

Characterization and crystallization of active domains of a novel luciferase from a marine dinoflagellate

Liyun Liu,^a Hookang Im,^b
Margaret Cegielski,^c Pierre
LeMagueres,^b L. Wayne Schultz,^c
Kurt L. Krause^{b*} and J. Woodland
Hastings^a

^aDepartment of Molecular and Cellular Biology,
Harvard University, 16 Divinity Avenue,
Cambridge, MA 02138-2020, USA,

^bDepartment of Biology and Biochemistry,
University of Houston, Houston, TX 77204,
USA, and ^cHauptman-Woodward Institute,
73 High Street, Buffalo, NY 14203-1196, USA

Correspondence e-mail: kkrause@uh.edu

Lingulodinium polyedrum luciferase is a bioluminescent protein found in the marine dinoflagellate formerly known as *Gonyaulax*. It is located in organelles called scintillons that emit brief and bright flashes of light that are regulated by an endogenous circadian clock. The complete luciferase molecule has a molecular mass of 136 994 Da and contains three homologous domains, each of which is a separately active luciferase. Two of these domains, D2-LCF and D3-LCF, have been cloned, expressed and crystallized. Crystals of D2-LCF were obtained from PEG 10 000 in space group $P2_12_12_1$, with unit-cell parameters $a = 49.1$, $b = 104.7$, $c = 180.3$ Å. They diffract to 2.9 Å on a rotating anode. Crystals of D3-LCF were grown from PEG 2000 in space group $P2_12_12_1$, with unit-cell parameters $a = 58.86$, $b = 63.98$, $c = 95.76$ Å. They diffract to 2.3 Å on a rotating anode.

Received 15 November 2002

Accepted 4 February 2003

1. Introduction

Lingulodinium polyedrum (Fig. 1; formerly *Gonyaulax polyedra*) is one of several marine unicellular dinoflagellates that emit light as brief (~100 ms) and bright (~ 10^{10} photons) flashes that can be observed when ocean water is disturbed at night (Hastings & Dunlap, 1986); the bioluminescence is regulated by an endogenous circadian clock (Hastings, 2001). Light is emitted from numerous (~400 per cell) small (~0.4 µm in diameter) organelles called scintillons, which occur as bulbous cytoplasmic projections into the vacuole (Nicolas *et al.*, 1987; Fritz *et al.*, 1990). These contain a luciferase (LCF) which, in concert with its tetrapyrrole luciferin substrate (Nakamura *et al.*, 1989), is responsible for light generation. In this species (but not all others; Schmitter *et al.*, 1976; Knaust *et al.*, 1998), the

scintillons contain a second protein, the luciferin-binding protein (LBP; Nicolas *et al.*, 1991; Morse *et al.*, 1989; Desjardins & Morse, 1993).

The luciferase of *L. polyedrum* has a molecular mass of 136 994 Da and comprises three intramolecularly homologous domains, each containing a catalytic site, preceded by an N-terminal region that is homologous to the N-terminal region of the LBP (Li *et al.*, 1997). Each homologous domain is a separately active luciferase; the sequences of the three are approximately 75% identical overall and ~95% identical in the more central regions, presumably the location of highly conserved catalytic domains. In addition, the frequency of synonymous (silent) nucleotide substitutions in the central regions of the repeats is extremely low, suggesting that those regions in either the DNA or RNA have a function in addition to coding for protein, possibly binding a regula-

