Volume 15 Number 24 1987 Nucleic Acids Research

Nucleotide sequence of the regulatory locus controlling expression of bacterial genes for bioluminescence

JoAnne Engebrecht and Michael Silverman

The Agouron Institute, 505 Coast Boulevard South, La Jolla, CA 92037, USA

Received June 12, 1987; Revised September 16, 1987; Accepted November 6, 1987 Accession no. Y00509

ABSTRACT

Production of light by the marine bacterium *Vibrio fischeri* and by recombinant hosts containing cloned *lux* genes is controlled by the density of the culture. Density-dependent regulation of *lux* gene expression has been shown to require a locus consisting of the *luxR* and *luxI* genes and two closely linked divergent promoters. As part of a genetic analysis to understand the regulation of bioluminescence, we have sequenced the region of DNA containing this control circuit. Open reading frames corresponding to *luxR* and *luxI* were identified; transcription start sites were defined by S1 nuclease mapping and sequences resembling promoter elements were located.

INTRODUCTION

Luminescent bacteria are common in the marine environment where they exist plantonically and as parasites and light organ symbionts. Light production by symbiotic bacteria living in association with higher organisms may serve to attract prey, for intraspecies communication or to escape from predators (1). Luminescence could function to provide a direct benefit to the bacteria. Light emitted by large aggregations of bacteria may attract feeders which ingest the bacteria into the nutrient-rich environment of the gut tracts. Another possibility is that the luminescence system is used as a terminal oxidase which allows the cells to disposal of excess reductant at low O_2 concentration (2). Luciferase, a mixed function oxidase consisting of α and β subunits, catalyzes the emission of light. In the generation of light, luciferase oxidizes a reduced flavin, $FMNH_2$, and a long chain fatty aldehyde producing oxidized flavin and the corresponding fatty acid (3). A fatty acid reductase unique to the bioluminescence system functions to synthesize or recycle the aldehyde substrate. Expression of cloned genes for luciferase and fatty acid reductase is sufficient for the production of light in a variety of non-luminous bacterial hosts (4,5) so functions which supply reduced flavin and precursors of the fatty aldehyde substrate are apparently not unique to the bioluminescence system.

Light production by most species of luminous bacteria is strongly influenced by the density of the cell culture. Light emission per cell can be as much as a thousand fold higher in dense cultures as compared to dilute cultures. Density-dependent regulation of luminescence has been investigated most thoroughly with the light organ symbiont *Vibrio fischeri* (6,7). This bacterium synthesizes a small extracellular signal molecule, called

autoinducer, which accumulates in the growth medium and induces expression of the luminescence phenotype. It is the concentration of autoinducer and not cell density *per se* which directly affects expression of luminescence. Autoinducer from *Vibrio fischeri* has been shown to be N-(β-ketocaproyl) homoserine lactone (8). Movement of autoinducer out of the cell and sensing autoinducer in the external environment probably does not require a specific transport system or a receptor for transmembrane communication because autoinducer rapidly diffuses across the bacterial membrane (9). Light production per cell actually increases exponentially once induction has commenced and reaches a plateau level before the culture has reached stationary phase of growth. Models to explain the control of luminescence should account for both of these properties.

All the genes (*lux*) necessary for light production in recombinant hosts have been cloned from *Vibrio fischeri* (strain MJ-1) on one contiguous 9-kilobase fragment of DNA (4). This fragment contains genes encoding regulatory functions and the luciferase and fatty acid reductase enzymes. Regulation of light production in recombinant *E. coli* containing *lux* genes mirrored that observed in *Vibrio fischeri* so the refined genetic techniques developed for *E. coli* have been used to explore the molecular bases of luminescence control. Figure 1 summarizes our previous work on the analysis of *lux* gene expression. Seven *lux* genes organized into two transcriptional units encode enzymes for bioluminescence and regulatory proteins (10,11). *LuxA* and *LuxB* encode the α and β subunits of luciferase, and *luxC*, *luxD* and *luxE* encode the components of the fatty acid reductase. These five genes belong to one transcriptional unit, operon R, which also includes the *luxI* gene. While all the genes whose function is known to be limited to luminescence (i.e. *lux*) reside on one recombinant molecule, other genes, specifically *cya* and *crp*, with global regulatory function have been found to be necessary for expression of luminescence (12).

The products of *luxI* and *luxR*, the latter gene located in divergently transcribed operon L, regulate the transcription of operon R. *LuxI* encodes a protein necessary for the synthesis of the autoinducer molecule. The *luxR* product is also necessary for transcription of operon R, and we hypothesize that this protein upon binding autoinducer acts as a positive effector of operon R transcription. So, *luxI* and *luxR* encode products which are both required for transcription of operon R which contains all the genes for luminescence enzymes. Since *luxI* is also a member of operon R, induction of transcription results in enhanced synthesis of autoinducer which then stimulates further transcription of operon R. The regulatory genes are organized into a positive feedback circuit which, when primed by a low constitutive level of transcription, leads to exponentially increasing expression of the genes encoding enzymes for bioluminescence. The two regulatory genes, *luxI* and *luxR*, which together positively regulate operon R transcription, negatively regulate expression of operon L. For example, the *luxR* product could be detected in minicell programming experiments only when *luxI* or *luxR* function was missing or defective. In contrast to operon R regulation, control of operon L expression by the *luxR* product and autoinducer appears to act post-transcriptionally (11). Furthermore, the c-AMP/CRP regulatory system probably controls bioluminescence by influencing transcription of operon L (12).

