

Cloning and expression of cDNA for the luciferase from the marine ostracod *Vargula hilgendorffii*

(bioluminescence/oxygenase/luciferin/aequorin/gene expression)

ERIC M. THOMPSON*, SHIGEKAZU NAGATA*, AND FREDERICK I. TSUJI*†

*Osaka Bioscience Institute, 6-2-4 Furuedai, Suita, Osaka 565, Japan; and †Marine Biology Research Division, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093

Communicated by Martin D. Kamen, June 7, 1989 (received for review May 12, 1989)

ABSTRACT The marine ostracod *Vargula hilgendorffii* ejects luciferin and luciferase into seawater to produce a bright luminous cloud. The light is due to the oxidation of luciferin, an imidazopyrazine compound, by molecular oxygen, catalyzed by luciferase. The mechanism of the reaction has been studied extensively and the 60 kcal/mol required for the blue emission have been shown to be derived from the oxidation of luciferin via a dioxetanone intermediate, in which the excited state oxyluciferin bound to luciferase is the emitter. However, only limited information is available regarding the properties of the enzyme. This paper reports the cloning and sequence analysis of the cDNA for *Vargula* luciferase and the expression of the cDNA in a mammalian cell system. The primary structure, deduced from the nucleotide sequence, consists of 555 amino acid residues in a single polypeptide chain with a molecular weight of 62,171. Two regions of the enzyme show significant amino acid sequence homology with an N-terminal segment of the photoprotein aequorin. The *Vargula* luciferase gene, which contains a signal sequence for secretion, should be well suited as a reporter in studies of gene expression.

The marine ostracod crustacean *Vargula hilgendorffii* (1), formerly *Cypridina hilgendorffii* (2), has played a central role in contributing to an understanding of the chemical and physical bases of bioluminescence. *V. hilgendorffii* is indigenous to the south coastal waters of Japan, where it is commonly referred to as "umi botaru" or the "sea firefly." In the past the animal occurred in great abundance, but their numbers have been reduced significantly in recent years. The earliest studies on the bioluminescence of *V. hilgendorffii* were carried out by Harvey (3), and subsequent studies by various investigators resulted in numerous publications (4).

V. hilgendorffii is a small animal (≈3 mm long) with nocturnal habits. It lives in sand at the bottom of shallow waters and becomes an active feeder at night. When disturbed it ejects a copious secretion of luciferin (substrate) and luciferase into sea water, producing a bright luminous cloud. The light results from an enzyme-substrate reaction, catalyzed by luciferase, in which luciferin, an imidazopyrazine compound, is oxidized by molecular oxygen (1, 5). The other products of the reaction are oxyluciferin and carbon dioxide (6, 7). The excited-state oxyluciferin bound to luciferase is the emitter in the reaction (8). Following the determination of the imidazopyrazine structure of *Vargula* luciferin (5), it was found that imidazopyrazine compounds are widely used as substrates in the bioluminescence reactions of marine organisms. One such reaction which has been studied in detail is that of the jellyfish *Aequorea victoria*, from which the cDNA for the calcium-binding photoprotein aequorin has recently been cloned (9, 10).

Interestingly, *V. hilgendorffii* luciferin, or a compound almost identical to it, is used as a substrate by luminescent fish. These include the North American batrachoidid fish *Porichthys notatus*, which possesses more than 700 dermal light organs (11); apogonid fishes of the Far East, in which the light organs are connected to the digestive tract by a duct (12); and the Japanese pempherid fish *Parapriacanthus ransonneti*, in which the light organ communicates directly with the pyloric caeca (13). The luciferins and luciferases of all of these animals give reciprocal light-emitting cross-reactions with *Vargula* luciferin and luciferase. Since luminescent *Vargula* is found in the same waters as these fishes, questions have been raised as to the origin of their luciferins and luciferases (14).

The present study reports the cloning and nucleotide sequence analysis of the cDNA for *Vargula* luciferase and the expression of this cDNA in a mammalian cell system.† Because of the use of imidazopyrazine compounds by taxonomically diverse marine organisms and since the primary structure of aequorin is the only one presently known, it is of interest to compare the structures of *Vargula* luciferase and aequorin. The study of the structures of these two proteins and others should lead to a better understanding of how bioluminescent reactions take place and how the luminescent systems of various marine organisms are related evolutionarily.

