Secretional luciferase of the luminous shrimp *Oplophorus gracilirostris*: cDNA cloning of a novel imidazopyrazinone luciferase[1]

Satoshi Inouye[a,*], Ken Watanabe[b], Hideshi Nakamura[c], Osamu Shimomura[d]

[a] *Yokohama Research Center, Chisso Corporation, Kanazawa-ku, Yokohama 236-8605, Japan*
[b] *National Institute for Longevity Sciences, Obu, Aichi 474-8522, Japan*
[c] *Graduate School of Bioagricultural Science, Nagoya University, Nagoya 464-8601, Japan*
[d] *Marine Biological Laboratory, Woods Hole, MA 02543, USA*

Received 22 June 2000; revised 11 August 2000; accepted 11 August 2000

Edited by Ned Mantei

Abstract The deep-sea shrimp *Oplophorus gracilirostris* secretes a luciferase that catalyzes the oxidation of coelenterazine to emit blue light. The luciferase (M_r approx. 106 000) was found to be a complex composed of 35 kDa and 19 kDa proteins, and the cDNAs encoding these two proteins were cloned. The expression of the cDNAs in bacterial and mammalian cells indicated that the 19 kDa protein, not the 35 kDa protein, is capable of catalyzing the luminescent oxidation of coelenterazine. The primary sequence of the 35 kDa protein revealed a typical leucine-rich repeat sequence, whereas the catalytic 19 kDa protein shared no homology with any known luciferases including various imidazopyrazinone luciferases.

Key words: Bioluminescence; Luciferin; Decapod; Leucine-rich repeat sequence; Reporter gene

1. Introduction

The deep-sea shrimp *Oplophorus gracilirostris* (H. Milne-Edwards, 1881) ejects a blue luminous cloud from the base of its antennae when stimulated [1–3], like various other luminescent decapod shrimps including those of the genera *Heterocarpus*, *Systellaspis* and *Acanthephyra* [4]. The mechanism underlying the luminescence of *Oplophorus* involves the oxidation of *Oplophorus* luciferin (coelenterazine) with molecular oxygen, which is catalyzed by *Oplophorus* luciferase [2,3] as follows:

$$\text{Coelenterazine} + \text{O}_2 \xrightarrow{\textit{Oplophorus}\ \text{luciferase}} \text{Coelenteramide} + \text{CO}_2 + \text{h}\upsilon\ (\lambda_{\max} = 454\ \text{nm})$$

It is well known that coelenterazine, an imidazopyrazinone compound, is involved in the bioluminescence of a wide variety of organisms as a luciferin or as the functional moiety of photoproteins. For example, the luciferin of the sea pansy *Renilla* is coelenterazine [5], and the calcium-sensitive photoprotein aequorin from the jellyfish *Aequorea* also contains coelenterazine as its functional moiety [6,7].

The previous studies of *Oplophorus* luciferase reported in 1978 [3] showed various interesting properties of this enzyme, such as a high specific activity (1.75×10^{15} photons/s mg at 23°C), a high quantum yield (0.34 at 22°C) and optimum light emission at a high temperature (40°C). Recently, it has been found that the substrate specificity of *Oplophorus* luciferase is unexpectedly broad [8], revealing that bisdeoxycoelenterazine, an analogue of coelenterazine, is an excellent substrate comparable to coelenterazine [9]. Moreover, *Oplophorus* luciferase is a secreted enzyme like the luciferase of the marine ostracod *Cypridina* (*Vargula*) *hilgendorfii* [10] which also uses an imidazopyrazinone-type luciferin to emit light. Some of these properties of *Oplophorus* luciferase clearly distinguish this enzyme from *Renilla* luciferase and aequorin. On the other hand, a luciferase functioning with an imidazopyrazinone-type substrate, such as that of *Oplophorus* or *Cypridina*, would be more favorable for use as a reporter than the luciferase of firefly or luminous bacteria, which requires co-factors in addition to luciferin in the luminescence reaction. The simplicity of the system, combined with the high sensitivity, may make the luciferase of *Oplophorus* a highly useful reporter.

Previously, the molecular weight of *Oplophorus* luciferase was reported to be 130 kDa (by gel filtration) for the native protein, and 31 kDa after treatment with SDS [3]. In our recent study, the luciferase showed a molecular weight of approximately 106 kDa in gel filtration, and we found that the molecule breaks down into 35 kDa and 19 kDa proteins upon sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) analysis. The present paper reports the molecular cloning of the cDNAs encoding the 35 kDa and 19 kDa proteins, and the identification of the protein component that catalyzes the luminescence reaction.

2. Materials and methods

2.1. Purification and amino acid sequence analysis of native Oplophorus luciferase

A purified preparation of *Oplophorus* luciferase obtained in 1978 [3] was further purified by two steps of chromatography. The first step was by hydrophobic interaction chromatography on a column of butyl Sepharose 4 Fast Flow (Pharmacia; 0.7×3.5 cm) using 20 mM Tris–HCl, pH 8.5, eluting with decreasing concentrations of ammonium sulfate starting at 1.5 M. The second step was by gel filtration on a column of Superdex 200 Prep (Pharmacia; 1×48 cm) in 20 mM Tris–HCl, pH 8.0, containing 50 mM NaCl. The specific activity of the purified luciferase was 1.3×10^{15} quanta/s mg, which was slightly lower than the value obtained in 1978 but the protein purity was

*Corresponding author. Fax: (81)-45-786 5512.
E-mail: sinouye@chisso.co.jp

[1] This paper is dedicated to the late Professor Kazuhiko Umesono (deceased 12 April 1999).