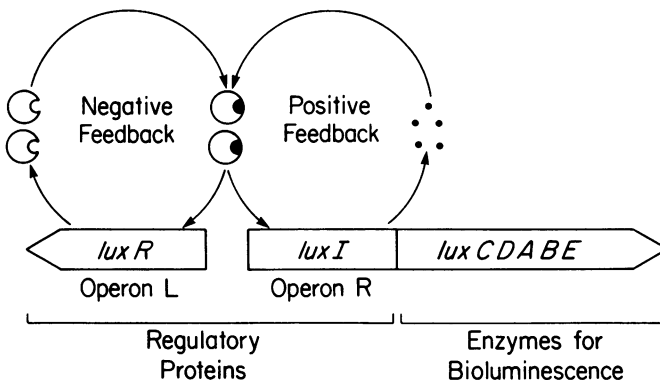

Figure 1. Model for regulation of bioluminescence. Synthesis of enzymes for the production of light is dependent on the transcription of operon R. Transcription of operon R is regulated by autoinducer, produced by an enzyme encoded by *luxI*, and a regulatory protein encoded by *luxR*. The concentration of autoinducer is dependent upon the density of the cell culture since this molecule freely diffuses through the cell membrane. A relatively small amount of autoinducer is produced by cells in dilute suspension due to weak constitutive transcription of operon R. As the cell density increases, a concentration of autoinducer is reached which is sufficient to induce transcription of operon R. However, since *luxI* is a member of operon R, synthesis of autoinducer is accelerated when operon R is induced. The genetic elements regulating operon R form a positive feedback circuit, and induction of operon R leads to an explosive increase (up to 1000-fold) in the amount of light produced by a cell. Initially, expression of operon R increases exponentially, but a compensatory circuit operates to modulate expression of operon R and subsequent light production. Modulation is achieved by limiting the expression of operon L containing *luxR* which encodes the regulatory protein necessary, with autoinducer, for transcription of operon R. This is accomplished by a negative control circuit which also requires the *luxR* and *luxI* functions, regulatory protein and autoinducer. Negative control appears to operate at the level of translation of the operon L mRNA.

Definition of the *lux* genes, operons, functions and regulatory interactions has been helpful in explaining the observed characteristics of luminescence induction. To further refine understanding of the molecular details of *lux* regulation we have sequenced the region of DNA containing the *luxR* and *luxI* genes and have identified coding sequences for these genes and sites for the promoter elements of the two *lux* operons.

METHODS

Bacterial strains and recombinant plasmids

Recombinant plasmids used for this study were derived from plasmid pJE202 which contains all the *lux* genes required for light production in a recombinant host. The details of construction of pJE202 and subclones derived from it have been described previously (4,10,11). Procedures for propagation of bacterial cultures, transformation, purification

of DNA, restriction enzyme digestion and ligation of DNA have also been reported (4,10). *E. coli* strain ED8654 (*supE supF met hsdR⁻ hsdM⁺*) or strain HB101 (hsdS recA ara proA lac galK rpsL) were used to harbor recombinant plasmids unless otherwise noted.

DNA sequencing

The use of M13 phage derivatives mp8 and mp9 was that of Messing (13). 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal) and isopropyl-beta-D-pyranoside (IPTG) were purchased from Boehringer Mannheim and Sigma, respectively. Phage DNA templates were prepared as described (14). Primers used were either the 20 bp universal primer (Bethesda Research Laboratory) or 20 bp primers designed from inital sequence data and synthesized on a 380a DNA Synthesizer (Applied Biosystems) by Linda Musik (Agouron Institute). The DNA sequencing protocol was essentially that of Biggin et al. (15), with the exception that the gel was fixed for at least 20 minutes in 10% acetic acid, 10% methanol and dried for 45 minutes at 80°C. [^{35}S]Adenosine '5-(α thio) triphosphate was obtained from New England Nuclear. Large fragment DNA polymerase I, dNTPs and ddNTPs were purchased from PL Biochemicals.

RNA purification RNA was purified from *E. coli* strain X9003 (Fsup−*lacZΔM15 thi rpsL recA*) containing plasmids pJE202(4) or pJE925(11) with a modification of the procedure of Aiba et al. (16). Strain X9003 was obtained from Dr. Cathy Lee, Harvard University, and a RecA$^-$ derivative was then constructed in our laboratory. Three hundred μl of an overnight culture of strain X9003 containing *lux* plasmid pJE202 were inoculated into flasks with 60ml or 30ml of L broth plus ampicillin (80μg/ml) and grown at 30°C until an OD_{600} of 0.35 (preinduction) or 0.70 (postinduction) was reached. Light measurements to monitor induction of *lux* genes were taken during growth of cells. Recombinant bacteria were harvested and resuspended in 3ml of a solution of 0.02M sodium acetate (pH5.5), 0.5% SDS and 1mM EDTA. Three ml of prewarmed phenol (equilibrated in 0.02M sodium acetate, pH5.5) were added, and the mixture was incubated at 60°C for 5 minutes with gentle shaking. After centrifugation the aqueous phase was re-extracted with prewarmed phenol. The RNA was then precipitated by adding 3 volumes of ethanol to the aqueous phase and chilling at -70°C for 30 minutes. The RNA precipitate was collected by centrifugation and dissolved in 500μl of 40mM tris-HCl (pH7.9), 10mM NaCl and 6mM $MgCl_2$. Ten units of of DNase (Promega Biotec) were added, and the mixture was incubated at 37°C for 15 minutes and then phenol extracted and ethanol precipitated. The RNA was resuspended in 200μl of distilled water, and the concentration of RNA was determined by measuring the optical density at 260nm. The RNA was stored in ethanol at -80°C. RNA was also isolated as described above from strain X9003 containing plasmids pJE925 or pJE980 except that 300μl of overnight cultures were inoculated into 60ml of L broth plus ampicillin (80μg/ml) and grown for 2 hours at 30°C. At this point the culture was split, and to one half 3μg of synthetic autoinducer (the gift of Dr. Anatol Eberhard) were added. The cells were then grown until the OD_{600} was equal to 0.7.

