Vol. 163, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1989 Pages 93-101

THE NUCLEOTIDE SEQUENCE OF THE LUXE GENE OF VIBRIO HARVEYI AND A COMPARISON OF THE AMINO ACID SEQUENCES OF THE ACYL-PROTEIN SYNTHETASES FROM V. HARVEYI AND V. FISCHERI

Timothy C. Johnston*, Kathleen S. Hruska, and Laura Faughn Adams

The Department of Biological Sciences, Murray State University
Murray, KY 42071

Received July 15, 1989

Summary: The luxE gene of bioluminescent bacteria encodes the acyl-protein synthetase component of the fatty acid reductase complex. The complex is responsible for converting tetradecanoic acid to the aldehyde which serves as a substrate in the luciferase-catalyzed reaction. The nucleotide sequence of the luxE gene of Vibrio harveyi was determined and the amino acid sequence of the acyl-protein synthetase deduced. The protein consists of 378 amino acid residues and has a molecular weight of 42,965 daltons. Alignment of the V. harveyi enzyme with the V. fischeri acyl-protein synthetase showed 62% identity.

Bioluminescence in bacteria results from the activities of four proteins. The heterodimeric enzyme luciferase catalyzes the oxidation of $FMNH_2$ and a long-chain aldehyde (tetradecanal) to produce FMN, tetradecanoic acid, H_2O, and light. The fatty acid reductase complex provides the tetradecanal for the luciferase-catalyzed, light-emitting reaction. The three components of the complex are acyl-transferase, which frees tetradecanoic acid from the fatty acid biosynthetic pathway (1); acyl-protein synthetase (2), which activates tetradecanoic acid; and fatty acid reductase (3), which reduces the activated acid to the aldehyde substrate.

The reductase, transferase, α and β subunits of luciferase, and the synthetase are encoded by genes designated luxC, luxD, luxA, luxB, and luxE, respectively and are transcribed from an operon in that order (4,5). The luxI gene precedes luxC in the

*To whom correspondence should be addressed.

Abbreviations: kbp, kilobase pair; rfDNA, replicative form DNA; SDS, sodium dodecyl sulfate.

0006-291X/89 $1.50

operon and encodes an enzyme that synthesizes the so-called autoinducer. The autoinducer diffuses through the cell membrane and only accumulates inside cells when the population becomes dense. The autoinducer is bound by an activator protein which then turns on the transcription of the operon. The activator is the product of the luxR gene, which lies upstream of the lux operon and is transcribed in the opposite direction (6, 7, 8, 9).

The sequence of the entire operon of Vibrio fischeri has been reported (10, 6, 7) and the sequences of luxC, luxD, luxA, and luxB from Vibrio harveyi have also been reported (11, 12, 8, 5). We have determined the nucleotide sequence of the luxE gene of V. harveyi and deduced the amino acid sequence of the acyl-protein synthetase. We compare these sequences with those from V. fischeri.

MATERIALS AND METHODS

Bacterial and Phage Strains. Escherichia coli strains TB1 (12) and JM103Y (13) were used in this study. TB1 (ara, Δ(lac proA,B), rpsL, lacZΔM15, hsdR) was used for the maintenance and

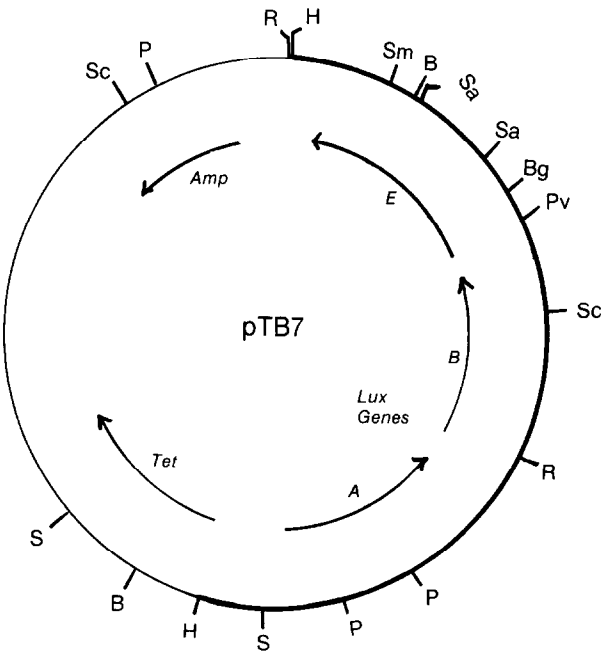

FIG. 1. Plasmid pTB7. The 4kbp V. harveyi insert is indicated by the thick line. The positions and direction of transcription of the luxA, luxB, and luxE genes are indicated. Restriction endonuclease sites are abbreviated BamHI (B), BglII (Bg), EcoRI (E), HindIII (H), PstII (P), PvuII (Pv), SalI (S), SacI (Sa), and ScaI (Sc).

