The nucleotide sequence of the 1.85-kilobase EcoRI fragment from \textit{Vibrio harveyi} that was cloned using a mixed-sequence synthetic oligonucleotide probe (Cohn, D. H., Ogden, R. C., Abelson, J. N., Baldwin, T. O., Nealon, K. H., Simon, M. I., and Mileham, A. J. (1983) \textit{Proc. Natl. Acad. Sci. U. S. A.} 80, 120–123) has been determined. The $\alpha$ subunit-coding region (luxA) was found to begin at base number 707 and end at base number 1771. The $\alpha$ subunit has a calculated molecular weight of 40,108 and comprises a total of 355 amino acid residues. There are 34 base pairs separating the start of the $\alpha$ subunit structural gene and a 669-base open reading frame extending from the proximal EcoRI site. At the 3' end of the luxA coding region there are 26 bases between the end of the genomic gene and the start of the luxB structural gene. Approximately two-thirds of the $\alpha$ subunit was sequenced by protein chemical techniques. The amino acid sequence implied by the DNA sequence, with few exceptions, confirmed the chemically determined sequence. Regions of the $\alpha$ subunit thought to comprise the active center were found to reside in two discrete and relatively basic regions, one from around residues 100–115 and the second from around residues 280–295.

Bacterial luciferase catalyzes the light-emitting reaction in luminous bacteria. Its synthesis is regulated by a complex control mechanism that has been termed autoinduction (1, 2), and in fully induced cells luciferase comprises up to 5% of the soluble protein (3). The enzyme is a heterodimer and catalyzes the following reaction.

\[
\text{FMNH}_2 + \text{RCHO} + \text{O}_2 \rightarrow \text{FMN} + \text{RCOOH} + \text{H}_2\text{O} + \text{light}
\]

The subunits of the enzyme from \textit{Vibrio harveyi} have molecular weights determined by SDS-polyacrylamide gel electrophoresis of 42,000 and 37,000 for $\alpha$ and $\beta$, respectively (4). Mutant enzyme analyses and chemical modification studies indicate that the single active center resides primarily if not exclusively on the $\alpha$ subunit (5). The specific role of the $\beta$ subunit is unknown, but it is absolutely required for bioluminescence activity.

Recently, the luciferase genes from \textit{V. harveyi} were isolated (6–8) and shown to be closely linked on the bacterial chromosome. luxA encodes the $\alpha$ subunit and luxB encodes the $\beta$ subunit. Partial sequence information at the nucleotide (7) and amino acid (9) levels suggests that the genes arose by tandem duplication of an ancestral gene.

Partial amino acid sequence information from regions thought to be associated with the active center has been obtained during the past five years (10), but determination of the entire sequence of the subunits has been hampered by the poor solubility of the proteolytic and chemically derived fragments. In order to determine the encoded sequence of the $\alpha$ subunit and to better understand the structure and regulation of the lux region of the \textit{V. harveyi} chromosome, we have determined the nucleotide sequence of the ~1.85-kb EcoRI fragment known to contain the entire luxA gene and part of the luxB gene. We report here the complete nucleotide sequence of the luxA gene and compare it with amino acid sequences obtained from analysis of peptide fragments comprising approximately two-thirds of the $\alpha$ subunit.

**MATERIALS AND METHODS**

Subcloning and DNA Sequencing—As the restriction map of the 1.85-kb EcoRI fragment was relatively well characterized, subcloning was done by purifying fragments from the recombinant plasmid pAG101 (7) and ligating them into the appropriate vector. During the course of the sequencing, additional restriction sites were identified, and these were then used to construct additional subclones. Subclones were constructed in the bacteriophage M13 derivatives mp7 and mp8 (11). DNA sequencing was by the chain-termination method (12).

Protein Sequencing—The procedures used to grow \textit{V. harveyi}, purify luciferase, and separate the $\alpha$ and $\beta$ subunits have been

*The abbreviations used are: SDS, sodium dodecyl sulfate; kb, kilobase; PTH, phenylthiohydantoin; DNS, 5-dimethylaminonaphthalene-1-sulfonyl.
published (13-15). Starting with 600 g of frozen cell paste, 1050 mg of luciferase was purified, and from this enzyme, 438 mg of α subunit was isolated. The α subunit was judged to be greater than 95% pure based on Coomassie Blue staining of SDS-polyacrylamide gels (16).