S1 nuclease protection

Plasmid pJE737 (10) containing the *lux* control region was used to isolate end-labeled restriction fragments employed to map the initiation sites of operon L and operon R tran-

scription. Probes were prepared by cleaving 10μg of pJE737 DNA with HindIII or DdeI followed by treatment with calf intestinal phosphatase to remove the terminal phosphates. After phenol extraction and ethanol precipitation the '5 ends were labeled with T4 polynucleotide kinase (Boehringer Mannheim) and 10-20μCi γ^{32}P-ATP (New England Nuclear). Unincorporated label was removed on a sephadex G-50 column according to Maniatis et al. (17), and the effluents were ethanol precipitated. The DNA was then cleaved with DdeI or HindIII, and the resulting fragments were separated on a 5% polyacrylamide gel, excised and eluted. The 335 fragment used to map the operon L transcript resulted from HindIII cleavage, end-labeling and then DdeI cleavage, and the 335bp fragment (labeled on the opposite end) used to map the operon R transcript was first generated by DdeI cleavage, then end-labeled and cut with HindIII.

The '5 end of the operon L and operon R transcripts were determined by the S1 nuclease mapping procedure of Barry et al. (18) except that the RNA mixture was hybridized for 3-5 hours at 42°C. In each reaction 100μg of RNA from the recombinant strains was hybridized to 50,000cpm of labeled fragment. S1 nuclease (Boehringer Mannheim, 200 units per reaction) digestion was allowed to proceed for 30 minutes at room temperature and was then phenol extracted and ethanol precipitated. The precipitate was washed with 70% ethanol, dissolved in 80% formamide including marker dyes, and was then electrophoresed on a denaturing 8M urea acrylamide gel. The standards for determining the size of the protected fragments were the Maxam and Gilbert G reactions (19) performed on the same fragments used in the S1 protection experiment.

RESULTS

DNA sequence of *luxR* and *luxI*

The approximate location, orientation of transcription and molecular weights of the protein products of the *lux* genes have been determined previously in our laboratory (4,10). Because we were particularly interested in the genetic control of luminescence, we chose to focus DNA sequence analysis on the region of DNA responsible for regulation of expression of *lux* genes. DNA sequence determination of the *luxR, luxI* region of DNA was performed using the dideoxy chain termination method (15,20). Specific restriction fragments obtained from a variety of recombinant plasmids were subcloned in both orientations into M13 sequencing vectors mp8 and mp9 (13). Sequencing employed the M13 universal primer for determination of the first 200-300 base pairs of sequence. Further sequencing then employed synthetic oligonucleotide primers designed from the initial sequence data. To insure that unambiguous results were obtained, both strands were sequenced. The M13 clones constructed, the strands and extents of sequence determined and the primers used for sequencing are shown in Figure 2. For reference, the sequence is numbered beginning with 1 at the HindIII site near the 3′ end of the *luxR* gene and proceeds through *luxR* and *luxI* into the *luxC* gene in operon R.

The nucleotide sequence of the region of DNA encoding *luxR*, *luxI* and part of *luxC* is shown in figure 3. Three large open reading frames were identified. The first contains a coding region, transcribed right to left, of 750 nucleotides in length which starts at an ATG at base pair 770 and ends at base pair 21. N-terminal amino acid sequence data for