purification of plasmid DNA, and JM103Y (Δ(lac proA,B) thi, strA, supE, endA, sbcB15, hsdR4, F'/traD36, proA,B, $lacI^q$, lacZΔM15) was used for the isolation of phage M13mp18 and mp19 (14), which contained inserts, and for the production of single-stranded DNA for sequencing.

Plasmids. DNA fragments for sequencing were subcloned from pTB7 (15) (Fig. 1). The plasmid is a derivative of pBR322 and was donated by Dr. Tom Baldwin. Plasmid purification was performed as described by Maniatis, et al.(16).

Nucleotide Sequencing. Restriction enzyme fragments from pTB7 were inserted into either M13mp18 or mp19 and subjected to the Sanger dideoxynucleotide chain termination protocol (17) as modified by Tabor and Richardson (18). Sequencing reagents were purchased from United States Biochemical Corp. in the form of the Sequenase Kit. Analyses of the nucleotide sequences of the luxE genes from V. harveyi and V. fischeri and of the synthetases they encode were performed using Microgenie software purchased from Beckman (19).

Materials. Chemicals were purchased from Fisher Scientific, Sigma, BioRad, and Research Organics. [α-^{35}S]dATP was purchased from Amersham Corp.. M13mp18 and M13mp19 rfDNA and restriction endonucleases were purchased from New England Biolabs, Bethesda Research Laboratories, and Boehringer Mannheim Biochemicals.

RESULTS AND DISCUSSION

Nucleotide sequence of the luxE gene and the amino acid sequence of acyl-protein synthetase of V. harveyi. The nucleotide sequence of the luxE gene and the amino acid sequence of the acyl-protein synthetase derived from it are shown in Fig. 2. Previously, the amino terminus was reported to begin 57 nucleotides 5' to the position indicated as position 1 in Fig. 2 (12). We determined that the start codon of the luxE gene is at the more 3' site, since the more 5' ATG codon has no discernable ribosome-binding site, the molecular weight of the encoded protein agrees with data from SDS polyacrylamide gels (4, 5), and the nucleotide sequence of the same region in the V. fischeri lux operon does not have an ATG at the 5' position (9). The encoded acyl-protein synthetase is comprised of 378 amino acids and has a calculated molecular weight of 42,965 daltons.

Table 1 shows the strategy used to sequence this region of the lux operon. Of the new information, 87% was sequenced on both strands. The other 13% was sequenced repeatedly on one strand and was unambiguous (data not shown). Also, in regions where only one strand was sequenced, the homology with the V. fischeri luxE gene is 66% while the homology of the entire gene with V. fischeri luxE is 62%.

luxB -80 -70 -60 -50 -40 -30 -20 -10 -1
TCGTAACGTTTAACTGATGCTGAAGGGGCAGCGATGCCCCTTATATCACCATTCTTTTCGCCGATAGAGCTAACTAATAGAGGCATTTAT
SerEnd
luxE 10 20 30 40 50 **ClaI** 60 70 80 90
ATGGACGTACTTTCAGCGGTTAAGCAGGAAAACATCGCAGCGAGCACAGAA**ATCGAT**GACTTGATTTTCATGGGAACTCCTCAGCAATGG
MetAspValLeuSerAlaValLysGlnGluAsnIleAlaAlaSerThrGluIleAspAspLeuIlePheMetGlyThrProGlnGlnTrp

100 110 120 130 140 150 160 170 180
TCATTGCAGGAACAAAAACAGCTGACATCTCGCCTTGTTAAAGGGGCATATCAATACCATTACCACAATAATGATGATTATCGTCAGTTC
SerLeuGlnGluGlnLysGlnLeuThrSerArgLeuValLysGlyAlaTyrGlnTyrHisTyrHisAsnAsnAspAspTyrArgGlnPhe