Abbreviations: CMV, cytomegalovirus; HPLC, high performance liquid chromatography; PCR, polymerase chain reaction; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis

0014-5793/00/$20.00
PII: S0014-5793(00)01963-3

clearly higher. A solution of purified luciferase (25 μg protein in 0.3 ml) dissolved in 0.1% SDS was subjected to high performance liquid chromatography (HPLC) on a gel filtration column, TSK 3000SW (Toso; 0.75×30 cm), using 20 mM Tris–HCl, pH 7.7, containing 0.1 M NaCl and 0.1% SDS. The elution profile monitored at 280 nm showed two major components, 35 kDa and 19 kDa proteins, and the results were confirmed by SDS–PAGE analysis (12% gel; Tefco, Tokyo, Japan) under reducing conditions [11]. The two protein bands separated by SDS–PAGE were transferred electrophoretically onto a polyvinylidene difluoride membrane (Millipore, Bedford, MA, USA) [12], and the proteins were subjected to amino acid sequence analysis on an Applied Biosystems model 470A gas phase sequencer. The 35 kDa protein was also obtained from the native luciferase (50 μg) by reversed phase HPLC on a 5C4 column (Waters; 0.39×15 cm; solvent: acetonitrile/water/0.1% trifluoroacetic acid). The protein was then digested with lysyl endopeptidase (Boehringer; sequencing grade) at an enzyme:substrate ratio of 1:50. The peptide fragments obtained were separated by reversed phase HPLC on a 5C8 column (Vydac; 0.46×25 cm) by gradient elution with increasing concentrations of acetonitrile (15–55% in 80 min; flow rate 0.5 ml/min) in the presence of 0.1% trifluoroacetic acid, monitoring at 220 nm.

2.2. mRNA preparation and cDNA library construction

Two specimens of *O. gracilirostris* (body size: 40 mm long) were obtained on board in Suruga Bay, Japan, on 19 November 1997, by selecting this species from a catch of the Sakura-ebi (*Sergestes* sp.) shrimp. The live specimens were frozen on dry ice and stored at −80°C until used. Total RNA was prepared by the guanidine isothiocyanate method [13,14], followed by precipitation with 2 M LiCl. The yield of total RNA from two whole specimens (2.8 g wet weight) was approximately 0.9 mg. Poly(A)$^+$ RNA (2 μg) isolated by oligo(dT)-cellulose spun column (Pharmacia, Piscataway, NJ, USA) was used in the synthesis of cDNA with dT_{12-18} primers using a cDNA synthesis kit (Pharmacia) [15]. cDNAs containing *Eco*RI/*Not*I linker sites (20 ng) were ligated to 1 μg of *Eco*RI digested/calf intestinal alkaline phosphatase-treated λZapII vector (Stratagene, La Jolla, CA, USA), then packaged in vitro using a Gigapack Gold III packaging kit (Stratagene). The titer of the cDNA library was 1.1×10^6 plaque-forming units.

2.3. Isolation and nucleotide sequence analyses of clones encoding the 35 kDa and 19 kDa proteins

The *Oplophorus* cDNA library was screened by plaque hybridization using synthetic oligonucleotide probes corresponding to the amino-terminal sequences of the proteins (OL-3 for the 35 kDa protein and SOL-2 for the 19 kDa protein; see Section 3.2) that were labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP (3000 Ci/mmol) [16]. Thirty-five thousand independent plaques (per 15 cm plate) were lifted onto Biodyne A membrane filters (1.2 μm; Pall Corp., Port Washington, NY, USA), then cross-linked with a Stratagene UV cross-linker. The filters were hybridized with the labeled probes in a solution of 1.1 M NaCl/60 mM NaH_2PO_4, pH 7.4/6 mM EDTA/0.2% Ficoll/0.2% SDS/0.02% salmon sperm DNA/0.2% bovine serum albumin at 50°C for 16 h, washed three times in a solution of 300 mM NaCl/30 mM sodium citrate at room temperature, and X-ray film with an intensifying screen was then exposed with the filters at −80°C for 18 h. After isolating positive phage clones, the cDNA inserts were excised as pBluescript phagemids [17], and nucleotide sequence was determined using Applied Biosystems DNA sequencers (models 373 and 377).

2.4. Western blot analysis

Highly purified native *Oplophorus* luciferase (80 μg) emulsified with Freund's complete adjuvant (Calbiochem, La Jolla, CA, USA) was used to immunize a female New Zealand White rabbit. Anti-*Oplophorus* luciferase serum (500 dilution) was used for Western blot analysis as previously described [18].

2.5. Homology search

A search of the non-redundant database of the National Center of Biotechnology Information (NCBI) using the program Gapped Blast [19] was performed. The molecular weight, isoelectric point (p*I*) and hydrophobicity were calculated using DNASIS software Ver. 3.7 (Hitachi Software Engineering, Yokohama, Japan).

2.6. Expression of cDNA for 35 kDa and 19 kDa proteins in Escherichia coli and COS7 cells

The *E. coli* expression system for histidine-tagged 35 kDa and 19 kDa proteins was essentially as previously described [8]. For the 35 kDa protein, an *Xba*I/*Sal*I fragment was obtained from pOL23 as a template by polymerase chain reaction (PCR) amplification (25 cycles; 1 min at 94°C, 1 min at 50°C, 1 min at 72°C) using Gene TaqNT polymerase (Nippon Gene, Toyama, Japan) with primer sets of OL-7 (5′-CCGTCTAGA-GCT-GTT-GCC-TGT-CCT-GCA-GCC-3′; *Xba*I site underlined) and OL-8 (5′-GCCGTCGAC-TTA-TTG-GCA-CAT-TGC-ATG-GAA-3′; *Sal*I site underlined). The fragment was ligated with the *Nhe*I/*Xho*I site of pTrcHis-B (Invitrogen, La Jolla, CA) to give the expression plasmid pHis-OL. For the 19 kDa protein, an *Nhe*I/*Xho*I fragment from pKAZ-412, obtained by PCR with KAZ-3 (5′-CCGGCTAGC-TTT-ACG-TTG-GCA-GAT-TTC-GTT-GGA-3′; *Nhe*I site underlined) and T7-BcaBEST (5′-TAATAC-GACTCACTATAGGG-3′), was inserted into *Nhe*I/*Xho*I site of pTrcHis-B to give the expression vector pHis-KAZ. The expressed protein had an extra 14 amino acid residues at the N-terminus, which included six histidine residues. The host strain used was BL21 (Novagen, Madison, WI, USA). Protein production was induced by the addition of isopropyl-β-thio-galactopyranoside (final concentration: 0.2 mM) at 37°C in Luria–Bertani broth. After incubation of 3 h, cells were harvested, then disrupted by sonication using a Branson model 250 sonifier (Danbury, CT, USA). After centrifugation at 12 000×*g* for 10 min, the supernatant was used for the luminescence assay.