Prior to digestion, the α subunit was alkylated by reaction with iodoacetate (17). The carboxymethylated subunit was digested with *Staphylococcus aureus* strain V8 protease (Miles) in ammonium bicarbonate, pH 7.8, 1 mM EDTA, at a substrate to enzyme ratio of 25:1 (w/w) for 16 h at room temperature. These conditions have been reported to yield cleavage at glutamyl residues only (18). Peptides were resolved by chromatography on columns of the cation exchange resins Aminex A-4 (Bio-Rad) and PA-35 (Beckman) using gradients of pyridine-acetate by methods that have been described (17, 19).

High voltage electrophoresis on paper was used extensively both to analyze fractions for purity and to purify contaminated peptides by published procedures (17, 19). Purity of peptides was also assessed on the basis of amino acid composition. Amino acid compositions of samples were determined with a Beckman model 121 amino acid analyzer following 24-h hydrolysis at 110 °C with 6 M HCl.

Edman degradation was performed manually as previously described (17, 19). Aliquots were removed after each cycle of degradation and dismantled, and DNS derivatives, released by acid hydrolysis, were identified by thin layer chromatography (20). Automated Edman degradation was performed with a Beckman Sequencer model 890C as described previously (9, 17, 19). PTH derivatives were determined by thin layer chromatography using a total of three solvent systems (9, 17, 19). The multiple determinations were considered necessary due to the lack of quantitative data from the chromatograms. Only residues identified by all three systems were considered to be accurate and are presented here.

RESULTS

DNA Sequencing

The clones used for sequencing are listed in Table I; the direction and extent of sequence derived from each clone are shown in Fig. 1. The nucleotide sequence of the 1.85-kb EcoRI fragment is shown in Fig. 2. The sequence of about 75% of the fragment was determined from both strands (see Fig. 1), and in regions where information from only one strand is presented it is from areas of the sequencing gels where the nucleotide assignment is unambiguous and/or regions for which protein sequence was available. Shown below the DNA sequence is the implied protein sequence.

Peptide Sequencing

After digestion with *S. aureus* protease, 33 α subunit peptides were isolated in sufficient yield and purity for sequence analysis. The amino acid compositions of these peptides are presented in Table II. While knowledge of the sequences of these peptides alone was not sufficient to deduce the sequence of the α subunit, the data were highly useful in checking the nucleotide sequence and ascertaining frameshift errors in reading the DNA-sequencing gels. The locations of most of the peptides presented in Table II are indicated in Fig. 2.

Determination of the Sequence of the α Subunit

The entire sequence of the α subunit was determined from the sequence of the *luxA* gene. We had also determined the sequence of numerous peptides derived from the protein, and knowledge of the sequences of those peptides was very helpful in confirming the DNA sequence. A detailed interpretation of the sequence follows.

Residues 1–26—The amino acid sequence of residues 1–26 was determined using a Beckman Sequencer, and the sequence has been published (9). Residue 13 was not determined in the earlier work, and the DNA sequence reported here allows us to assign Pro to that position (see Table I and Fig. 1, clone 3). This assignment is consistent with the sequence of the peptide SAP 2. Residue 17 was reported to be Glu (9), but the DNA sequence indicated Gln. It is likely that the DNA sequence is correct, and the error was due to the well-known deamidation of Gln. The peptide SAP 1 was placed at residues 1–4 based on sequence identity. The DNA sequence in this region (residues 707–784) was determined from both strands using clones 3, 4, and 15.

It was of interest that while the *S. aureus* V8 protease, under the conditions of the digestion, has been reported to be specific for bonds on the carboxyl side of Glu residues (18), we observed cleavages at reasonably high yield at other residues, most notably Gly. SAP 1 resulted from cleavage of a glycy1-asparagine bond, and SAP 2 resulted from cleavage of a threonyl-tyrosine bond, as well as a glutamyl-leucine bond. The cleavages observed and reported here indicate a preference, but hardly a specificity, for Glu.