190 200 210 **BglII** 220 230 240 250 260 270
TGCGAGAGGCTGGGAGTCGGAGAGGTGGTGGA**AGATCT**CAACGATATCCCCGTTTTCCCTACTTCTATTTTTAAGTTGAAGACCCTATTA
CysGluArgLeuGlyValGlyGluValValGluAspLeuAsnAspIleProValPheProThrSerIlePheLysLeuLysThrLeuLeu

280 290 **TaqI** 300 310 320 330 340 350 360
ACACTTGACGATGACGAGG**TCGA**GAATCGCTTTACTAGCAGTGGCACTAGTGGCATCAAAAGTATTGTTGCACGAGATAGACTCAGTATT
ThrLeuAspAspAspGluValGluAsnArgPheThrSerSerGlyThrSerGlyIleLysSerIleValAlaArgAspArgLeuSerIle

370 380 390 400 410 420 430 440 450
GAGCGACTTCTTGGCTCAGTAAATTTCGGTATGAATTACGTTGGTGATTGGTTTGACCATCAGATGGAGTTGGTGAACTTAGGCCCAGAT
GluArgLeuLeuGlySerValAsnPheGlyMetAsnTyrValGlyAspTrpPheAspHisGlnMetGluLeuValAsnLeuGlyProAsp

460 470 480 490 **TaqI** 500 **SacI** 510 520 530 540
CGCTTTAATGCCAACAATATTTGGTTCAAGTACGTCATGAGCTTAG**TCGAGCTC**CTTTATCCGACCGCATTTACTGTCACTGAGGATGAG
ArgPheAsnAlaAsnAsnIleTrpPheLysTyrValMetSerLeuValGluLeuLeuTyrProThrAlaPheThrValThrGluAspGlu

550 560 570 580 590 600 610 620 630
ATCGACTTTGAGGCGACGCTAGCTAATATGAATCGTATTAAGCAGTCTGGTAAAACCATTTGTCTTATCGGCCCTCCTTATTTTATCTAT
IleAspPheGluAlaThrLeuAlaAsnMetAsnArgIleLysGlnSerGlyLysThrIleCysLeuIleGlyProProTyrPheIleTyr

640 650 660 670 680 690 700 710 720
CTACTGTGCTGTTTCATGCGCGAGCAAGGTCAAACTTTCAATGGTGGTCGCGATCTTTACATCATCACTGGCGGCGGCTGGAAAAAACAT
LeuLeuCysCysPheMetArgGluGlnGlyGlnThrPheAsnGlyGlyArgAspLeuTyrIleIleThrGlyGlyGlyTrpLysLysHis

730 740 750 760 770 780 790 800 810
CAGGATCAATCGCTCGATAGAGACGAGTTCAACCAACTTTTGTGTGAGACTTTTACCTTAGAAAGCCCAGAGCAGATTCGAGACACATTT
GlnAspGlnSerLeuAspArgAspGluPheAsnGlnLeuLeuCysGluThrPheThrLeuGluSerProGluGlnIleArgAspThrPhe

820 830 840 850 860 870 880 **SacI** 890 **BamHI** 900
AATCAAGAAGAACTGAACACCTGCTTTTTTGAGGATACAGAACACAAAAACGGTGTACCGCCCTGGGTCTTTGCAA**GAGCTCTGGATCC**T
AsnGlnGluGluLeuAsnThrCysPhePheGluAspThrGluHisLysAsnGlyValProProTrpValPheAlaArgAlaLeuAspPro

910 920 930 940 950 960 970 980 990
AAAACATTAAAGCCGCTTCCGCATGGTCAGTCAGGACTGATGAGCTATATGGATGCCTCGGCGGTCAGCTATCCATGCTTTCTAGTGACG
LysThrLeuLysProLeuProHisGlyGlnSerGlyLeuMetSerTyrMetAspAlaSerAlaValSerTyrProCysPheLeuValThr

1000 1010 1020 1030 **SmaI** 1040 1050 1060 1070 1080
GATGATATAGGCATCGTGCGAGAAGAAGAAGGCGATCG**CCCGGG**AACCACCGTTGAGATCGTTAGAAGAGTGAAGACGCGGGGTATGAAG
AspAspIleGlyIleValArgGluGluGluGlyAspArgProGlyThrThrValGluIleValArgArgValLysThrArgGlyMetLys