For gene expression in COS7 cells, expression plasmids with and without a putative signal peptide sequence were constructed from pRL-cytomegalovirus (CMV) (Promega, Madison, WI) by replacement of the *Nhe*I/*Xba*I region of the *Renilla* luciferase gene. *Nhe*I/*Xba*I fragments derived from the coding sequence of 35 kDa and 19 kDa proteins was obtained by PCR and inserted into the *Nhe*I/*Xba*I site of pRL-CMV. For the 35 kDa protein, pSOL-CMV was constructed by primer sets of OL-4 and OL-6, and pOL-CMV by primer sets OL-5 and OL-6, where OL-4 is 5′-CCGGCTAGCCACC-ATG-GCT-GTC-AAC-TTC-AAG-TTT-3′ (*Nhe*I site underlined), OL-5 is 5′-CCGGCTAGCCACC-ATG-GCT-GTT-GCC-TGT-CCT-GCA-GCC-3′, OL-6 is 5′-CCGCTCTAGAA-TTA-TTG-GCA-CAT-TGC-ATG-GAA-3′ (*Xba*I site underlined). For the 19 kDa protein, pSKAZ-CMV was produced using primer sets of KAZ-1 and KAZ-5, and pKAZ-CMV by primer sets of KAZ-2 and KAZ-5, where KAZ-1 is 5′-CCGGCTAGCCACC-ATG-GCG-TAC-TCC-ACT-CTG-TTC-ATA-3′, KAZ-2 is 5′-CCGGCTAGCCACC-ATG-TTT-ACG-TTG-GCA-GAT-TTC-GTT-GGA-3′, and KAZ-5 is 5′-CC-GCTCTA-GAA-TTA-GGC-AAG-AAT-GTT-CTC-GCA-AAG-CC-T-3′.

COS7 cells (2×10^5) were grown in a 35 mm well plate containing 3 ml of Dulbecco's modified Eagle's medium (Gibco BRL, Rockville, MD, USA) supplemented with 10% (v/v) heat-inactivated fetal calf serum (Gibco BRL), 100 U/ml penicillin and 100 μg/ml of streptomycin. Cells were cultured at 37°C under a humidified 5% CO_2 atmosphere for 24 h, then transfected with 2 μg of plasmid DNA using FuGENE6 transfection reagent (Rosche Diagnostics, Mannheim, Germany). After further incubation for 36 h, cultured medium and cells were separated by centrifugation. The cells were suspended in 0.5 ml of phosphate-buffered saline (Ca^{2+} and Mg^{2+} free), then stored at −80°C. Before assay, the cells were disrupted by three cycles of freezing and thawing, and 2 μl of the cell extracts and 10 μl of the culture media were used in the luminescence assay.

2.7. In vitro translation

In vitro transcription-coupled translation was carried out using a TNT T7 Quick transcription/translation system with canine pancreatic microsomal membranes (Promega) using pOL-CMV, pSOL-CMV, pKAZ-CMV and pSKAZ-CMV. 25 μl of reaction mixture containing 0.5 μg of plasmid, 20 μl of rabbit reticulocyte lysate, 1 μl of 1 mM methionine and 2.5 μl of microsomal membranes was incubated at 30°C for 90 min, then 1 μl of the reaction mixture was assayed for luminescence activity.

2.8. Assay for luminescence activity and measurement of emission spectra

Coelenterazine and bisdeoxycoelenterazine were chemically synthesized [8]. The total reaction mixture (100 μl) contained coelenterazine or bisdeoxycoelenterazine (1 μg) in 50 mM Tris–HCl, pH 7.6, con-

taining 10 mM EDTA. The reaction was started by the addition of a test sample and the initial intensity of the luminescence was measured by a Lab Science (Tokyo, Japan) model TD-4000 luminometer, equipped with a Hamamatsu R-268 photomultiplier. The initial intensity emitted by 1 μg of native *Oplophorus* luciferase was 1.0×10^5 rlu (relative luminescence unit; 1 rlu = 1.25×10^7 quanta/s).

Luminescence emission spectra were measured on a Hitachi model F4010 fluorescence spectrophotometer at room temperature with the excitation light source turned off.

2.9. Data deposition

The sequences reported in this paper have been deposited in the DDBJ/EMBL/GenBank databases, with accession numbers AB030245 (35 kDa protein) and AB030246 (19 kDa protein).

3. Results and discussion

3.1. Two protein components of native Oplophorus luciferase

The molecular weight of highly active, purified native *Oplophorus* luciferase was estimated to be about 106–000 by gel filtration on a column of Superdex 200 (Fig. 1A). SDS–PAGE analysis of the same luciferase gave two major protein bands corresponding to molecular weights of 35 kDa and 19 kDa, respectively, under both reducing and non-reducing conditions (Fig. 1B). These two proteins were also obtained using reversed phase HPLC. Thus, native *Oplophorus* luciferase is considered to be a complex of 35 kDa and 19 kDa proteins, bound together in a non-covalent manner. The molecular weights and the estimated relative yields of the 35 kDa and 19 kDa proteins suggested that the luciferase is possibly composed of two subunits each of the 35 kDa and 19 kDa proteins.