MATERIALS AND METHODS

Purification and Sequencing of *Vargula* Luciferase. *Vargula* luciferase, partially purified as described (15), was purified to homogeneity, first by using a tryptamine affinity column (Pierce) equilibrated in 2.0 M NaCl/0.07 M Tris-HCl, pH 7.2, and eluting stepwise with 30% (vol/vol) ethylene glycol/0.17 M NaCl/0.07 M Tris-HCl, pH 7.2, and second, after concentration by ultrafiltration, by using a *p*-aminobenzamide affinity column (Pierce) under the same chromatographic conditions (F.I.T., unpublished work). Aliquots (120 μg) of the purified protein were subjected to endopeptidase digestions by trypsin (Boehringer Mannheim), lysyl endopeptidase (Wako Pure Chemical, Osaka, Japan) or arginyl endopeptidase (Takara Shuzo, Kyoto, Japan). Peptide fragments were separated by reverse-phase chromatography on a C₁₈-protein-peptide HPLC column (Vydac). Undigested luciferase and purified peptides were sequenced by Edman degradation using a gas-phase protein sequenator (Applied Biosystems; model 477A).

mRNA Preparation and cDNA Library Construction. *V. hilgendorffii* were collected (15) at Chiba, Japan, and frozen immediately in liquid nitrogen. The ostracods (wet weight, 5 g) were ground to a fine powder in liquid nitrogen with an ultraturax homogenizer (Janke & Kunkel, Staufen, F.R.G.). Total cellular RNA was extracted from the powder by the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M25666).

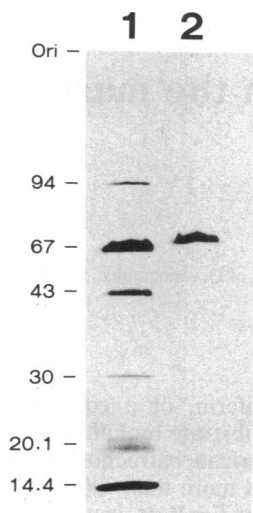


FIG. 1. NaDodSO₄/polyacrylamide gradient gel electrophoresis of purified *Vargula* luciferase. Samples were analyzed on a 10–15% PhastGel (Pharmacia) gradient and proteins were visualized by using a PhastGel silver staining kit (Pharmacia). Lane 1, Pharmacia low molecular weight markers (M_r given $\times 10^{-3}$); lane 2, 50 ng of affinity chromatography purified *Vargula* luciferase.

guanidine thiocyanate/cesium chloride method (16) and poly(A)⁺ RNA was purified by oligo(dT)-cellulose column chromatography (17). The cDNA library was constructed as described (18) except that the double-stranded *Eco*RI digested cDNA was size-selected by two purifications on low-melting agarose (Bio-Rad).

Screening and Analysis of cDNA Clones. Peptide sequences with minimum codon ambiguity were used to synthesize oligonucleotide probes by the phosphoramidite method (19), using a DNA synthesizer (Applied Biosystems; Model 381A). The oligonucleotide probes were 5'-end-labeled (specific activity, $5\text{--}6 \times 10^6$ cpm/pmol) with T4 polynucleotide kinase (Takara Shuzo) and [γ -³²P]ATP (222 TBq/mmol; New England Nuclear) and used to screen 1×10^6 recombinant phages by plaque hybridization (20). Hybridization conditions were as described (21) except that the hybridization temperature was lowered to 28°C and the filters were washed at 37°C with $1 \times$ SSC (0.15 M NaCl/15 mM sodium citrate, pH 7.0). *Eco*RI fragments of positive clones were subcloned in pUC8 and nucleotide sequence analyses were carried out by the dideoxynucleotide chain termination technique (22, 23) using a 7-deaza sequencing kit (Takara Shuzo) and [α -³²P]dCTP (222 TBq/mmol) (New England Nuclear).