1090 1100 1110 1120 1130 1140 1150 1160 1170
GGGTGCGCTCTCAGTATGTCTCAAGCGTTTACAGCTAAGAATGACGGAGGTAACTGAGATGTTATGTTCGATAGAAAAAATTGAGCCGTT
GlyCysAlaLeuSerMetSerGlnAlaPheThrAlaLysAsnAspGlyGlyAsnEnd MetLeuCysSerIleGluLysIleGluProLe

1180 1190 1200 1210 1220
AACTAGCTTCATATTCCGAGTGTTGCTCAAGCCGGATCAGCCTTTTGAATTTAGGG
uThrSerPheIlePheArgValLeuLeuLysProAspGlnProPheGluPheArg

FIG. 2. Nucleotide sequence of the luxE gene and the amino acid sequence of acyl-protein synthetase. The 3′ end of luxB, luxE, the 5′end of a potential open reading frame and the intergenic regions of the lux operon of V. harveyi are indicated by the translation. Restriction endonuclease recognition sequences used to sequence this DNA are indicated by bold face type.

Amino acid alignment of the acyl-protein synthetases from V. harveyi and V. fischeri. The amino acid sequences of the acyl-protein synthetases from V. harveyi and V. fischeri were aligned by similarity using the Microgenie software (Fig. 3). There are 378 residues in each protein and by inserting three gaps into the carboxy-terminal portion of each protein, there

TABLE 1
Sequencing strategy for the *luxE* gene of *Vibrio harveyi*

Nucleotides	Fragment Sequenced	Vector
MESSAGE EQUIVALENT STRAND		
1-280	Previously Reported[a]	
213-556	BglII to BamHI	M13mp19[b]
499-817	SacI to SacI	M13mp18[b]
892-1227	BamHI to HindIII	M13mp19
MESSAGE COMPLEMENT STRAND		
1226-920	HindIII to BamHI	M13mp18
1034-886	SmaI to SacI	M13mp18[b]
892-540	SacI to SacI	M13mp18[b]
500-290	TaqI to TaqI	M13mp18
280-1	Previously Reported[a]	

[a]Reference 12.
[b]The SacI fragment was inserted in M13mp18 in both orientations.

are 237, (or 63%) identical residues. These data are consistent with the alignment of the α subunits of luciferase at 64% identity and slightly higher than the alignment of the β subunits of luciferase at 49% identity. Since the active site of luciferase is located exclusively on the α subunit (11), the low level of homology found in the β subunits might be expected. Between residues 114 and 176 of the synthetases only 5 pairs of residues are not identical and no two pairs are consecutive. Other conserved regions of more than 15 residues are positions 192 to 213 with 5 pairs of nonidentical, nonconsecutive residues; positions 265 to 282 with only 1 nonidentical pair; positions 289 to 305 with 2 pairs of

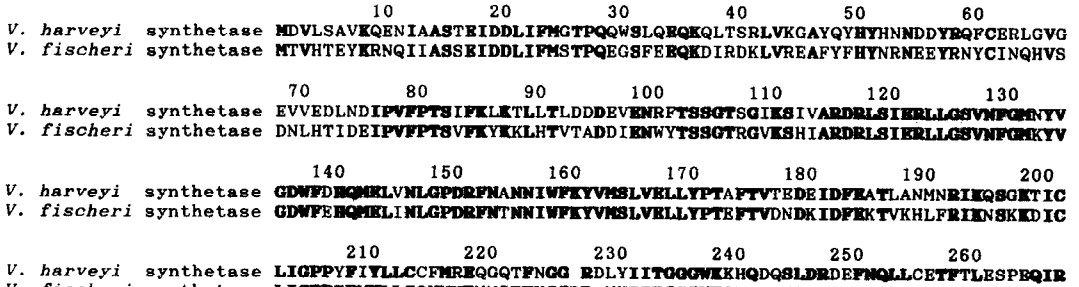

FIG. 3. Amino acid alignment of the acyl-protein synthetases of *V. harveyi* and *V. fischeri*. The alignment was performed using the Microgenie software. Identical residues are highlighted by boldface type.

nonidentical, nonconsecutive residues; and positions 324 to 338 with 3 pairs of nonidentical, nonconsecutive residues. Close inspection of Fig. 3 reveals other regions of homology that contain two consecutive, nonidentical residues. To identify the active site from these data with any certainty is impossible; however, the conserved regions mentioned above are probably very important to the function of the enzyme.