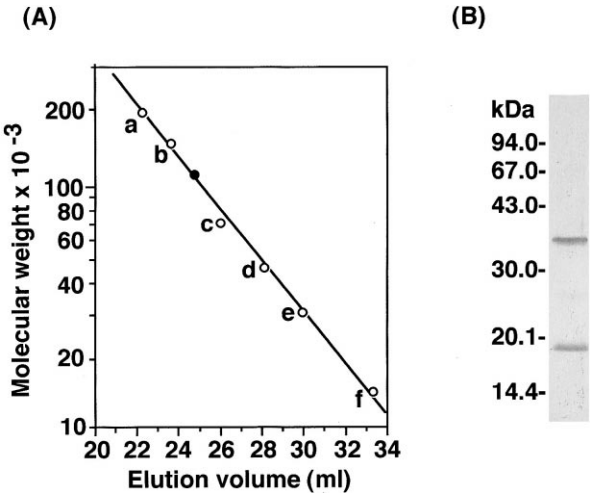

Fig. 1. Determination of molecular weight of native *Oplophorus* luciferase. A: Molecular weight of native *Oplophorus* luciferase determined by Superdex 200 gel filtration. The filled circle represents *Oplophorus* luciferase. The molecular weight markers (Pharmacia Biotech) are as follows: a, amylase (200 000); b, alcohol dehydrogenase (150 000); c, bovine serum albumin (67 000); d, ovalbumin (45 000); e, carbonic anhydrase (29 000); f, ribonuclease (13 700). B: Western blot analysis of purified native *Oplophorus* luciferase on SDS–PAGE gel using anti-*Oplophorus* luciferase antibody. The numbers on the left margin represent the molecular weights of the marker (Pharmacia Biotech): phosphorylase b (94 000), bovine serum albumin (67 000), ovalbumin (43 000), carbonic anhydrase (30 000), soybean trypsin inhibitor (20 100) and α-lactalbumin (14 400).

3.2. cDNA cloning and nucleotide sequence analysis

3.2.1. 35 kDa protein. The amino-terminus of the 35 kDa protein had the sequence Ala-Val-Ala-(Xaa)-Pro-Ala-Ala-Glu-Asp-Ile-Ala-Pro-(Xaa)-Thr-(Xaa)-Lys-Val-Gly-Glu-Gly-Asp-Val-Met-Asp-Met-Asp-(Xaa)-Ser-Lys-Val-Thr-Ser-Asp-Ala-Glu-Leu-Ala-Ser-Phe (Xaa: amino acid not determined). From the underlined sequence, a probe was designed for screening: OL-3, 5′-GTN-GT(T/C)-GTN-ATG-GA(T/C)-ATG-TC-3′, and nine positive clones were isolated from a library of 70 000 independent clones using the plaque hybridization technique. Restriction enzyme analysis of all the clones showed identical restriction maps; one clone, designated pOL23, was subjected to DNA sequence analysis. The complete DNA sequence and the deduced amino acid sequence for pOL23 are shown in Fig. 2A. The open reading frame encodes a protein of 359 amino acids. The experimentally determined amino-terminus of the mature protein begins 39 amino acids downstream, suggesting that the primary translation product contains either an unusually long signal sequence, a propeptide, or both. However, we cannot exclude the possibility that the sample of protein subjected to sequence analysis had undergone some degradation at the amino-terminal end. The amino acid sequences of the peptide fragments derived from lysyl endopeptidase digestion were in agreement with those deduced from the nucleotide sequence, including the amino acid sequence at carboxy-terminus (underline in Fig. 2A). Therefore, the 35 kDa protein should consist of 320 amino acid residues, with a calculated molecular mass of 34 837.08 and an estimated p*I* value of 4.61.

3.2.2. 19 kDa protein. The amino-terminal sequence was found to be Phe-Thr-Leu-Ala-Asp-Phe-Val-Gly-Asp-(Xaa)-Gln-Gln-Thr-Ala-Gly-Tyr-Asn-Gln-Asp-Gln-Val-Leu-Glu-

Table 1
In vitro translation of the 19 kDa and 35 kDa proteins using expression plasmids with and without a signal sequence, in the presence (+) and absence (−) of microsomal membranes, at 30°C for 90 min

Expression plasmid	Expressed protein	Microsomal membranes	Luminescence activity (rlu)
pRL-CMV	*Renilla* luciferase	−	3253.3
pOL-CMV	35 kDa	−	< 0.001
pSOL-CMV	35 kDa*	−	< 0.001
pSOL-CMV	35 kDa*	+	< 0.001
pKAZ-CMV	19 kDa	−	1242.0
pSKAZ-CMV	19 kDa*	−	3.0
pSKAZ-CMV	19 kDa*	+	65.0
pOL-CMV+pKAZ-CMV	35 kDa and 19 kDa	−	581.2
None	None	−	< 0.001

The luminescence activities of the reaction mixtures were assayed using coelenterazine as the substrate. See Section 2 for the details of the expression plasmids. An asterisk indicates 35 kDa or 19 kDa protein with a signal sequence.