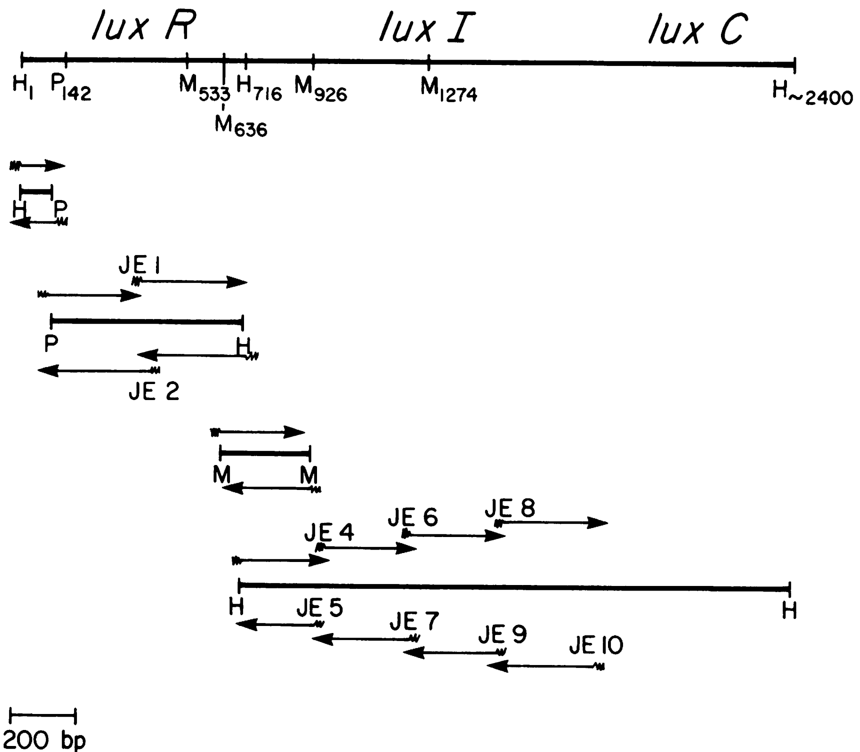

Figure 2. Sequencing strategy. The clones for sequencing were constructed in M13mp8 and M13mp9 utilizing the indicated restriction sites. H: HindIII. P: Pst I. M: Mbo I. The primers generated from initial sequence data are: JE1, ACCTGATGTTTCGCTT; JE2, TGGACTACAAAGCGAA; JE4, GGGAGGTTGGTATGACTATA; JE5, ATCATTATAGTCATACCAAC; JE6, ATTAAGTCGTTTTGCTGTAG; JE7, TTTGAGCTATTTTACCTACA; JE8, AATACTAAGTATATTATAGG; JE9, ACTCTTATATGCATATTATC; JE10, GGTTCGTCGACCGATTATCC. Sequence from termini of the cloned DNA utilized the M13 universal primer. The arrows indicate the extent of sequence generated from each primer.

the *luxR* product (personal communication, Drs. E. P. Greenberg, and H. B. Kaplan, Department of Microbiology, Cornell University) was consistent with the utilization of this initiation codon, rather than another possible site at position 756. Fifteen amino acid residues at the N-terminus have been determined, and these correspond exactly to the first fifteen amino acids predicted from the DNA sequence. Overproduction of the *luxR* gene product used for sequencing was directed by a recombinant vector in which an exogenous promoter and SD sequence (*ptac*) replaced the native control element. So, although the *luxR* product produced by this method was apparently functional, the possibility that a slightly truncated protein was used for sequencing can not be excluded. The

molecular weight calculated for the 250 residue polypeptide encoded by this DNA is 28,539. This size is in agreement with the 27,000 M_r value obtained for the *luxR* product from minicell protein programming experiments (10). The location and direction of transcription of this putative gene also corresponds to that obtained from previous genetic analysis.

The second open reading frame contains a coding region, read left to right, of 579 nucleotides which begins at ATG at position 989 and ends at position 1567. The molecular weight calculated for the 193 residue polypeptide deduced from this sequence is 21,934. An experimental value of 25,000 M_r was obtained for the product of the *luxI* gene from minicell programming experiments. The disparity in molecular weights obtained by these two methods appears to be slightly greater than would be expected to result from the inaccuracy commonly experienced in measuring protein sizes by SDS polyacrylamide electrophoresis, and this could mean that the *luxI* product migrates anomalously on SDS gels. Since the location and direction of transcription of the putative *luxI* coding sequence corresponds closely to the results of earlier genetic analysis, we are confident that the open reading frame identified in the DNA sequence represents the *luxI* gene.

The third open reading frame begins at position 1624 and continues to the right extending to the end of the sequenced region. This coding sequence has the same location and direction of transcription as the *luxC* gene, the next gene downstream from *luxI* in operon R. The three open reading frames identified in the DNA sequence, thus, correspond to the *luxR* and *luxI* genes and the 5′ end of the *luxC* gene. The amino acid sequences for the *luxR* and *luxI* products were compared to the sequences of a data library of known proteins with the FASTP program developed by Lipman and Pearson (21). The database survey did not detect any homologies with the *luxR* or *luxI* product which we judged to be significant.

Transcription initiation sites

The 5′ ends of the transcripts from operon L and operon R were mapped by the S1 nuclease protection procedure (18). A 335 base pair HindIII - DdeI fragment spanning the promoter regions of both operons was used to map both transcripts (see "methods"). RNA for the S1 protection experiment was obtained from *E. coli* containing recombinant *lux* plasmids pJE202 and pJE925. Plasmid pJE202 contains both *lux* operons in their entirety, and RNA was isolated from preinduction and postinduction cells. Plasmid pJE925 contains all of operon L (including *luxR*) and the divergent promoters of both operons, but operon R is truncated and only part of the *luxI* gene is present. Thus, transcription from this latter plasmid resembles preinduction expression, but mRNA characteristic of the postinduction pattern of expression was obtained by growing pJE925-containing *E. coli* in the presence of exogenous autoinducer.