Codon usage. Johnston, et al. (12) compared the codon usage of the luxA and luxB genes of V. harveyi and found several interesting differences. The luxB gene indicates a bias for the codons UUU (for phenylalanine), GAU (for aspartic acid), and AUC (for isoleucine) in each respective codon group. The luxA gene shows no bias for codons in these groups. The luxA gene uses the lysine codon AAA twice as often as AAG and avoids using the codon AGU out of 17 serine codons. The luxB gene shows no preference in lysine codons and uses AGU more than any other serine codon. The luxE gene (Table 2) shows the same bias or lack of bias as the luxB gene in each of these cases except for the isoleucine codons. Both luxE and luxA show no bias in choosing isoleucine codons.

Codon usage in the luxE gene is different from luxA and luxB in some codon groups. The luxE gene uses the leucine codon CUU 10 out of 35 times while luxA and luxB prefer UUG. Tyrosine is encoded by UAC 7 out of 10 times in luxB, 10 out of 16 times in luxA, but only 5 times out of 12 in luxE. The luxA and luxB

TABLE 2
Codon usage in the lux genes of V. harveyi and V. fischeri

	a	b	c	d		a	b	c	d		a	b	c	d		a	b	c	d
UUU Phe	14	19	24	34	UCU Ser	4	5	12	9	UAU Tyr	7	13	9	24	UGU Cys	3	2	12	8
UUC Phe	8	4	16	4	UCC Ser	0	0	2	4	UAC Tyr	5	3	17	5	UGC Cys	5	3	2	4
UUA Leu	5	12	7	29	UCA Ser	4	4	9	10	UAA End	0	1	2	2	UGA End	1	0	0	0
UUG Leu	5	4	18	2	UCG Ser	2	0	5	4	UAG End	0	0	0	0	UGG Trp	5	5	6	7
CUU Leu	10	7	6	2	CCU Pro	5	3	6	12	CAU His	4	9	11	17	CGU Arg	2	6	8	8
CUC Leu	5	2	5	3	CCC Pro	2	2	0	2	CAC His	2	1	12	3	CGC Arg	6	2	5	3
CUA Leu	4	1	10	6	CCA Pro	3	8	11	8	CAA Gln	9	10	14	27	CGA Arg	4	1	8	7
CUG Leu	6	1	7	3	CCG Pro	5	2	3	0	CAG Gln	10	4	13	5	CGG Arg	1	2	0	1
AUU Ile	8	20	14	31	ACU Thr	9	11	12	18	AAU Asn	12	23	20	33	AGU Ser	5	7	7	9
AUC Ile	11	3	17	5	ACC Thr	7	2	8	6	AAC Asn	8	2	19	7	AGC Ser	6	3	4	3
AUA Ile	1	7	0	10	ACA Thr	7	9	7	19	AAA Lys	8	19	22	29	AGA Arg	5	8	4	3
AUG Met	11	9	18	21	ACG Thr	3	0	9	7	AAG Lys	9	6	13	11	AGG Arg	1	1	1	2
GUU Val	7	10	15	21	GCU Ala	4	7	12	9	GAU Asp	19	23	26	38	GGU Gly	10	5	22	15
GUC Val	7	5	11	4	GCC Ala	2	2	13	3	GAC Asp	9	7	20	7	GGC Gly	10	0	11	3
GUA Val	3	7	7	15	GCA Ala	5	2	15	16	GAA Glu	11	18	38	36	GGA Gly	6	8	4	11
GUG Val	6	3	14	3	GCG Ala	5	3	20	10	GAG Glu	16	8	12	13	GGG Gly	2	5	6	10

a. V. harveyi luxE.
b. V. fischeri luxE.
c. V. harveyi luxA + luxB.
d. V. fischeri luxA + luxB.

genes show a very strong bias for the glutamic acid codon GAA and the cysteine codon UGU. The luxE gene uses GAG slightly more often than GAA, 16 and 11 respectively, for encoding glutamic acid. The luxE gene uses UGC 5 times and UGU 3 times for encoding cysteine. Finally, in the glycine codon group, luxE uses the codon GGC and GGU 10 times each, while using GGA and GGG a total of 8 times. The luxB gene avoids GGC just as it does GGA and GGG, while luxA shows a slight bias for GGU over GGC.