Expression of *Vargula* Luciferase. *Hind*III and *Bgl* II linkers (Takara Shuzo) were ligated to *Vargula* luciferase cDNA at the 5' and 3' ends, respectively. The plasmid pRSVL (obtained from S. Subramani, Univ. of California, San Diego) (24) was digested with *Sma* I and ligated with *Bgl* II linker. The linearized plasmid and the cDNA were digested with

*Hind*III and *Bgl* II and the *Hind*III–*Bgl* II fragment containing *Vargula* luciferase cDNA was inserted in place of the firefly luciferase cDNA. The expression vector was designated pRSVVL.

Monkey COS cells (25) (7×10^6) were seeded on 10-cm dishes in 10 ml of Dulbecco's modified Eagle's medium (Nissui Seiyaku, Tokyo) containing 10% fetal calf serum (HyClone). The cells were transfected with 10 μ g of plasmid DNA containing the cDNA insert by using the calcium phosphate method (26, 27). After 48 hr of incubation, the medium was collected and the cells were harvested. Cell extracts were prepared by repeated cycles of freezing and thawing followed by centrifugation (24). To assay for *Vargula* luciferase activity, aliquots of the medium and of the cell extract were diluted with 200 mM Tris-HCl, pH 7.6 (total volume of 1.5 ml) in a 20-ml scintillation vial and placed in a Mitchell-Hastings photometer (28). *Vargula* luciferin, prepared as described (15), was dissolved at 50 nM in 200 mM sodium phosphate buffer, pH 6.8, and 1.5 ml of this solution was injected into the scintillation vial. Conversion of light intensity to quanta per second was made by calibrating the photometer with a [¹⁴C]-hexadecane light standard (29).

RESULTS

Peptide Sequences and Probe Design. When analyzed by polyacrylamide gel electrophoresis in the presence of NaDodSO₄, the purified *Vargula* luciferase showed a single band of M_r 68,000 (Fig. 1). Edman degradation of the intact protein did not yield an assignable amino acid, suggesting that the N terminus is blocked. The luciferase was then digested with endopeptidases, and several peptides were subjected to Edman degradation.

A portion of one of the peptide fragments had the sequence Thr-Met-Glu-Asn-Leu-Asp-Gly-Gln-Lys, which was used to design the complementary oligodeoxynucleotide probe 5'-(T/C)TT(T/C)TGICC(A/G)TCIAGGTT(T/C)TCCATIGT-3'. Deoxyinosine (30) was included at three positions where there was high codon degeneracy. A second complementary probe with A substituted for G at position 15 was also synthesized, and a mixture of these two probes was used to screen the *Vargula* cDNA library. Using the plaque hybridization technique (20), we isolated eight positive clones from the library; two of these, designated VL16 and VL18, with insert lengths of 1.2 and 1.5 kilobases (kb), were selected for further analysis.

Nucleotide Sequence of *Vargula* Luciferase. The restriction map of *Vargula* luciferase cDNA was constructed by using clones VL16 and VL18 as shown in Fig. 2. Clones VL16 and VL18 each had a single internal *Eco*RI site and had a