Residues 27–117—The DNA sequence through this region (residues 785–1063) was determined from both strands (see Fig. 1), and the encoded amino acid sequence was confirmed by peptides SAP 3 through SAP 14. The corresponding region in the DNA sequence was residues 785–1063; the sequence was determined using clones 3, 4, 5, and 15 and by using the

---

**Table I**

Details of the cloning strategy used to determine the sequence of the 1.85-kb EcoRI fragment

| Sequence designation* | Fragment cloned | Position of first baseb | Vector |
|------------------------|-----------------|-------------------------|--------|
| 1*                     | 1.8-kb EcoRI-EcoRI | 1 mp7                  |        |
| 2*                     | 1.6-kb HindIII-EcoRI | 266 mp8               |        |
| 3*                     | 1.3-kb SalI-EcoRI | 572 mp8               |        |
| 4*                     | 0.2-kb Abl-Alul | 749 mp8               |        |
| 5*                     | 0.9-kb PstI(partial)-EcoRI | 929 mp8       |        |
| 6*                     | 0.4-kb RsuI-RsuI | 1061 mp8             |        |
| 7*                     | 0.6-kb PstI-EcoRI | 1259 mp8             |        |
| 8*                     | 0.3-kb NruI-NruI | 1341 mp8             |        |
| 9*                     | 0.2-kb NruI-EcoRI | 1632 mp8             |        |
| 10*                    | 1.5-kb EcoRI-EcoRI | 1838 mp7             |        |
| 11*                    | 0.3-kb NruI-NruI | 1632 mp8             |        |
| 12                    | 0.08-kb Sau3AI-Sau3AI | 1403 mp8       |        |
| 13                    | 0.7-kb NruI-NruI | 1341 mp8             |        |
| 14*                    | 1.8-kb EcoRI-EcoRI | 1092 mp7             |        |
| 15*                    | 0.9-kb PstI-EcoRI | 929 mp8              |        |
| 16*                    | 0.3-kb NruI-NruI | 611 mp8              |        |
| 17*                    | 0.3-kb NruI-EcoRI | 291 mp8              |        |
| 18*                    | 0.1-kb Sau3AI-EcoRI | 130 mp8          |        |

*Designations pertain to Fig. 1.

a Position in the 1.85-kb fragment of the first base of the fragment that was being sequenced.

b Sequences 1 and 10 resulted from isolation of both orientations of the 1.85-kb EcoRI fragment in mp7 (see Ref. 7).

c Sequences 8 and 11 resulted from isolation of both orientations of the 0.3-kb NruI fragment in mp8.

This sequence was determined by priming the same template DNA as was used in sequence 10 with the mixed-sequence synthetic oligonucleotide that was used to isolate the original clone (7).

---

**Fig. 1.** Sequencing strategy for the 1.85-kb EcoRI fragment. The clones were constructed using the indicated restriction sites (see also Table I). Arrows indicate the direction and extent of sequence derived from each clone.
Fig. 2. Nucleotide sequence of the 1.85-kb EcoRI fragment and the encoded amino acid sequence of the luciferase α subunit.

For the luxA and luxB structural genes (nucleotides 707-1771 and 1901-1838, respectively), known peptide sequences are indicated below the encoded sequence (see also Table II). Underlined amino acid residues indicate those that were determined by automatic Edman degradation of either the whole α or β subunit (9) or of large proteolytic fragments of the α subunit (10). Locations of peptides derived from digestion of the α subunit by trypsin (T) or the Staphylococcus aureus protease (SAP) are indicated by the labeled brackets. Regions of sequence of peptides enclosed within brackets were not determined, and alignments are by amino acid content.

Regions of sequence of peptides enclosed within brackets were not determined, and alignments are by amino acid content. Numerous other peptides were isolated and sequenced, but they were of no additional help in elucidation of the sequence and are not shown here for the sake of clarity.

synthetic oligonucleotide that was used in the original cloning (Ref. 7; see sequence 14 in Fig. 1). The position of the peptides SAP 3 through 14 was based on the DNA sequence, with the exception of SAP 10 and 12 which were placed relative to each other on the basis of the sequence of the reactive thiol-containing tryptic peptide Phe-Gly-Ile-Cys-Arg (21). The cleavage of bonds associated with the carboxyl side of glycyl residues by the SAP enzyme was quite evident here as well.
Peptides SAP 3, SAP 5, SAP 7, SAP 8, SAP 10, SAP 12, and SAP 13 all resulted from cleavage of bonds associated with the carboxyl side of glycyl residues.

