of the luxF protein sequence of 231 residues is homologous to the carboxyl end of the luciferase β subunits from the Vibrio genus (324-326 residues). Consequently, a gap of 106 residues has been introduced between amino acids 17 and 18 of the luxF protein to demonstrate the high degree of homology between the proteins coded by the luxB and luxF genes. As the luxA gene product (α subunit of luciferase) is larger than the β subunit, 28 amino acid residues (positions 259-286 in V. fischeri luxA) were deleted to afford maximum homology between luxA,B and F. Under these alignment conditions, the carboxyl terminals of all three gene products end within a few amino acid residues of one another.

The degree of homology between the amino acid sequences coded by the luxF gene and the luxA and luxB genes from different species is given in Table I. The sequence homology between the luxF and luxB proteins is very high (29-34% identity) whereas only 18% of the protein sequence encoded by luxF is identical to the luxA proteins. Homology of the luxF gene to both the luxA and luxB genes is not unexpected since the α and β luciferase subunits have been shown to be very similar (29-31% identity) in both V. harveyi (5,6) and V. fischeri (7). Surprisingly, the homology of the luxF gene product to the P. phosphoreum luxB gene product is less than to the Vibrio luxB gene products possibly reflecting a strong conservation of the amino acid sequence of the luxF protein during evolution.

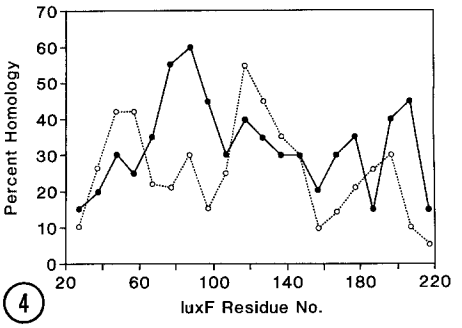
Comparison of the relative identity in the amino acid sequence between the V. fischeri α and β luciferase subunits to that between the luxF protein and the β subunit of V. fischeri luciferase is given in Figure 4. Over this region of sequence (luxF, residues 18-231; luxB, residues 121-326), the degree of homology between the luciferase subunits is only 24%, significantly less than found between the luxF and luxB proteins (31% identity). In the

TABLE I. Homology Between the LuxF Protein and the Luciferase Subunits[a]

Species	β Subunit (luxB)	α Subunit (luxA)
V. fischeri	34%	18%
V. harveyi	31%	18%
P. phosphoreum	29%[b]	

[a]The percentage homology across the highly conserved carboxyl regions (residues 34 to 226 of luxF) was calculated from the number of identical amino acids at the same positions aligned as given in Figure 3. The same relative values were obtained for the Vibrio proteins if the amino terminal regions were also included.

[b]The amino acid sequence coded by P. phosphoreum (NCMB844) luxB complementary to the luxF protein was determined during this work (unpublished data). Maximum homology was obtained using the same alignment with the luxF protein as the Vibrio luxB proteins.



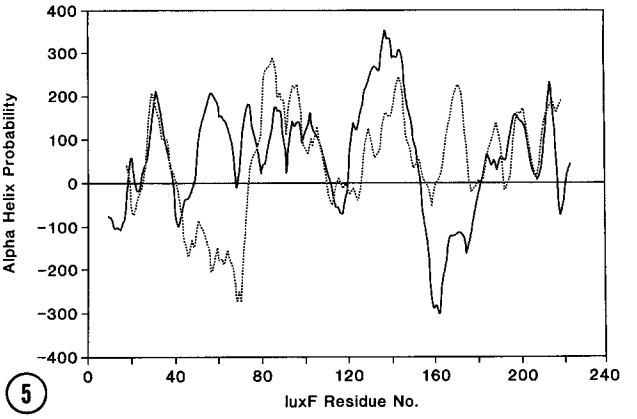


Figure 4. Comparison of the relative homology between the lux proteins. The percentages of identical amino acid residues between the luxF protein and Vf luxB protein (—) and between the Vf luxB and luxA proteins (---) in a window of 20 amino acid residues are plotted versus the residue number of luxF. Alignment of sequences is given in Figure 3.

Figure 5. Prediction of α-helix formation by the luxF and V. fischeri luxB proteins. The probability of helix formation by the luxF (——) and the luxB (....) proteins was computed by the method of Garnier et al. (16) and plotted versus the luxF amino acid position. The plots are aligned such that residue 18 of luxF corresponds to residue 118 of the luxB protein.

central (residues 70-100) and carboxyl terminal (residues 195-210) regions of luxF, the degree of homology is very high between the luxF and luxB proteins, reaching 60% identity by this analysis whereas the percent identity in these specific areas between the luxA and luxB proteins is less than 30%. The major area of homologies between the luciferase subunits occurs in their amino terminal regions and not in the sequences homologous to the luxF protein.

A comparison of the probability of forming an α-helix for residues 18-231 of the luxF protein and residues 118-326 of the V. fischeri luxB protein is given in Figure 5. Three regions have high structural homologies in terms of potential for forming an α-helix; a region extending from residues 18 to 40 (of luxF); a central region from residues 69 to 158 and a carboxyl terminal region from residues 185 to 214. The latter two regions of the luxF protein are also very similar to the corresponding areas in the luxB protein in relative hydrophobicity (data not given) whereas residues 18-40 of the luxF protein differs from the corresponding region in the luxB protein both in hydrophobicity and primary structure (Figure 3). The two regions which differ significantly in probability of forming an α-helix (residues 41 to 68 and 158 to 184) also differ in hydrophobicity (data not shown).