An autoradiogram of DNA fragments protected from S1 nuclease by hybridization to RNA is shown in figure 4. Chemical cleavage ladders of the same DNA fragment used for transcript mapping are positioned next to the lanes containing the fragments protected from S1 nuclease digestion. A precise start at the operon L and operon R promoters was not observed. Several protected fragments differing by one base were detected for the P_L

```
                                                                      59
TTCGAAATGAATGCATGAAT TAA AAA TTT CAT ACC CGT TAG TTA ACG AGG ACA ATT TTA
                      N   K   F   Y   P   C   D   I   A   G   T   L   I
                                                                     113
ACG AAA TCT TTA TGA AAC CGT CGC CAA ACA ACA TAA CTC AAA GTA AAC GCG TAA
 A   K   S   I   S   Q   C   R   N   T   T   N   L   K   M   Q   A   N
                                                                     167
CCA ATT TAC CTT TCA CTG TCA TGC GAG TGA CGT CGG ATT ATA AAA ACT TTA TAG
 T   L   H   F   T   V   T   R   E   S   C   G   L   I   K   S   I   D
                                                                     221
GGT TCT CGA AAA AGG AAG CGT ACG GGT GCG ATT TGT AAG AAA AAG AGA AAA CCA
 W   S   S   K   G   E   C   A   W   A   L   C   E   K   E   R   K   T
                                                                     275
ATT TAG CAA CAA ACT AAA TAA TAA ACG ATA TAA ATA AAA AGC TAT TAA TAG TTG
 L   D   N   N   S   K   N   N   A   I   N   I   K   R   Y   N   D   V
                                                                     329
ATC TCT TCC TTG TTA ATT ACC ATA CAA GTA TGT GCG TAC ATT TTT ATT TGA TAG
 L   S   P   V   I   L   P   I   N   M   C   A   H   L   F   L   S   D
                                                                     383
ATA TAT CAA CAG AAA AAG ACT TAC ACG TTT TGA TTC GTA AGG CTT CGG TAA CAA
 I   Y   N   D   K   E   S   H   A   F   S   L   M   G   F   G   N   N
                                                                     437
TCG GCA TAC TTA TCC CTT TGA TTT GGG TCA CTA TTC TGG ACT ACA AAA GCG AAG
 A   T   H   I   P   F   S   F   G   T   I   L   G   S   T   K   A   E
                                                                     491
AAA TTA ATG TAA ACC TCT AAA AAA TAA ATG TCG TAA CAA AAG TTT ATA TAA GGT
 K   I   V   N   P   S   K   K   N   V   A   N   N   E   F   I   N   W
                                                                     545
TAA TTA ACC ACT TAC TAA CCT CAA TCT TAT TAG ATG ATA TCC TAG TAT AAA ATA
 N   I   P   S   H   N   S   N   S   Y   D   V   I   P   D   Y   K   I
                                                                     599
ATT TAA TCG CAG TAG TAT TAT AAC GGA GGT AAA AAA TCC CAT TAA TAG ATC TTA
 L   N   A   D   D   Y   Y   Q   R   W   K   K   P   Y   N   D   L   I
                                                                     653
ACT TTA TAG TCT AAA TTG GTA TCT TAC TCC TAT TTA CTA GCG CTC ACA ATT TAT
 S   I   D   S   K   V   M   S   H   P   Y   I   I   A   L   T   L   Y
                                                                     707
TAT AAG TGT TAC ATG GTA AAA TCA GTA TAG TCT ATT CGT AAC TAA TTA TAG TAT
 Y   E   C   H   V   M   K   T   M   D   S   L   C   Q   N   I   D   Y
                                                                     761
ACG AGA TGT TCG AAA TTA AAA TAA TTA ATA AGA CAT ACA CAG CAG CCG TAA ATA
 A   R   C   A   K   I   K   N   I   I   R   Y   T   D   D   A   N   I
                                                                     828
            ACCCATCTCTTTATCCTTACCTATTGTTTGTCGCAAGTTTTGCGTGTTATATATCATT
CAA AAA GTA TGGGTAGAGAAATAGGAATGGATAACAAACAGCGTTCAAAACGCACAATATATAGTAA
 N   K   M      SD                                                -10
                                                                     900
AAAACGGTAATGGATTGACATTTGATTCTAATAAATTGGATTTTTGTCACACTATTGTATCGCTGGGAATAC
TTTTGCCATTACCTAACTGTAAACTAAGATTATTTAACCTAAAAACAGTGTGATAACATAGCGACCCTTATG
               -35                         CAP
                                                                     972
AATTACTTAACATAAGCACCTGTAGGATCGTACAGGTTTAGCGAAGAAAATGGTTTGTTATAGTCGAATAAA
TTAATGAATTGTATTCGTGGACATCCTAGCATGTCCAAATCGCTTCTTTTACCAAACAATATCAGCTTATTT
                                                             -10
                                                                    1030
CGCAAGGGAGGTTGGT ATG ACT ATA ATG ATA AAA AAA TCG GAT TTT TTG GCA ATT CCA
GCGTTCCCTCCAACCA  M   T   I   M   I   K   K   S   D   F   L   A   I   P
    SD

                                                                    1084
TCG GAG GAG TAT AAA GGT ATT CTA AGT CTT CGT TAT CAA GTG TTT AAG CAA AGA
 S   E   E   Y   K   G   I   L   S   L   R   Y   Q   V   F   K   Q   R
                                                                    1138
CTT GAG TGG GAC TTA GTT GTA GAA AAT AAC CTT GAA TCA GAT GAG TAT GAT AAC
 L   E   W   D   L   V   V   E   N   N   L   E   S   D   E   Y   D   N
                                                                    1192
TCA AAT GCA GAA TAT ATT TAT GCT TGT GAT GAT ACT GAA AAT GTA AGT GGA TGC
 S   N   A   E   Y   I   Y   A   C   D   D   T   E   N   V   S   G   C
                                                                    1246
TGG CGT TTA TQA CCT ACA ACA GGT GAT TAT ATG CTG AAA AGT GTT TTT CCT GAA
 W   R   L   L   P   T   T   G   D   Y   M   L   K   S   V   F   P   E
                                                                    1300
TTG CTT GGT CAA CAG AGT GCT CCC AAA GAT CCT AAT ATA GTC GAA TTA AGT CGT
 L   L   G   Q   Q   S   A   P   K   D   P   N   I   V   E   L   S   R
                                                                    1354
TTT GCT GTA GGT AAA AAT AGC TCA AAG ATA AAT AAC TCT GCT AGT GAA ATT ACA
 F   A   V   G   K   N   S   S   K   I   N   N   S   A   S   E   I   T
                                                                    1408
ATG AAA CTA TTT GAA GCT ATA TAT AAA CAC GCT GTT AGT CAA GGT ATT ACA GAA
 M   K   L   F   E   A   I   Y   K   H   A   V   S   Q   G   I   T   E
                                                                    1462
TAT GTA ACA GTA ACA TCA ACA GCA ATA GAG CGA TTT TTA AAG CGT ATT AAA GTT
 Y   V   T   V   T   S   T   A   I   E   R   F   L   K   R   I   K   V
                                                                    1516
CCT TGT CAT CGT ATT GGA GAC AAA GAA ATT CAT GTA TTA GGT GAT ACT AAA TCG
 P   C   H   R   I   G   D   K   E   I   H   V   L   G   D   T   K   S
                                                                    1571
GTT GTA TTG TCT ATG CCT ATT AAT GAA CAG TTT AAA AAA GCA GTC TTA AAT TAAT
 V   V   L   S   M   P   I   N   E   Q   F   K   K   A   V   L   N  ATTA
                                                                    1638
GTTGTTAAATCATTAATTTATTTTAAATACTAAGTATATTATAGGGAAAATA ATG AAT AAA TGT ATT
CAACAATTTAGTAATTAAATAAAATTTATGATTCATATAATATCCCTTTTAT  M   N   K   C   I
                                                   SD               1692
CCA ATG ATA ATT AAT GGA ATG ATT CAA GAT TTT GAT AAT TAT GCA TAT AAA GAA
 P   M   I   I   N   G   M   I   Q   D   F   D   N   Y   A   Y   K   E
                                                                    1746
GTT AAA CTA AAT AAT GAT AAT AGA GTA AAA TTA TCT GTC ATT ACT GAA AGT TCA
 V   K   L   N   N   D   N   R   V   K   L   S   V   I   T   E   S   S
                                                                    1800
GTT TCA AAA ACA TTA AAT ATC AAA GAT AGA ATT AAT CTA AAT TTA AAT CAG ATT
 V   S   K   T   L   N   I   K   D   R   I   N   L   N   L   N   Q   I
                                                                    1854
GTG AAT TTT TTA TAT ACC GTT GGT CAA CGA TGG AAA AGT GAA GAA TAT AAT CGG
 V   N   F   L   Y   T   V   G   Q   R   W   K   S   E   E   Y   N   R
                            1866
CGA CGA ACC TAT
 R   R   T   Y
```