**Residues 118–143**—The DNA sequence in this region (residues 1063–1139) was determined from both strands, clones 5, 6, and 13. Sequence 14 (Fig. 1) was derived by priming the template DNA from a clone that carried the entire 1.85-kb EcoRI fragment with the mixture of 8 sequences used in the original cloning (7). The protein sequence to which the synthetic probe was directed was Met-Asp-Cys-Trp-Tyr-Asp (amino acid residues 128–133 (10)). The 17-base oligonucleotide was constructed to have 2 base ambiguities at positions 6, 9, and 15, giving a total of 8 sequences. The DNA sequence demonstrated that the correct base in the ambiguous position in the Asp codon was C, while it was T for both Cys and Tyr. The amino acid sequence in this region was determined using a Beckman Sequencer and the large proteolytic fragment demonstrated that the correct base in the ambiguous position in the Asp codon was C, while it was T for both Cys and Tyr.

The DNA sequence indicated the sequence Met-Lys for these positions, consistent with the peptide predicted from the nucleotide sequence. Numerous other sequences would be determined in the following order.

**Residues 144–355**—The corresponding region from the DNA sequence (residues 1140–1771) was determined based on clones 7, 8, 9, 10, 11, 12, and 13. From about 1170 to base 1259 (the PstI site in Fig. 1), the DNA sequence was of the message-complementary strand (clone 13 in Fig. 1). From base 1196–1231, the DNA sequence was confirmed by the sequence of SAP 18. The sequence of SAP 18, determined by manual Edman degradation with the DNS-CI technique, identified Ala (rather than Pro) at position 169 and Pro (rather than Ala) at position 174. The errors were probably due to the similarity in migration of DNS-Pro and DNS-Ala on polyamide sheets (20). SAP 19 was shown to contain a Trp based on absorbance spectroscopy, but of course DNS-Trp was not detected due to the acid hydrolysis step. The Trp was tentatively placed at the amino-terminal end of the peptide since no DNS derivative was obtained from the undegraded peptide. The DNA sequence from 1250–1261 confirms the location of the sequence of SAP 19 and the location of the Trp residue. Due to the alignment of peptides SAP 18 and 19 and the unambiguous sequences read from the DNA sequencing gels, we are confident that the sequence in this region is correct in spite of the fact that the sequence was derived from only one strand.

The sequence from position 1259 through the end of the 1.85-kb EcoRI fragment was determined from both strands with the exception of a stretch of ~10 bases from 1403 to ~1415, where the sequence was exclusively from clone 8 (Fig. 1). The alignment of SAP 20, 21, and 22 (amino acid residues 201–214; nucleotide residues 1307–1348) before this region and tryptic peptide T1 (amino acid residues 238–244; nucleotide residues 1421–1438) after the region of single strand data gives us confidence that the region is correct.

The proteolytic fragment resulting from chymotryptic cleavage around residue 280 (position ~1250 in the DNA sequence) has been designated the light $\delta$ fragment (10). Isolation of this fragment by SDS-gel electrophoresis and sequence analysis on a Beckman Sequencer led to an erroneous sequence due to contamination of the sample (10). The sample actually contained a mixture of two fragments, one beginning with residue 281 and one beginning with residue 283. The two sequences would be determined in the following order.

1. Val-Leu-Lys-Gly-His-Lys-Asp-Thr-Aas-Arg-Arg-Ile-Asp-Tyr-
   Ser-Tyr-Glu-Ile-Asn-Pro-Val-
2. Lys-Gly-His-Lys-Asp-Thr-Asn-Arg-Arg-Ile-Asp-Tyr-Ser-Tyr-
   Glu-Ile-Asn-Pro-Val-Gly-Thr-
Luciferase α Subunit Sequence