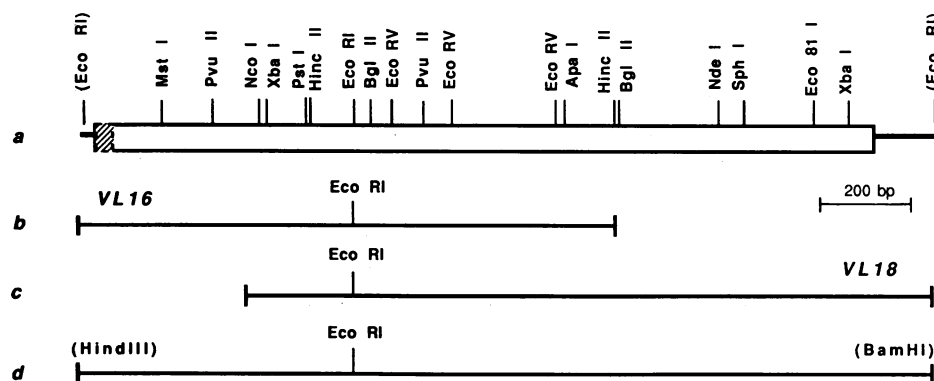


FIG. 2. Restriction map of *Vargula* luciferase cDNA and strategy for the construction of a full-length clone. The restriction map (a) was generated from nucleotide sequence analysis of clones VL16 (b) and VL18 (c). After digestion with *Eco*RI, the 5' fragment of VL16 was ligated to the 3' fragment of VL18 to form the full-length clone (d). The full-length clone was inserted into the expression vector (described in the text) via a 5' *Hind*III linker and a 3' *Bam*HI linker. The shaded portion of the restriction map (a) indicates the portion of the nucleotide sequence believed to code for the putative signal sequence. bp, Base pairs.