DISCUSSION

The lux structural genes, located in the order luxC,D,A,B,E in the Vibrio lux systems (8-10), are disrupted in the P. phosphoreum lux system by

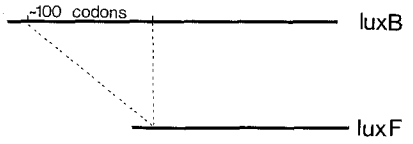


Figure 6. Relation between the luxB and luxF genes.

a new gene designated as luxF. This gene appears to have evolved by duplication of the luxB gene followed by a deletion of about 100 codons near the amino terminal end (Figure 6). A very high degree of homology was found between the central and carboxyl terminal regions of the luxF protein and the β subunit of luciferase in terms of primary structure, relative hydrophobicity, and potential capacity of α-helix formation. Alignment of these two regions, extending over 150 residues (69-226 of luxF), can be obtained without the need for any insertions or deletions, indicating that the luxF protein may retain important structural features possessed by the β-subunit of luciferase.

Although the level of conserved sequence upstream of residue 69 of luxF appears less significant, the strong homology at the amino terminal clearly indicates that the luxF gene arose by a large deletion within the luxB gene penultimate to the 5' terminal. The luxF gene would likely have evolved after the formation of luxA and luxB by gene duplication (2) in view of the greater similarity between the luxF and luxB proteins than between either the proteins coded by luxF and A or luxB and A. Interestingly, the luxF protein is more homologous to the luxB proteins from the Vibrio genus than to that from P. phosphoreum. This may reflect the relative rates of divergence of the luxF and luxB proteins from an ancestral form of the luciferase genes. In any case, it appears that the luxF gene was never produced in the Vibrio lux systems or was lost during evolution. Southern blot experiments support the absence of luxF elsewhere in the chromosomes of V. fischeri or V. harveyi (data not shown).

It appears that the luxF protein is not essential for light emission in all luminescent bacteria although it must play an essential role in the biochemistry, physiology, or ecology of the luminescent system in P. phosphoreum. Recently, the determination of the amino terminal sequences of the non-fluorescent flavoproteins (17,18) from P. leiognathi S1 and P. phosphoreum A_{13} has demonstrated that the first 22 residues of both flavoproteins are identical to the amino terminal sequence of the luxF protein with only two changes in the next 13 residues (D.J. O'Kane and J. Lee, personal communication). This sequence identity clearly shows that the luxF gene codes for the same protein in P. phosphoreum (NCMB844). The

location of the luxF gene between luxB amd luxE and the strong homology with the luxB protein would implicate this protein in a structural role directly related to the light-emitting reaction.

ACKNOWLEDGMENTS

We would like to thank Dennis O'Kane and John Lee for communicating to us the amino terminal sequences of the non-fluorescent flavoproteins from the Photobacterium species. We very much appreciate the effort made by Josie D'Amico in typing this manuscript. This work was supported by a grant from the Medical Research Council of Canada (MT-4314).

REFERENCES

1. Engebrecht, J., Nealson, K.H., and Silverman, M. (1983) Cell 32, 773-781.
2. Hastings, J.W. (1986) in Light Emission by Plants and Bacteria (Govindjee, J., Amesz, and Fork, D.C., eds), Vol. 13, pp. 363-398.
3. Ziegler, M.M., and Baldwin, T.O. (1981) in Current Topics in Bioenergetics (Sanadi, D.R., ed.), Vol. 12, pp. 65-113.
4. Meighen, E.A. (1988) Ann. Rev. Microbiol. 42, 151-176.
5. Cohn, D.H., Mileham, A.J., Simon, M.I., Nealson, K.H., Rausch, S.K., Bonam, D., and Baldwin, T.O. (1985) J. Biol. Chem. 260, 6139-6146.
6. Johnston, T.C., Thompson, R.B., and Baldwin, T.O. (1986) J. Biol. Chem. 261, 4805-4811.
7. Foran, D.R., and Brown, W.M. (1988) Nucl. Acids Res. 16, 777.
8. Boylan, M., Graham, A.F., and Meighen, E.A. (1985) J. Bacteriol. 163, 1186-1190.
9. Engebrecht, J., and Silverman, M. (1984) Proc. Natl. Acad. Sci. USA 81, 4154-4158.
10. Miyamoto, C.M., Boylan, M., Graham, A.F., and Meighen, E.A. (1988) J. Biol. Chem. (in press).
11. Hastings, J.W., Baldwin, T.O., and Nicoli, M.Z. (1978) Methods Enzymol. 57, 135-152.
12. Miyamoto, C.M., Graham, A.F., and Meighen, E.A. (1988) Nucl. Acids Res. 16, 1551-1562.
13. Mancini, J.A., Boylan, M., Soly, R.R., Graham, A.F., and Meighen, E.A. (1988) J. Biol. Chem. (in press).
14. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning, Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. Davis, L.G., Dibner, M.D., and Battey, J.F. (1986) Methods in Molecular Biology. Elsevier Science Publishing Co., Inc., New York, N.Y.
16. Garnier, J., Osguthorpe, D.J., and Robson, B. (1978) J. Mol. Biol. 120, 97-120.
17. Visser, A.J., Vervoort, J., O'Kane, D.J., Lee, J., and Carreira, L.A. (1983) Eur. J. Biochem. 131, 639-645.
18. O'Kane, D.J., Vervoort, J., Muller, F., and Lee, J. (1987) in Flavins and Flavoproteins (Edmondson, D.E. and McCormick, D.B., eds), pp. 641-645.