Table III

Comparison of the amino acid composition determined from the sequence with the experimentally measured composition (21)

| Amino acid | Encoded composition | Measured composition |
|------------|---------------------|----------------------|
| Lysine     | 20                  | 19.2                 |
| Histidine  | 11                  | 10.6                 |
| Arginine   | 13                  | 13.4                 |
| Aspartic acid | 28                | 49.1                 |
| Asparagine | 19                  |                      |
| Threonine  | 21                  | 22.2                 |
| Serine     | 17                  | 17.3                 |
| Glutamic acid | 24              | 40.4                 |
| Glutamine  | 13                  |                      |
| Proline    | 12                  | 12.5                 |
| Glycine    | 26                  | 27.0                 |
| Alanine    | 26                  | 27.0                 |
| Cysteine   | 8                   | 8.6                  |
| Valine     | 19                  | 19.2                 |
| Methionine | 9                   | 8.6                  |
| Isoleucine | 22                  | 19.2                 |
| Leucine    | 28                  | 27.9                 |
| Tyrosine   | 16                  | 16.4                 |
| Phenylalanine | 17            | 17.3                 |
| Tryptophan | 6                   | 4.8                  |

III Val-Leu-Arg-Gly-Asp-Thr-Asn-Thr-Asn-Ile-Asp-Ile-Asp-Tyr-Glu-Tyr-Glu-Ile-Asp-Pro-Val-

Sequence I begins with residue 281, sequence II begins with residue 283, and sequence III is the reported sequence of δL (10). Re-evaluation of the data necessitated by the lack of agreement with the DNA sequence demonstrated the existence of the two sequences (I and II given above). The error was due to the use of gas chromatography (22) and thin-layer chromatography to identify the PTH derivatives and the differential stabilities of the various PTH derivatives. The sequences of SAP 25, 26, and 27, and tryptic peptide T2 (see Table II) confirm the DNA sequence and are consistent with the above hypothesis explaining the earlier error (10).

The sequence from position ~1600 in the DNA sequence to the end of the α subunit coding region was largely confirmed by the sequences of SAP peptides 28, 29, 30, 31, 32, and 33 (see Table II and Figs. 1 and 2).

Structure of the lux Region

Three parallel open reading frames are seen in the DNA sequence. The only complete reading frame is that encoding the α subunit, from nucleotides 707–1771. It encodes a protein of 355 amino acids with a calculated molecular weight of 40,108. This agrees well with the published molecular weight of 42,000 (4). In addition, the composition of the encoded protein (Table III) closely corresponds with the measured amino acid composition (21). Only 26 nucleotides separate the stop codon of the luxA gene and the beginning of the luxB coding region, suggesting that the two genes are cotranscribed. Upstream from luxA was an incomplete reading frame of 223 codons which had the same polarity as luxA and luxB. There are 34 nucleotides between the stop codon of the upstream open reading frame and the beginning of the luxA structural gene, suggesting that this gene may be transcribed with both luxA and luxB.

DISCUSSION

The 1.85-kb EcoRI fragment described in this paper was isolated from a genomic clone bank of V. harveyi DNA on the basis of hybridization with a mixed sequence synthetic oligonucleotide probe designed from α subunit amino acid sequence

(7, 10). The complete sequence of the fragment, reported here, showed that it contained the entire α subunit-coding region, a region encoding the carboxyl-terminal 223 residues of the polypeptide of unknown function and the amino-terminal 13 codons of the β subunit. The amino-terminal coding regions of the luxA and luxB genes imply amino acid sequences identical to those of the mature polypeptides (9), demonstrating that neither subunit undergoes post translational processing in the amino-terminal region.

Chemical modification and limited proteolysis studies with bacterial luciferase have shown that a highly reactive sulfhydryl group, thought to reside in or near the flavin-binding site, is located close to a region that is highly sensitive to proteases (10, 23, 24). In the complete sequence, the reactive cysteinyl residue is in position 106 (Fig. 3).