and P_R transcripts. The appearance of such multiple starts has been observed by other investigators studying different transcripts, and this effect is thought to result from S1 nuclease shortening of the DNA probe from the initiating base (22). Therefore, we have used the size of the largest of the protected fragments to compute the location of the site of transcription initiation.

From these results the transcription start site was calculated to be at base pair 813 for operon L and base pair 968 for operon R (see figure 3). The location of sequences resembling elements of promoters relative to these initiation sites is discussed later. Protection of the labeled transcript probe from S1 nuclease depended upon the source of the RNA, and the amounts of protected fragments gave a relative measure of the abundance of transcription initiation products from preinduction and postinduction recombinant *E. coli*. It is apparent (figure 4, lanes B, C, D, E) that transcripts initiated from P_R are abundant in postinduction cells but not in preinduction cells. Transcription from P_L (figure 4, lanes F, G, H, I) was generally lower than from P_R (postinduction) and was not influenced markedly by the induction state of the culture used to isolate the RNA. Thus, the presence of autoinducer strongly affected the initiation of transcription of operon R but had little effect on initiation of transcription of operon L.

Promoter sequences

Sequences of bases within the *lux* control region resembled consensus sequences inferred from examination of numerous *E. coli* promoters. Regions of homology between *E. coli* promoters include the -10 sequence (Pribnow box) represented by the consensus sequence **TA**TAA**T**G and the -35 sequence (polymerase recognition site) represented by the sequence **TTG**ACA (23). The bases in bold type are those which are most strongly conserved. The interval between the -10 sequence and the transcriptional start site (usually A or G) is generally 5 to 8 base pairs, and the interval between the -35 sequence and the -10 sequence is generally 15 to 18 base pairs.

The region of DNA '5 to the transcription start site of operon L contains sequences similar to the -10 and -35 *E. coli* promoter consensus sequences. The sequence TGATATA is found six base pairs from the site of transcription initiation, and the sequence ATGTCA is located 17 base pairs farther upstream (see figure 3). The composition, orientation and spacing of the *lux* DNA sequences supports the idea that they are elements of the P_L promoter. The sequence TGTTATA is similar to the -10 consensus sequence and is located 5 base pairs from the site of initiation of transcription of

Figure 3. The complete nucleotide sequence of the *lux* control region. The deduced protein sequences for *luxR, luxI,* and part of *luxC* are shown below the DNA sequence. Amino acids are designated with one-letter symbols as follows: A, alanine; V, valine; L, leucine; I, isoleucine; P, proline; F, phenylalanine; W, tryptophan; M, methionine; G, glycine; S, serine; T, threonine; C, cysteine; Y, tyrosine; N, asparagine; Q, glutamine; D, aspartic acid; E, glutamic acid; K, lysine; R, arginine; H, histidine. Sequences bracketed by lines represent possible ribosome binding sites and elements of promoters. Arrows mark the bases at which transcription of operon L and operon R is initiated.