(A)

```
          5'  (EcoRI/NotI)TAGCGTAGCTGCATCCTGGTGTCGTCGACCCTCTCCAGCATCATCAT -146
CTGTGGAAGTTCGAACATCTCGCAGAGCAAA ATG GCT GTC AAC TTC AAG TTT AGC CTC CTT -86
                                M   A   V   N   F   K   F   S   L   L  -30
ACC ATA ACC ATT GTT GTT AAT ATC TTA GTC TAT TGC AAT GCA TCA GCA ATT AAA -32
 T   I   T   I   V   V   N   I   L   V   Y   C   N   A   S   A   I   K  -12
TTC GAT GTT GAT TTG GAG AAG GTT CCC TCT AAT GCT GTT GCC TGT CCT GCA GCC  21
 F   D   V   D   L   E   K   V   P   S   N   A   V   A   C   P   A   A   7
                                    -1  +1
GAA GAT ATT GCC CCT TGC ACC TGC AAA GTG GGT GAA GGC GAC GTT ATG GAT ATG  75
 E   D   I   A   P   C   T   C   K   V   G   E   G   D   V   M   D   M  25
GAT TGC TCC AAA GTA ACA AGT GAC GCT GAA CTT GCT TCC ATA TTT AGT AAA ACG 129
 D   C   S   K   V   T   S   D   A   E   L   A   S   I   F   S   K   T  43
TTT CCC TCT AAC ACC TTC CGT GAA TTA TTT ATT GAA TTC AAT CGC GAG ATT ACG 183
 F   P   S   N   T   F   R   E   L   F   I   E   F   N   R   E   I   T  61
ACT CTG ACA GCT GAT AGT TTG GGA GCA GCA ACA TTT ACA AAA ATC GCT ATT ACT 237
 T   L   T   A   D   S   L   G   A   A   T   F   T   K   I   A   I   T  79
AGT TGT ACT CAA TTG AAG ACC ATA GAA GAA AAT GCT TTT ATG GCC AGT GCT GCC 291
 S   C   T   Q   L   K   T   I   E   E   N   A   F   M   A   S   A   A  97
ACA CTC GAG AAA CTC GTG CTC TTA AAA AAT GAT CTT TCC TCT TTT CCT TTT GAA 345
 T   L   E   K   L   V   L   L   K   N   D   L   S   S   F   P   F   E 115
GAA ATG TCA CAA TAC ACA AAA TTA AAT TGG CTT GAA TTA TCC GTA AAT AGC ATT 399
 E   M   S   Q   Y   T   K   L   N   W   L   E   L   S   V   N   S   I 133
ACA GGA TGG CCA GCT CTC TCA TCG GAT ACA CTA GCT AAC CTT ATT TTG TTC CGT 453
 T   G   W   P   A   L   S   S   D   T   L   A   N   L   I   L   F   R 151
AAT CCT ATT GGT AAT ATT CCA GTT GAT GCC TTC CAG ACT CTT CCT AAT ATC GAA 507
 N   P   I   G   N   I   P   V   D   A   F   Q   T   L   P   N   I   E 169
CAA TTC AAC TGC TTC GAT TGT AGC ATC ACC GAA GTG GAA GCA GGT ACT TTT ACT 561
 Q   F   N   C   F   D   C   S   I   T   E   V   E   A   G   T   F   T 187
AGA TCA CCA AAA CTC CAA AAG CTT GTG TTA GGT TAT AAC GGT CTG ACT AGC CTT 615
 R   S   P   K   L   Q   K   L   V   L   G   Y   N   G   L   T   S   L 205
CCC GTA GGC GCC ATC AAA CTC CAT GGA CAT GGC CCA ACC ACT TCC AAC TTG GGT 669
 P   V   G   A   I   K   L   H   G   H   G   P   T   T   S   N   L   G 223
ATC ACC AAT AAT CAG ATC ATC AGT TTC CCC GAG GGT GCT GTT GAA GGC ATC CAA 723
 I   T   N   N   Q   I   I   S   F   P   E   G   A   V   E   G   I   Q 241
GGC ATC CTT GGA ATT GAC TTT AAT CGT GTA ACA TCT CTA AGT GAG GAA GTG TGG 777
 G   I   L   G   I   D   F   N   R   V   T   S   L   S   E   E   V   W 259
CGA CCA ATT TTA GAA AAT CTT TTC CAA TTC AGC TTG CTT AAC AAC CCA CTA GCA 831
 R   P   I   L   E   N   L   F   Q   F   S   L   L   N   N   P   L   A 277
TGT GTA TGT GAC GTA ATG TGG CTT ATT GAT AGC CCA GAA TTG CTG GCA AAA ATT 885
 C   V   C   D   V   M   W   L   I   D   S   P   E   L   L   A   K   I 295
AAA GGC AAT CCC CGA TGT GCC GGT GGA AAA AGA CTC AAG AAT TTG GAT CCA GCT 939
 K   G   N   P   R   C   A   G   G   K   R   L   K   N   L   D   P   A 313
GTT TTC CAT GCA ATG TGC CAA TAA GAAGAAGAAGAAGAATTGAGTCCTCCTGTATCTACTTCT 1002
 V   F   H   A   M   C   Q   *                                         320
GAAGAAGAAGAAGAAGAAGAATCATTAAAATAACAACTAATATTTTTTAAATATAAATCACAATGTATTTA 1074
TACAGTGTAGTGGCAAATACAGTA (NotI/EcoRI) 3'
```

(B)

```
          5'  (EcoRI/NotI) TGTTTGGGTTATAGGTGGTATATCATTAACTCTACTTGAGAGAAG -82
ATG GCG TAC TCC ACT CTG TTC ATA ATT GCA TTG ACC GCC GTT GTC ACT CAA GCT -28
 M   A   Y   S   T   L   F   I   I   A   L   T   A   V   V   T   Q   A -10
TCC TCA ACT CAA AAA TCT AAC CTA ACT TTT ACG TTG GCA GAT TTC GTT GGA GAC  27
 S   S   T   Q   K   S   N   L   T   F   T   L   A   D   F   V   G   D   9
                                -1  +1
TGG CAA CAG ACA GCT GGA TAC AAC CAA GAT CAA GTG TTA GAA CAA GGA GGA TTG  81
 W   Q   Q   T   A   G   Y   N   Q   D   Q   V   L   E   Q   G   G   L  27
TCT AGT CTG TTC CAA GCC CTG GGA GTG TCA GTC ACG CCC ATA CAG AAA GTT GTA 135
 S   S   L   F   Q   A   L   G   V   S   V   T   P   I   Q   K   V   V  45
CTG TCT GGG GAG AAT GGG TTA AAA GCT GAT ATT CAT GTC ATA ATA CCT TAC GAG 189
 L   S   G   E   N   G   L   K   A   D   I   H   V   I   I   P   Y   E  63
GGA CTC AGT GGT TTT CAA ATG GGT CTA ATT GAA ATG ATC TTC AAA GTT GTT TAC 243
 G   L   S   G   F   Q   M   G   L   I   E   M   I   F   K   V   V   Y  81
CCC GTG GAT GAT CAT CAT TTC AAG ATT ATT CTC CAT TAT GGT ACA CTC GTT ATT 297
 P   V   D   D   H   H   F   K   I   I   L   H   Y   G   T   L   V   I  99
GAC GGT GTA ACA CCC AAC ATG ATT GAC TAC TTT GGA AGA CCT TAC CCT GGA ATT 351
 D   G   V   T   P   N   M   I   D   Y   F   G   R   P   Y   P   G   I 117
GCT GTA TTT GAC GGC AAG CAG ATC ACA GTT ACT GGA ACT CTG TGG AAC GGC AAC 405
 A   V   F   D   G   K   Q   I   T   V   T   G   T   L   W   N   G   N 135
AAG ATC TAT GAT GAG AGG CTA ATC AAC CCT GAT GGT TCA CTC CTC TTC AGA GTT 459
 K   I   Y   D   E   R   L   I   N   P   D   G   S   L   L   F   R   V 153
ACT ATC AAT GGA GTC ACG GGA TGG AGG CTT TGC GAG AAC ATT CTT GCC TAA ATT 513
 T   I   N   G   V   T   G   W   R   L   C   E   N   I   L   A   *     169
ACATCTCGAGAATTGCTTAAAGCCTTTTTATGTCTATAAATTGGAGTGGAAAATGTATAATACATATGATT 584
TTTAGGACAGTTATTTTATTTAATTGCTCACTTAAATTTAAATCTGAAGACCACTATAACTGTTCAGAATG 655
GAACTGTAGTCAAACTGTATTAAATGCATTAAAGATCTTATCATATGATTTAGAAAAAAAAAAAAAAAAAA 726
AATAAAAAAAAAAA (NotI/EcoRI)  3'
```