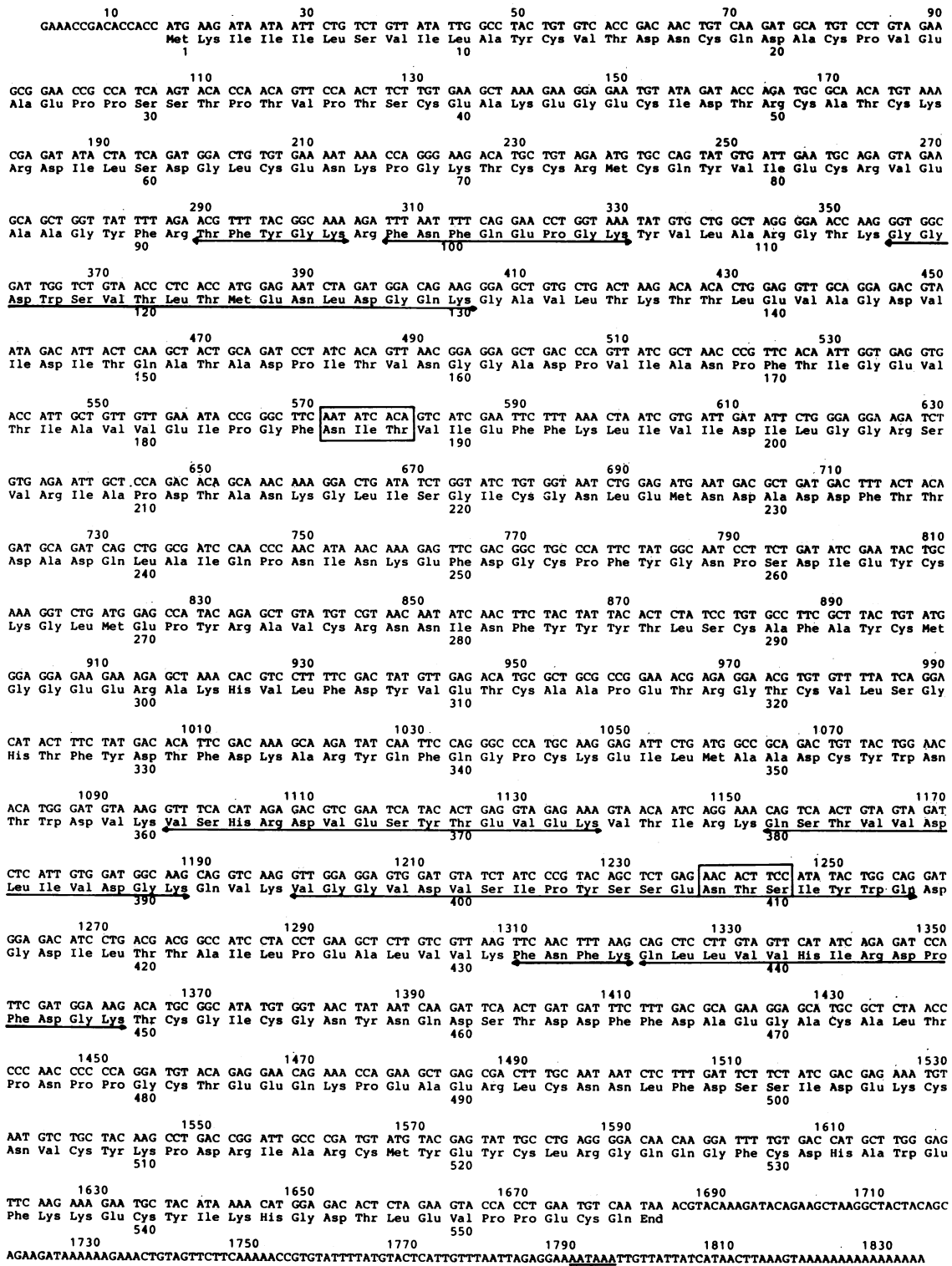


FIG. 3. Nucleotide sequence and predicted amino acid sequence of *Vargula luciferase* cDNA. Numbers above each line refer to nucleotide position and numbers below each line refer to amino acid position. Horizontal arrows indicate regions where the amino acid sequences were identical with those obtained by endopeptidase digestions. Consensus sequences for N-glycosylation are boxed, and the AATAAA polyadenylation signal is underlined.

sequence overlap of 830 nucleotides. Nucleotide sequence analysis demonstrated that clone VL16 encodes the N-terminal portion of luciferase and clone VL18, the C-terminal portion. The complete nucleotide sequence of the *Vargula luciferase* cDNA is shown in Fig. 3 together with the deduced amino acid sequence.

The sequence contains an open reading frame of 1665 nucleotides coding for a protein of 555 amino acids with a

calculated molecular weight of 62,171. The open reading frame contains the amino acid sequence used to construct the oligonucleotide probe. In addition, the amino acid sequences of seven other peptides determined by Edman degradation were in complete agreement with those deduced from the nucleotide sequence. The translation initiation codon was assigned to the first ATG codon at nucleotide position 16. The nucleotide sequence around the putative initiation codon is in

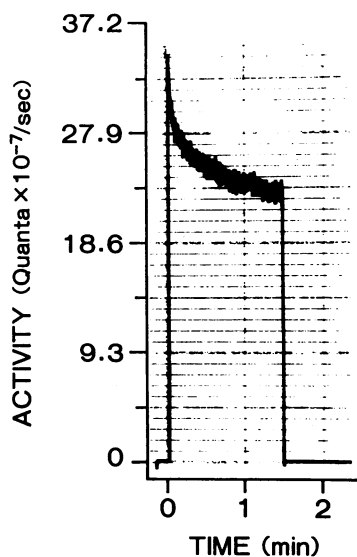


FIG. 4. Assay of active *Vargula* luciferase synthesized and secreted by COS cells. The cells, transfected with the expression plasmid pRSVVL, were incubated for 48 hr and the culture medium was assayed for luciferase activity in the presence of *Vargula* luciferin. Light emission was monitored with a Mitchell-Hastings (28) photometer attached to a Chromatocorder 12 (System Instruments). Shown above is the assay of the culture medium in which the transfected cells were incubated. A scintillation vial containing 100 μ l of the culture medium was placed in the photometer and the shutter was opened. At time zero, 50 nM luciferin was injected into the vial and at 1.5 min the shutter was closed.

good agreement with the consensus sequence CC(A/G)CCAUGG, characteristic of many eukaryotic mRNAs (31). The N-terminal amino acid sequence also has many of the features of a signal sequence (32), in accord with the biological role of luciferase as a secretory protein.

Consistent with the positive staining of luciferase by the periodic acid-Schiff reaction, there are two potential N-glycosylation sites (Asn-Xaa-Ser/Thr), at amino acid residues 186 and 408. At residue 258 the sequence Asn-Pro-Ser occurs, but this sequence is generally not efficiently glycosylated (33). N-glycosidase F was found to reduce the molecular weight of luciferase by 2000–3000, while no reduction in molecular weight was detected in the digestions specific for O-glycosylation (data not shown). These results indicate that *Vargula* luciferase is N-glycosylated and that carbohydrate moieties may account for the difference in the calculated molecular weight of the polypeptide deduced from the nucleotide sequence and that measured for the native protein by gel electrophoresis (Fig. 1) and by gel filtration and sedimentation equilibrium analysis (34). The other notable feature of *Vargula* luciferase is the very cysteine-rich region present in the N-terminal portion, where nine cysteine residues can be found between amino acid residues 39 and 82.