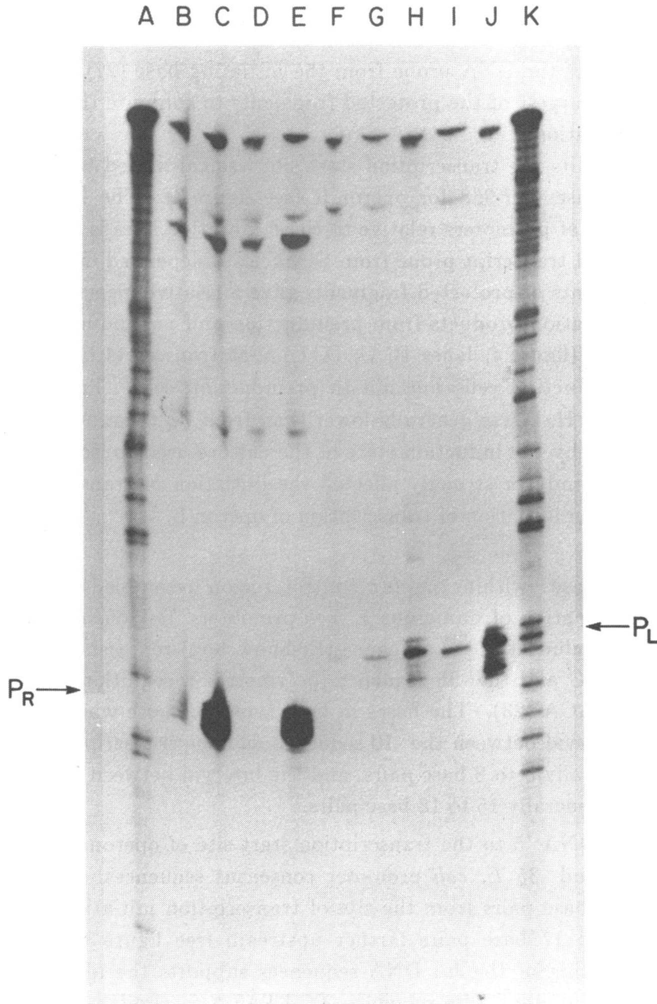

Figure 4. S1 nuclease mapping of transcripts from operon L and operon R. End-labeled fragments protected by RNA isolated from recombinant *E. coli* are marked with arrows, P_R for the fragments protected by operon R mRNA and P_L for the fragments protected by operon L mRNA. Lanes A and K contain the G reaction chemical cleavage ladders of the end-labeled probes used for operon R and operon L transcript mapping respectively. Lanes B, C, D and E contain the probe for operon R transcript mapping, and lanes F, G, H, I and J contain the probe for operon L transcript mapping. RNA for S1 protection was isolated from *E. coli* containing *lux* plasmids: lanes B and F, preinduction cells with pJE202; lanes C and G, postinduction cells with pJE202; lanes D and H, cells with pJE925 (no autoinducer); lanes E and I, cells with pJE925 (exogenous autoinducer added); and lane J cells with plasmid pJE980, a construction which overproduces *luxR* gene product.

operon R. However, no sequence homologous to a -35 sequence was identified in the region of the P_R promoter.

Expression of luminescence requires cAMP and the cAMP receptor protein (12). A consensus DNA sequence for the site of interaction with the cAMP-CRP complex has been inferred from comparisons of various promoter regions, location of mutations affecting transcription, footprinting experiments and modeling studies (24). This prototype sequence is $TGT^G/_CACA$, and an identical sequence is centered at position 876 in the *lux* sequence, 59 base pairs '5 to the site of transcription initiation of operon L. (see figure 3). Examination of the DNA sequence '5 to the translation initiation codon of the three open reading frames revealed sites with homology to the ribosome binding site of *E. coli* (Shine and Dalgarno sequence). The consensus sequence for this binding site, AAGGAGGT, usually occurs 5 to 9 base pairs from the initiation codon (25). The sequence preceding the *luxI* initiation codon has very strong homology to the *E. coli* consensus sequence (7 of 8 base pairs), and the sequence preceding the *luxR* open reading frames also shows strong homology (6 of 8 base pairs). However, the sequence preceding *luxC* shows poor homology to the SD consensus sequence (3 of 8 base pairs). The presence of stem and loop structures characteristic of terminators of transcription could not be evaluated because sequencing was not performed past the HindIII restriction site close to the '3 end of the *luxR* coding region.

DISCUSSION

The goal of our research is to understand the molecular interactions which control the expression of *lux* genes from *Vibrio fischeri*. Regulation of luminescence appears to be unusual because expression of *lux* genes is controlled by a mechanism which senses cell density so we expect novel features of gene regulation will be revealed in the course of this study. Earlier work utilizing recombinant DNA techniques and classical genetic methods resulted in the identification of *lux* genes, operons, gene products and gene functions. This information was integrated into a model for the regulation of *lux* expression (figure 1). The DNA sequence of the *lux* control region and the sites of transcription initiation reported here can be readily aligned with the *lux* regulatory genes and transcriptional units defined previously. The three open reading frames found in the DNA sequence correspond to *luxR*, *luxI* and the '5 end of *luxC*. The sites of initiation of transcription map to the region between *lux* operon L and operon R. The abundance of operon L and operon R transcripts in preinduction and postinduction cells was evident from the S1 nuclease protection experiment. Initiation of transcription of operon R increased dramatically in the postinduction state, but initiation of transcription of operon L was influenced little by the state of induction. These results are in agreement with those obtained by using transcriptional fusions formed *in vivo* with transposon mini-Mu*lac* (11) and with Northern analysis (J. Engebrecht, Ph.D. thesis, Scripps Institution of Oceanography, 1986) which indicated that autoinducer stimulated the transcription of operon R but had little effect on the transcription of operon L.