Fig. 2. Nucleotide sequence and predicted amino acid sequence of 35 kDa and 19 kDa proteins derived from *Oplophorus* luciferase. A: 35 kDa protein from pOL23. B: 19 kDa protein from pKAZ-412. Underline indicates the region where the amino acid sequence was identical with that obtained by lysyl endopeptidase digestion. ←

Gln-Gly-Gly-Leu-Ser. The underlined sequence was used to design the probe for screening, SOL-2: 5′-GCN-GGN-TA-(T/C)-AA(T/C)-CA(A/G)-GA(T/C)-CA-3′. From 300 000 independent clones, one positive clone was isolated and the nucleotide sequence was determined. This clone was designated pKAZ-412. As shown in Fig. 2B, the 19 kDa protein contained 196 amino acid residues including a putative signal peptide sequence for secretion. Judging from the results of the N-terminal sequence analysis of the protein, the mature protein should consist of 169 amino acid residues with a calculated molecular mass of 18 689.50 and a p*I* of 4.70.

As for the 35 kDa protein, it is not certain whether the N-terminal extension represents a signal sequence, a propeptide, or both, and again the possibility of amino-terminal degradation cannot be excluded.

3.3. In vitro translation of cDNAs

To identify the catalytic component involved in the luminescence reaction of *Oplophorus* luciferase, cDNAs encoding the 35 kDa and 19 kDa proteins were expressed by means of an in vitro transcription-translation system. Expression plasmids were constructed for the 35 kDa and 19 kDa proteins, with and without a putative signal sequence at the N-terminus, and used as the templates. As summarized in Table 1, significant luminescence activity was found in the 19 kDa protein expressed with pKAZ-CMV, but no activity was observed in the 35 kDa protein expressed with pSOL-CMV or pOL-CMV. These results indicate that the 19 kDa protein is the catalytic component involved in the *Oplophorus* luciferase reaction. The activity of the 19 kDa protein expressed with the pSKAZ-CMV plasmid having a putative signal peptide sequence was very weak. However, an increase of activity was observed when canine pancreatic microsomal membranes with the ability to cleave the signal peptide sequence were included, indicating that the 19 kDa protein had contained the signal peptide sequence. When both pOL-CMV (35 kDa) and pKAZ-CMV (19 kDa) were co-expressed in this system, no enhancement of luminescence activity was observed.

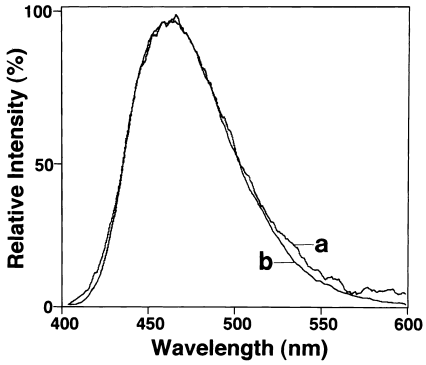

Fig. 3. Bioluminescence spectra of His-tagged 19 kDa protein and of native *Oplophorus* luciferase. Curve a: native *Oplophorus* luciferase (1 μg); curve b: His-tagged 19 kDa recombinant protein extracted from 20 ml of culture, respectively with 1 μg of coelenterazine in 500 μl of 50 mM Tris–HCl, pH 7.6, containing 1 mM EDTA.

3.4. Heterologous expression of cDNAs in E. coli and COS7 cells

Expression plasmids for His-tagged 35 kDa and 19 kDa proteins without the putative signal sequence were constructed and expressed in *E. coli*. However, both proteins were expressed mainly as inclusion bodies (data not shown) and the luminescence activity was detected only in the soluble fraction from the pHis-KAZ expression. The amount of the active 19 kDa protein calculated on the basis of luminescence activity was approximately 50 ng per ml of cell extracts. As shown in Fig. 3, the luminescence spectrum measured with the crude extract of recombinant His-tagged 19 kDa protein was in a good agreement with that of native *Oplophorus* luciferase, showing the light emission peak at about 455 nm.

For expression in mammalian cells (COS7), the plasmids for the 35 kDa and 19 kDa proteins with and without putative signal peptide sequence under the control of the CMV promoter were transfected. The luminescence activities found in the cell extracts and in the culture media are summarized in Table 2. Activity was found only in the extracts of cells transfected with plasmids encoding the 19 kDa protein. In the case of the pSKAZ-CMV expression construct with the putative signal sequence for secretion, the culture medium showed a very low luminescence activity and no band corresponding to the 19 kDa protein could be detected by Western blot analysis using anti-*Oplophorus* antibodies (data not shown). These results suggest that either the putative signal sequence of the

Table 2
Expression of cDNA for 35 kDa and 19 kDa proteins in COS7 cells

Expression plasmid	Luminescence activity (rlu)		
	Medium	Cell extract	
Substrate:	Coelenterazine	Coelenterazine	Bisdeoxycoelenterazine
pRL-CMV	2.9	2059.5	2.6
pOL-CMV	2.5	<0.001	<0.001
pSOL-CMV	2.1	<0.001	<0.001
pKAZ-CMV	3.6	2273.0	2124.0
pSKAZ-CMV	1.9	297.5	217.0
None	1.9	<0.001	<0.001

The plasmids for expressing the 35 kDa and 19 kDa proteins, with and without a signal sequence, were transfected under the control of the CMV promoter. The luminescence activities of the culture media and cell extracts were assayed using coelenterazine or bisdeoxycoelenterazine as the substrate.