The proteolytic fragment resulting from chymotryptic cleavage around residue 280 has been designated the light δ fragment (10). We have re-evaluated the light δ protein se-
FIG. 3. The complete amino acid sequence of the α subunit of luciferase from V. harveyi and the predicted secondary structure of the subunit. Amino acid residues within predicted α-helical regions are indicated as such, residues within predicted β-sheet regions are indicated within broad arrow-like forms, and random coil regions are indicated within parallel lines. Amino acid residues found more often on the surfaces of proteins are indicated by white backgrounds and amino acid residues found more often within globular proteins are indicated by black backgrounds (38).

sequence data because of a lack of agreement with the DNA sequence; we found that the light sample contained a mixture of 2 peptides resulting from cleavage by chymotrypsin at residues 280 and 282, resulting in errors in the protein sequence determination. The differential stabilities of the PTH amino acids derived in the sequence of the mixture account for the errors. The secondary structure of the α subunit predicted by the modified method (25) of Chou and Fasman (26) is shown in Fig. 3. The prediction is that 34% of the residues should form α-helix while 12% should be in the β sheet configuration. These values are consistent with the measurements of 28% α-helix and 14% β-sheet reported by Holzman and Baldwin for the dimer (27).

Luciferase is a highly soluble protein. For a protein of ~40,000 daltons, the α subunit has a high proportion of hydrophilic residues (aspartic acid, asparagine, glutamic acid, glutamine, lysine, and arginine). Compared with two other polypeptides of similar size, carboxypeptidase A (a monomer) and horse liver alcohol dehydrogenase (one subunit of the dimer), luciferase α subunit has about 25% more external residues (33% versus 26%). The proportion of hydrophobic residues (leucine, methionine, isoleucine, valine, cysteine, phenylalanine, and tyrosine) is about the same in all three examples (~34%), but luciferase α subunit has a lower fraction (33%) of "neutral" residues (alanine, methionine, isoleucine, valine, cysteine, phenylalanine, and tyrosine) is about the same in all three examples (~34%), but luciferase α subunit has a lower fraction (33%) of "neutral" residues (al alanine, threonine, glycine, proline, serine, histidine, and tryptophan) than the other two, with about 40% spiece. The amino-terminal third is the most nonpolar region of the subunit. From residues 49–84, 71% of the residues are nonpolar. There is only a single trypsin-sensitive bond between residues 29 and 98. A second area of high nonpolarity is from residues 166–199, which contains 59% nonpolar residues. These nonpolar regions are likely to be either internal in the native structure or involved in interactions with the β subunit. Between these two nonpolar regions is a highly polar region from residues 107–153 containing 36% charged residues as well as the reactive cysteine and the first protease labile region. A second polar region, from residues 262–297, has 42% charged residues and contains the second protease labile region. It will be especially interesting to compare the amino acid sequences in these regions with the same regions of the β subunit, since the two subunits appear to be homologous (9) but have very different functions.

In the nucleotide sequence, an open reading frame is seen upstream from luxA. The reading frame has the same polarity as luxA and extends beyond the proximal EcoRI site at the end of the cloned fragment. The reading frame has 223 codons and ends 34 nucleotides upstream from the luxA initiation codon. The random chance of not finding a stop codon in a sequence of 223 codons is less than 0.002%, strongly suggesting that this region encodes a polypeptide. Previous analyses of the proteins encoded by the 5-kb BamHI fragment that encompasses the 1.85-kb fragment (6) support this view. The BamHI fragment has about 1.2 kb upstream from the proximal EcoRI site. A clone carrying the fragment synthesizes an M, 35,000 protein in addition to the 2 luciferase subunits. Deletion from the proximal BamHI site to the HindIII site 265 nucleotides downstream from the proximal EcoRI site eliminates this protein. Assuming that the open reading frame upstream from luxA encodes the carboxyl-terminal 223 amino acids of this protein, the structural gene should extend about 230 nucleotides upstream from the EcoRI site, to account for the observed molecular weight of 35,000. A protein with approximately the same mobility that is coinduced with lucif-
binding sites results in identical rates of translational initiation, resulting in synthesis of equal amounts of each subunit. The codon usage of the luxA gene of V. harveyi, presented in Table IV, shows a bias in codon selection, as does E. coli and other organisms for which genes have been sequenced (36). Parker et al. (37) have shown a relationship between codon usage and translational fidelity and suggest that codons are selected in order to reduce third position misreading. In codon groups where the third position misreading would result in amino acid substitution, G or C is preferred in the third position. luxA does not follow the rule. The codons for Asn, Glu, His, and Phe show no bias while the codons for Lys show a preference for A over G. The lack of codon bias in these groups suggests that either luxA is not representative of codon usage in V. harveyi or that the parameters dictating codon usage in V. harveyi are significantly different from those dictating codon usage in other organisms such as E. coli. Until more is known about V. harveyi tRNAs, it will be difficult to determine what pressures have been at work in codon selection.