G+C:A+T usage in V. harveyi and V. fischeri. Table 2 also contains the codon usage of the luxA, luxB, and luxE genes of V. fischeri. Overall, V. fischeri shows a stronger bias for codons that use A's or U's than V. harveyi. For example, in the valine codon group where the base in the third position is redundant, GUA and GUU are used 53 times out of the 68 valine codons. Similarly, in groups shared by two amino acids, U or A is prefered in the third position. In the AAN group, lysine is most often encoded by AAA, 48 out of 65 codons, and asparagine is usually encoded by AAU, 56 out of 65 codons. These data suggest that one of the more important selection pressures involved in codon selection in V. fischeri was a requirement for a high A+T content in DNA. Further, the data in Table 2 indicate that not only has synonymous codon choice been effected by this requirement for a high A+T content, but also amino acid choice. V. harveyi has 71 glycine residues, encoded by GGN in the proteins encoded by these three genes, while V. fischeri has only 57. V. harveyi also uses 76 alanine residues, encoded by GCN, while V. fischeri has 52. On the other hand, the V. fischeri versions of these three proteins have a total of 76 isoleucine residues (encoded by AUC, AUU, or AUA) while V. harveyi proteins have only 51. The luxA, luxB, and luxE genes in V. fischeri have an A+T content of 65%. In V. harveyi these genes have an A+T content of 55%. We suggest that the primary force involved in the divergence of these sequences, and, therefore, in the divergence of these organisms, is the requirement of V. fischeri to obtain or perhaps maintain a higher A+T content. The high A+T content may be required by V. fischeri since it is found predominantly in cold waters, whereas V. harveyi is found in warm waters. In the laboratory, V. fischeri does not grow well at temperatures higher than 30° C. V. harveyi grows very well at temperatures as high as 37° C.

Open Reading Frame. Fig. 2 shows an open reading frame 3′ to the luxE gene which extends for 29 codons. Whether or not this open reading frame is translated or even extends significantly farther is unknown. Baldwin, et al. (10) also reported an open reading frame 3′ to the luxE gene of V. fischeri that begins 4 nucleotides downstream of the termination codon. No sequence homology exists between the encoded proteins, however, and the nucleotide sequences match at only 42%. While there is enough room on the mRNA for another gene since a portion of lux mRNA terminates approximately 3,000 bases downstream of luxE (5), the encoded protein is not required for either bioluminescence or regulation of bioluminescence. Clones that terminate at a HindIII site 3′ to the luxE gene (see Fig. 1) exhibit both activities (5).

ACKNOWLEDGMENTS: This work was supported by Public Health Service grant GM-39946 from the National Institute of General Medical Sciences and by grant C-2334 from the Research Corporation. We also thank Dr. Tom Baldwin, Jerry Devine, Lorenzo Chen, Dr. Ken Nealson and Dr. Ted Meighen for thoughtful discussion and Edmund Rucker and Leanne Cochrum for technical assistance.

REFERENCES

1. Carey, L. M., Rodriguez, A., and Meighen, E. (1984) J. Biol. Chem. 259, 10216-10221.

2. Rodriguez, A., Wall, L., Riendeau, D., and Meighen, E. (1983) Biochemistry 22, 5604-5611.

3. Rodriguez, A., Riendeau, D., and Meighen, E. (1983) J. Biol. Chem. 258, 5233-5237.

4. Engebrecht, J. and Silverman, M. (1984) Proc. Natl. Acad. Sci. USA 81, 4154-4158.

5. Miyamoto, C. M., Boylan, M., Graham, A. F., and Meighen, E. A. (1988) J. Biol. Chem. 26, 1551-1562.

6. Devine, J. H., Countryman, C., and Baldwin, T. O. (1988) Biochemistry 27, 837-842.

7. Engebrecht, J. and Silverman, M. (1987) Nucleic Acids Res. 15, 10455-10467.

8. Miyamoto, C. M., Graham, A. F., and Meighen, E. A. (1988) Nucleic Acids Res. 16, 1551-1562.

9. Nealson, K. H. (1977) Arch. Microbiol. 112,73-79.

10. Baldwin, T. O., Devine, J. H., Heckel, R. C., Lin, J.-W., and Shadel, G. F. (1989) J. Bioluminescence and Chemiluminescence 4, 326-341.

11. 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.

12. Johnston, T. C., Thompson, R. B., and Baldwin, T. O. (1986) J. Biol. Chem. 261, 4805-4811.

13. Messing, J., Crea, R., and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309-321.

14. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Gene 33, 103-119.

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

16. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) Molecular Cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y. p. 89-94.

17. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

18. Tabor, S. and Richardson, C. C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767-4771.

19. Queen, C. and Korn, L. J. (1984) Nucleic Acids Res. 12, 581-599.