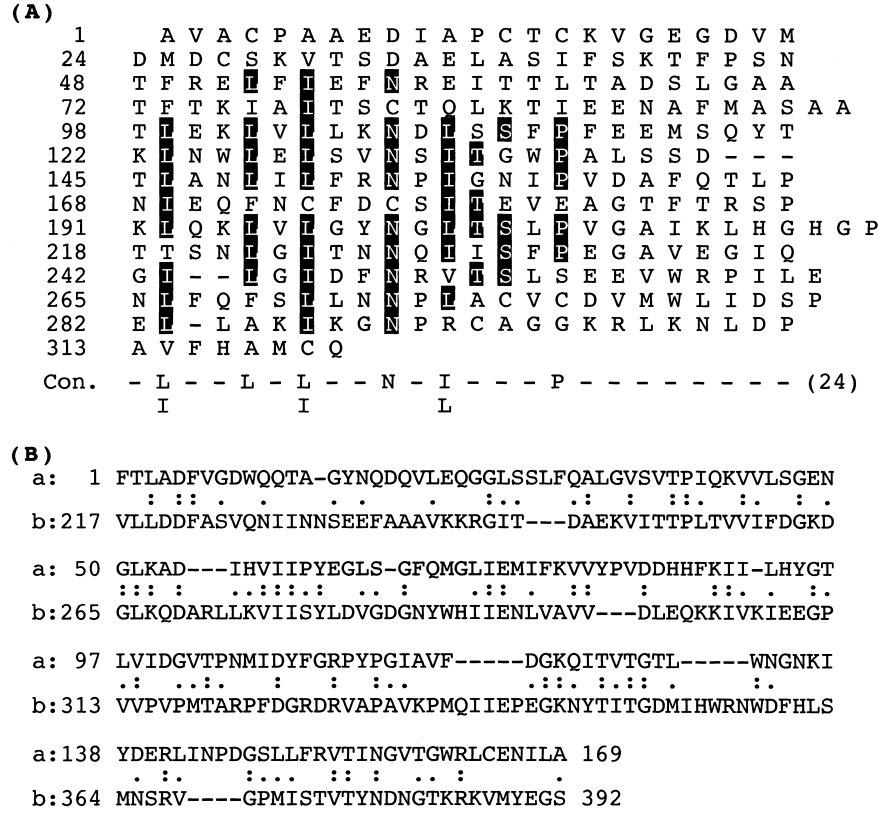

Fig. 4. Structures of the 35 kDa and 19 kDa proteins. A: Alignment of the leucine-rich repeats of 35 kDa protein. The leucine-rich repeats in 35 kDa protein are shown aligned with one another. The criterion for the consensus sequence is that the respective amino acid be present in at least 40% of the cases. The amino acid sequence location is shown on the left. B: Sequence homology between (a) 19 kDa protein and (b) amine oxidase (pir 140924). The double dots indicates identical residues and the single dots indicate amino acids with homologous scores in the Dayhoff mutation data matrix [33]. Numbers indicate the position from the N-terminus of each protein.

19 kDa protein functioned poorly in COS7 cells, as was the case in the in vitro translation system, or the secreted 19 kDa protein was affected by a degradative enzyme.

In the luminescence reaction with the 19 kDa protein, both coelenterazine and bisdeoxycoelenterazine were highly efficient substrates, in agreement with the substrate specificity of native *Oplophorus* luciferase previously reported [8]; this fact may suggest that the function of the 35 kDa protein in the luciferase is unrelated to substrate specificity. When the 19 kDa protein was expressed in COS7 cells using pKAZ-CMV, the activity of cell extracts was high, exceeding that of *Renilla* luciferase expressed with pRL-CMV (Table 2). This suggests that the 19 kDa protein may be a good candidate for use as a new reporter protein in mammalian cell systems.

3.5. Structure and function of the 35 kDa and 19 kDa proteins

As shown in Fig. 4A, the 35 kDa protein contains leucine-rich repeat sequences with the consensus sequence (Leu/Ile)-Xaa-Xaa-Leu-Xaa-(Leu/Ile)-Xaa-Xaa-Asn-Xaa-(Leu/Ile)-Xaa-Xaa-Xaa-Pro (Xaa: any amino acid residue), which is common in many members of the leucine-rich repeat family [20,21]. A homology search of the deduced amino acid sequence using the gapped Blast program revealed a high similarity between the 35 kDa protein and chaoptin, a photoreceptor cell-specific membrane adhesion protein [22]; the calculation gave an E value (homology score) of 8×10^{-9} and showed that 25% of amino acids are identical. The 35 kDa protein contains 11 cysteine residues, but there is no potential *N*-linked glycosylation site with a canonical sequence of Asn-Xaa-(Ser/Thr).

In an effort to clarify the interaction between the 35 kDa and 19 kDa proteins, native luciferase was treated with various detergents and denaturants. Nevertheless, the 35 kDa and 19 kDa proteins formed, as well as the mixture of them, were all found to be inactive in catalyzing the luminescence of coelenterazine, and no catalytic activity could be restored under various conditions tested (data not shown). Thus, the function of the 35 kDa protein remains unknown, although the protein may have a role in the stabilization of the 19 kDa protein.

The 19 kDa protein of *Oplophorus* luciferase is the smallest catalytic component having the luciferase function presently known [23] and its primary structure has no significant homology with any reported luciferase including imidazopyrazinone luciferases [24,25]. The overall amino acid sequence of the 19 kDa protein appears similar to that of *E. coli* amine oxidase (757 amino acid residues; pir 140924) in the region of residues 217–392 (domain of D3–S1) [26], as shown in Fig. 4B, whereas the amino-terminal region (3–49) of the same protein is homologous to the amino-terminal region (1–47) of a fatty acid binding protein (132 amino acid residues; GenBank, L23322) [27] (data not shown). However, it is difficult to rationalize the relationship between the functions and the structural similarities in these examples at present.