The sequence of the luxA gene and its surrounding DNA sequences indicates that the genes of Vibrioharveyi are not radically different from those of other bacteria. Information on the fine structure of the luxA region suggests that the luciferase genes are cotranscribed and that they are in an operon with at least one additional gene. Recent work by Engbrecht and Silverman (30) indicates that the lux gene family in V. fischeri is composed of two operons comprising 7 distinct complementation groups, the two luciferase subunits, three polypeptides involved in aldehyde metabolism, and two genes involved in regulation of the expression of bioluminescence.

The amino acid sequence of the α subunit has allowed insight into the structure and function of the enzyme, supporting the biochemical data. The sequence of the luxB gene should expand our understanding further and perhaps indicate the manner in which the two subunits interact. It will also be important to define the limits of the operon and identify the functions of the additional genes in V. harveyi, for which there is such a wealth of biochemical data, as has been accomplished for V. fischeri (30).

Acknowledgments—We thank Dr. A. Boyd for assistance with DNA sequencing and secondary structure prediction. Dr. Rosemarie Swan-son offered numerous insightful comments on the hydrophobicity considerations of the amino acid sequence, and Dr. Timothy Johnston provided the essence of our interpretation of the codon usage.

REFERENCES

1. Nealson, K. H., Platt, T., and Hastings, J. W. (1970) J. Bacteriol. 104, 313-322
2. Eberhard, A., Burlingame, A. L., Eberhard, C., Kenyon, G. L., Nealson, K. H., and Oppenheimer, N. J. (1981) Biochemistry 20, 2444-2449
3. Hastings, J. W., Riley, W. H., and Massa, J. (1965) J. Biol. Chem. 240, 1473-1481
4. Gussalus-Miguele, A., Meighen, E. A., Nicoli, M. Z., Nealson, K. H., and Hastings, J. W. (1972) J. Biol. Chem. 247, 398-404
5. Ziegler, M. M., and Baldwin, T. O. (1981) Curr. Top. Bioenerg. 12, 65-115
6. Belas, R., Mileham, A., Cohn, D., Hilton, M., Simon, M., and Silverman, M. (1982) Science 218, 791-793
7. Cohn, D. H., Ogden, R. C., Abelosn, J. N., Baldwin, T. O., Nealson, K. H., Simon, M. I., and Mileham, A. J. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 120-123
8. Baldwin, T. O., Berends, T., Bunch, T. A., Holzman, T. F., Rausch, S. K., Shamansky, L., Treat, M. L., and Ziegler, M. M. (1984) Biochemistry 23, 3663-3667
9. Baldwin, T. O., Ziegler, M. M., and Powers, D. A. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 4887-4889

2 A. C. Bieger, personal communication.

Table IV
Codon usage in the luxA gene of Vibrio harveyi

| U | C | A | G |
|---|---|---|---|
| U | UUU Phe 8 | UCU Ser 7 | UAU Tyr 6 | UGU Cys 7 |
| UUC Phe 9 | UCC Ser 1 | UAC Tyr 10 | UGC Cys 1 |
| UUA Leu 3 | UCA Ser 5 | UAA End 1 | UGA End 0 |
| UUG Leu 9 | UGG Ser 1 | UAG End 0 | UGG Trp 6 |
| U | C | A | G |
| C | CUU Leu 3 | CUC Pro 4 | CAU His 5 | CGU Arg 4 |
| CUC Leu 3 | CCC Pro 0 | CAC His 6 | CGC Arg 3 |
| CUA Leu 7 | CCA Pro 5 | CAA Gln 7 | CGA Arg 5 |
| CUG Leu 3 | CCG Pro 3 | CAG Gln 6 | CGG Arg 0 |
| A | AUU Ile 12 | ACC Thr 7 | AAU Asn 9 | AGU Ser 0 |
| AUC Thr 10 | ACC Thr 7 | AAC Asn 10 | AGC Ser 3 |
| AUA Ile 12 | ACA Thr 6 | AAA Lys 14 | AGA Arg 1 |
| AUG Met 9 | ACU Thr 7 | AAG Lys 14 | AGG Arg 0 |
| G | GGU Val 6 | GCC Ala 4 | GGU Gly 12 |
| GUG Val 5 | GCA Ala 3 | GAC Asp 14 | GCC Gly 9 |
| GUA Val 3 | GCA Ala 4 | GAG Glu 18 | GGA Gly 2 |
| GUG Val 5 | GCC Ala 4 | GAG Glu 18 | GGG Gly 3 |
Luciferase α Subunit Sequence