Expression of *Vargula* Luciferase in COS Cells. To establish that the cloned cDNA actually encodes *Vargula* luciferase, a mammalian cell system was used to express the cDNA. As shown in Fig. 2, a full-length cDNA was constructed from clones VL16 and VL18, and it was placed under the promoter of the Rous sarcoma virus long terminal repeat. The hybrid gene was introduced into COS cells and the transfected cells were incubated for 48 hr. As shown in Fig. 4, luciferase activity was clearly detected in the culture medium of transfected COS cells with some luciferase activity also present in the cell extract. Luminescence from the media and cell extracts of COS cells which were either mock (no plasmid DNA) transfected or transfected with pSV0CAT (35) or pRSVL (24) was more than two orders of magnitude less intense than that shown in Fig. 4. These results confirm that the reconstructed cDNA is the full-length cDNA for *Vargula* luciferase and that it also encodes the signal for protein secretion. The luminescence emitted from the culture medium of COS cells was directly proportional to the volume of medium assayed, and a clear signal above background could be detected from as little as 5 μ l of the 10 ml of culture medium obtained from COS cells transfected with the *Vargula* luciferase expression vector.

DISCUSSION

The present results are of interest not only because they contribute to a further understanding of the bioluminescent reaction in *Vargula* but also because they provide a more general understanding of the bioluminescent reactions of a large number of marine organisms in which structurally analogous luciferins and similar reaction mechanisms are used. Among these organisms, a direct comparison can now be made with the bioluminescent reaction of the jellyfish *Aequorea victoria* (36).

Aequorea possesses, in the margin of its umbrella, a calcium-binding protein, aequorin, which emits light in the presence of calcium ions. Aequorin is a complex of apoaequorin (apoprotein), coelenterazine, and molecular oxygen. When aequorin binds calcium ions, a conformational change takes place, converting the protein into an oxygenase. Coelenterazine is then oxidized via an intramolecular reaction, as shown in Fig. 5. The emitter in the reaction is the excited-state coelenteramide bound to apoaequorin (37). The cDNA for apoaequorin has been cloned and expressed, and subsequent analyses have shown apoaequorin to be composed of 189 amino acid residues ($M_r = 21,400$) with three EF-hand structures characteristic of Ca^{2+} -binding sites (9, 10, 38, 39).

Despite the close similarities in substrate structure and mechanisms of the bioluminescent reaction in *Vargula* and *Aequorea*, the enzymes and substrates of the two systems show only slight light-emitting cross-reactions. However, the two proteins are seen to show limited but significant amino acid sequence homologies in two regions of *Vargula* luciferase (residues 97–154 and 353–411) with one in apoae-

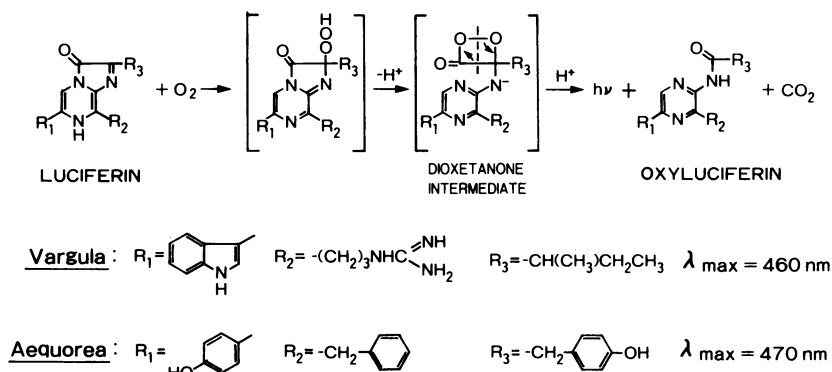


FIG. 5. Bioluminescence reactions of *Vargula* and *Aequorea*. In literature describing the *Aequorea* reaction, luciferin and oxyluciferin are also referred to as coelenterazine and coelenteramide, respectively.