Previously, the primary structures were determined for several enzymes and proteins that are functional with an imid-

azopyrazinone compound, including *Renilla* luciferase (36 kDa) [24], aequorin (21.5 kDa) [28], *Renilla* luciferin binding protein (20.5 kDa) [29] and *Cypridina* luciferase (58.5 kDa) [25]. Among these proteins, *Renilla* luciferase (GenBank, M63501) is structurally similar to haloalkane dehalogenase (GenBank, M26950) [30]; aequorin (GenBank, L29571) and *Renilla* luciferin binding protein (SwissProt, P05938) are similar to EF hand-type calcium binding proteins such as calmodulin [31]; and *Cypridina* luciferase (GenBank, M25666) is similar to the von Willebrand factor precursor (GenBank, X04146) [32]. Although the relationship between the structural similarities and the functions of the catalytic domain remains unclear, it is intriguing and possibly significant to find no apparent similarity between the primary structures of *Oplophorus* luciferase and *Renilla* luciferase despite the fact that both luciferases catalyze the oxidation of coelenterazine to emit blue light.

Acknowledgements: We thank Mr. T. Hanai of Shizuoka Prefectural Fisheries Experiment Station and Dr. Y. Shizuri of Marine Biotechnology Institute for assistance in the collection of *Oplophorus*. This work was supported in part by a grant from the National Science Foundation (MCB-9722982) to O.S.

References

[1] Harvey, E.N. (1952) in: Bioluminescence, pp. 345–354, Academic Press, New York.
[2] Johnson, F.H., Stachel, H.-D., Shimomura, O. and Haneda, Y. (1966) in: Bioluminescence in Progress (Johnson, F.H. and Haneda, Y., Eds.), pp. 523–532, Princeton University Press, Princeton, NJ.
[3] Shimomura, O., Masugi, T., Johnson, F.H. and Haneda, Y. (1978) Biochemistry 17, 994–998.
[4] Herring, P.J. (1976) J. Mar. Biol. Assoc. UK 56, 1029–1067.
[5] Inoue, S., Kakoi, H., Murata, M., Goto, T. and Shimomura, O. (1977) Tetrahedron Lett. 31, 2685–2688.
[6] Shimomura, O. and Johnson, F.H. (1975) Nature 256, 236–238.
[7] Head, J.F., Inouye, S., Teranishi, K. and Shimomura, O. (2000) Nature 405, 372–376.
[8] Inouye, S. and Shimomura, O. (1997) Biochem. Biophys. Res. Commun. 233, 349–353.
[9] Nakamura, H., Wu, C., Murai, A., Inouye, S. and Shimomura, O. (1997) Tetrahedron Lett. 38, 6405–6406.
[10] Johnson, F.H. and Shimomura, O. (1978) Methods Enzymol. 57, 331–364.
[11] Laemmli, U.K. (1970) Nature 227, 680–685.
[12] Towbin, H., Staehelin, T. and Gordon, J. (1977) Proc. Natl. Acad. Sci. USA 76, 4350–4354.
[13] Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294–5299.
[14] Inouye, S. and Tsuji, F.I. (1993) FEBS Lett. 315, 242–246.
[15] Kakizuka, A., Yu, R., Evans, R.M. and Umesono, K. (1993) in: Essential Developmental Biology (Stern, C.D., Ed.), pp. 223–232, IRL Press, Oxford.
[16] Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E.H. and Itakura, K. (1981) Nucleic Acids Res. 9, 879–897.
[17] Short, M.J., Fernandez, M.J., Sorge, J.A. and Huse, W.D. (1988) Nucleic Acids Res. 16, 7583–7600.
[18] Inouye, S. and Tsuji, F.I. (1992) Anal. Biochem. 201, 114–118.
[19] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[20] Kobe, B. and Deisenhofer, J. (1994) Trends Biochem. Sci. 19, 415–421.
[21] Buchanan, S.G.St.C. and Gay, N.J. (1996) Prog. Biophys. Mol. Biol. 65, 1–44.
[22] Reinke, R., Krantz, D.E., Yen, D. and Zipursky, S.L. (1988) Cell 52, 291–301.
[23] Wilson, T. and Hastings, J.W. (1998) Annu. Rev. Cell Dev. Biol. 14, 197–230.
[24] Lorenz, W.W., McCann, R.O., Longiaru, M. and Cormier, M.J. (1991) Proc. Natl. Acad. Sci. USA 88, 4438–4442.
[25] Thompson, E.M., Nagata, S. and Tsuji, F.I. (1989) Proc. Natl. Acad. Sci. USA 86, 6567–6571.
[26] Parson, M.R., Convery, M.A., Wilmot, C.M., Yadav, K.D.S., Blakeley, V., Corner, A.S., Phillips, S.E.V., McPherson, M.J. and Knowles, P.F. (1995) Structure 3, 1171–1184.
[27] Becker, M.M., Kalinna, B.H., Waine, G.J. and McManus, D.P. (1994) Gene 148, 321–325.
[28] Inouye, S., Noguchi, M., Sakaki, Y., Takagi, Y., Miyata, T., Iwanaga, S., Miyata, T. and Tsuji, F.I. (1985) Proc. Natl. Acad. Sci. USA 82, 3154–3158.
[29] Kumar, S., Walsh, K.A., Cormier, M.J. and Charbonneau, H. (1990) FEBS Lett. 268, 287–290.
[30] Janssen, D.B., Pries, F., Proeg, J., Kazemier, B., Terpstra, P. and Witholt, B. (1989) J. Bacteriol. 171, 6791–6799.
[31] Tsuji, F.I., Ohmiya, Y., Fagan, F.T., Toh, H. and Inouye, S. (1995) Photochem. Photobiol. 62, 657–661.
[32] Verweij, C.L., Diergaarde, P.J., Hart, M. and Pannekoek, H. (1986) EMBO J. 5, 1839–1847.
[33] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: Atlas of Protein Sequence and Structure, Vol. 5, pp. 345–352, National Biomedical Research Foundation, Washington, DC.