The putative promoter region preceding the initiation site of operon L transcription contains sequences with homology to and with spacing characteristic of the -10 and -35

consensus sequences of *E. coli* promoters. A sequence resembling a -10 sequence was found in the operon R promoter region, but no sequence homologous to a -35 polymerase recognition sequence was apparent. However, this is a common observation with some positively regulated promoters, and it is thought that binding of a positive regulator to this region of a promoter compensates for the absence of a -35 sequence by interacting with the polymerase directly or by altering the conformation of the DNA in such a way that polymerase binding is enhanced. A site which closely matches the consensus sequence of a cAMP-CRP recognition site was found '5 to the operon L initiation site. Studies by Dunlap and Greenberg (12) and by our laboratory (unpublished results) indicated that cAMP-CRP was required for full expression of operon L. In addition, the former group reported that cAMP-CRP had a small negative influence on transcription of operon R (in the absence of *luxR* function). This raises the possibility that cAMP-CRP can affect the expression of the promoters of both operons. The cAMP-CRP binding site is situated much closer to the operon L transcription initiation site, but it is conceivable that binding to this site could regulate the function of both promoters.

The studies reported here have set the stage for further studies on *lux* regulation. Additional refinements in our understanding will require a characterization of the interaction of regulatory proteins such as the *luxR* product and CRP with specific sites in the control region. We have already used gel shift assays to demonstrate that the *luxR* product and CRP can bind to the *lux* regulatory region (unpublished). DNA footprinting experiments should identify the particular contacts necessary for promoter activation. Because cAMP, CRP, the *luxR* product, autoinducer, polymerase and two linked promoters interact to regulate luminescence, complexity is expected. Post-transcriptional regulation of *luxR* expression has also been implicated, and possibilities such as mRNA processing must also be considered.

ACKNOWLEDGEMENT

This research was supported by grant DMB-8507874 from the National Science Foundation. We thank Dr. Shella Fuhrman for advice on S1 nuclease protection experiments and Dr. Jim Hoch for assistance in searching the protein sequence data base for sequence homologies. Dr. Anatol Eberhard generously provided synthetic autoinducer, and Drs. E. Peter Greenberg and Heidi B. Kaplan kindly shared their amino acid sequence data prior to publication. Mrs. Leticia Wolpert ably assisted in the preparation of the manuscript.

REFERENCES

1. Nealson, K. H. and Hastings, J. W. (1979) Microbiol. Rev. **43**, 496-518.
2. Makemson, J. C. and Hastings, J. W. (1982) Curr. Microbiol. **7**, 181-186.
3. Ziegler, M. M. and Baldwin, T. O. (1981) Curr. Top. Bioenerg. **12**, 65-113.
4. Engebrecht, J., Nealson, K. and Silverman, M. (1983) Cell **32**, 773-781.
5. Shaw, J. J. and Kado, C. I. (1986) Biotech. **4**, 560-564.
6. Kempner, E. S. and Hanson, F. E. (1968) J. Bacteriol. **95**:975-979.
7. Nealson, K. H. (1977) Arch. Microbiol. **112**, 73-79.

8. Eberhard, A., Burlingame, A. L., Eberhard, C., Kenyon, G. L., Nealson, K. H. and Oppenheimer, N. J. (1981) Biochemistry **20**, 2444-2449.
9. Kaplan, H. B. and Greenberg, E. P. (1985) J. Bacteriol. **163**, 1210-1214.
10. Engebrecht, J. and Silverman, M. (1984) Proc. Natl. Acad. Sci. USA **81**, 4154-4158.
11. Engebrecht, J. and Silverman, M. (1986) *In* Setlow, J. K. and Hollander, A. (eds), Genetic Engineering, Plenum Publishing Corp., New York, Vol. 8, pp. 31-44.
12. Dunlap, P. V. and Greenberg, E. P. (1985) J. Bacteriol. **164**: 45-50.
13. Messing, J. (1983) Methods Enzymol. **101**, 20-78.
14. Messing, J, Crea, R. and Seeburg, P. H. (1981) **9**, 309-321.
15. Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983) Proc. Natl. Acad. Sci. USA **80**, 3963-3965.
16. Aiba, H., Adhya, S. and De Crombrugghe, B. (1981) J. Biol. Chem. **256**, 11905-11910.
17. Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) Molecular Cloning, a Laboaratory Manual. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York. pp. 464-467.
18. Barry, G., Squires, C. and Squires, C. L. (1980) Proc. Natl. Acad. Sci. USA **77**, 3331-3335.
19. Maxam, A. M. and Gilbert, W. (1980) Methods. Enzymol. **65**, 499-560.
20. Sanger, F., Nicklen, S. and Coulsen, A. R. (1979) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
21. Lipman, D. J. and Pearson, W. R. (1985) Science **227**, 1435-1441.
22. Burton, Z. F., Gross, C. A., Watanabe, K. K. and Burgess, R. R. (1983) Cell **32**, 335-349.
23. Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. **13**, 319-353.
24. O'Neill, M. C., Amass, K. and De Crombrugghe, B. (1981) Proc. Natl. Acad. Sci. USA **78**, 2213-2217.
25. Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. USA **71**, 1342-1346.