10. Rausch, S. K., Dougherty, J. J., Jr., and Baldwin, T. O. (1982) in Flavins and Flavoproteins (Massey, V., and Williams, C. H., eds) pp. 375–378, Elsevier/North-Holland, New York
11. Messing, J., Crea, R., and Seeburg, P. H. (1981) Nucleic Acids Res. 9, 309–321
12. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463–5467
13. Baldwin, T. O., Nicoli, M. Z., Becvar, J. E., and Hastings, J. W. (1975) J. Biol. Chem. 250, 2763–2768
14. Hastings, J. W., Baldwin, T. O., and Ziegler-Nicoli, M. M. (1978) Methods Enzymol. 57, 135–152
15. Tu, S.-C. (1978) Methods Enzymol. 57, 171–174
16. Laemmli, U. K. (1970) Nature 227, 680–685
17. Baldwin, T. O., and Riggs, A. (1974) J. Biol. Chem. 249, 6110–6118
18. Drapeau, G. R. (1977) Methods Enzymol. 47, 189–191
19. Imamura, T., Baldwin, T. O., and Riggs, A. (1972) J. Biol. Chem. 247, 2785–2797
20. Woods, K. R., and Wang, K.-T. (1967) Biochim. Biophys. Acta 133, 369–370
21. Ziegler-Nicoli, M. M. (1972) Ph.D. thesis, Harvard University
22. Pisano, J. (1972) Methods Enzymol. 25, 27–44
23. Dougherty, J. J., Jr., Rausch, S. K., and Baldwin, T. O. (1982) in Flavins and Flavoproteins (Massey, V., and Williams, C. H., eds) pp. 379–382, Elsevier/North-Holland, New York
24. Baldwin, T. O., Dougherty, J. J., Jr., Rausch, S. K., and Merritt, M. V. (1981) in Bioluminescence and Chemiluminescence: Basic Chemistry and Analytical Applications (DeLuca, M. A., and McElroy, W. D., eds) pp. 121–128, Academic Press, New York
25. Garrett, J., Oguthorpe, D. J., and Robson, B. (1978) J. Mol. Biol. 120, 97–120
26. Chou, P. Y., and Fasman, G. D. (1974) Biochemistry 13, 222–245
27. Holzman, T. F., and Baldwin, T. O. (1980) Biochem. Biophys. Res. Commun. 94, 1199–1206
28. Michalaszy, G. A., and Meighen, E. A. (1976) J. Biol. Chem. 251, 2541–2549
29. Lamfrom, H., Sarabhai, A., and Abelson, J. (1978) J. Bacteriol. 133, 354–365
30. Engebrecht, J., and Silverman, M. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 4154–4158
31. Shine, J., and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U. S. A. 71, 1342–1346
32. Gold, L., Pribnow, D., Schneider, J., Shinedling, S., Singer, B. S., and Stormo, G. (1981) Annu. Rev. Microbiol. 35, 365–403
33. Tu, S.-C., Makemson, J. C., Becvar, J. E., and Hastings, J. W. (1977) J. Biol. Chem. 252, 803–805
34. Oppenheim, D. S., and Yanofsky, C. (1980) Genetics 85, 785–795
35. Schumperli, D., McKenney, K., Sobieski, D. A., and Rosenberg, M. (1982) Cell 30, 865–871
36. Post, L. E., and Nomura, M. (1980) J. Biol. Chem. 255, 4660–4666
37. Parker, J., Johnston, T. C., Borgia, P. T., Holtz, G., Remaut, E., and Fiers, W. (1983) J. Biol. Chem. 258, 10007–10012
38. Swanson, R. (1984) Bull. Math. Biol. 46, 187–203