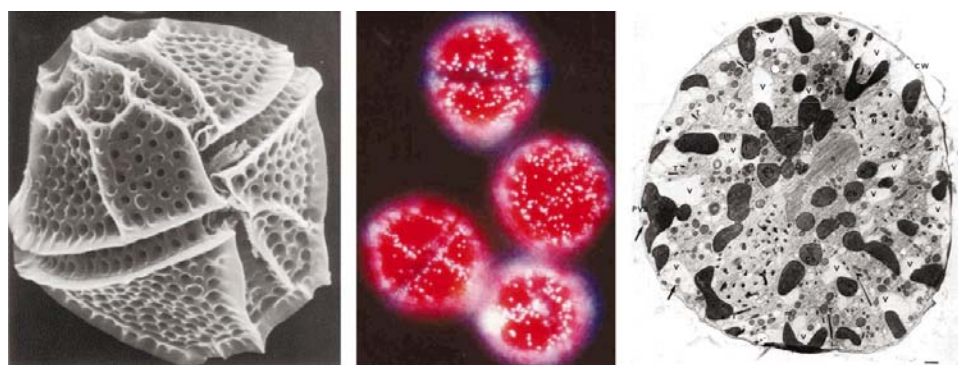


Figure 1
(a) Illustration of the test or shell of *L. polyedrum* (after Kofoid). (b) Fluorescence microscopy of four *L. polyedrum* cells showing the many light-emitting organelles (scintillons) with a background red fluorescence of chlorophyll from chloroplasts. Zeiss filter set #487718; excitation, 395–420 nm; emission, 450–560 nm (Fritz *et al.*, 1990). (c) Transmission electron micrograph of whole cell of *L. polyedrum*. Arrows point to the locations of scintillons (Nicolas *et al.*, 1987).

tory element. In the genome, the *lcf* gene is organized as tandem repeats (Li & Hastings, 1998) and, like the gene for LBP (Lee *et al.*, 1993), does not contain introns. The spacer region between the *lcf* genes has no TATA box or other known conserved eukaryotic promoter elements, but a candidate promoter sequence has been identified by comparison with the dinoflagellate peridinin-chlorophyll binding protein (Le *et al.*, 1997; Li & Hastings, 1998).

An important feature of the system is that a pH change appears to trigger the flash (Krieger & Hastings, 1968; Morse *et al.*, 1989). At a high pH the luciferin is bound to LBP and LCF is not active, a property shared by peptides of each of its three domains (Li *et al.*, 2001). As a consequence, the reaction occurs maximally at pH 6.3 and virtually not at all at pH 7.5 and above. Flashing is postulated to be the result of a rapid and transient change in pH within the scintillon caused by the opening of channels in the membrane, allowing protons to enter rapidly from the acidic vacuole (Nicolas *et al.*, 1987).

Recent searches yielded no sequences in the database that had even the slightest indication of sequence similarity to the three repeat domains of dinoflagellate LCF (Hastings & Wood, 2001). The discovery that the luciferase contains three homologous repeat units, referred to as D1, D2 and D3, led to their individual expression and characterization. The luciferase activity of each domain is less than that of the full-length protein, owing at least in part to the fact that only one active site is present (Li *et al.*, 1997). Activity, but a decreased amount, is retained upon the removal of the first ~50 N-terminal amino acids of any one of the domains. Notably, these truncated domains retain significant activity at pH 8, although it is decreased from that at pH 6.3. Four histidines in this N-terminal region have been implicated in the regulation of activity by pH (Li *et al.*, 2001).

Knowledge of the individual structures of these domains, as well as their configuration in the full-length molecule, should result in a clearer understanding of the mechanism of the light-emitting reaction and its regulation by pH change. Here, we report the cloning and expression of individual domains, the crystallization of these domains and their initial characterization using X-ray diffraction.

2. Materials and methods

Earlier studies established that the full-length *L. polyedrum* luciferase gene, as well

as each of its three individual repeat domains, can produce enzymatically active proteins when expressed as a GST fusion protein in *Escherichia coli*. We produced two different GST luciferase constructs called GST-D2-LCF and GST-D3-LCF. Derived from the second and third repeat domains of LCF, each construct has a molecular mass of about 44 000 Da. When efforts to obtain crystals were not successful, it was considered that the GST tag might be interfering with crystallization.

As an alternative, domains 2 and 3 were overexpressed in pQE30 (Qiagen, CA, USA) tagged with six histidines at the N-terminus. The yields were quite high (>6 mg l⁻¹) and protein that was >95% pure was eluted from the nickel-agarose resin. After gel filtration with Sephacryl 100, the purities of the eluted constructs were 97–99%, as judged by densitometry of the stained PAGE gels. Both domains proved crystallizable, but the needle-shaped crystals were inadequate for diffraction analysis. Again, it was considered that the tag might be interfering with the formation of favorable crystals.

We therefore produced tag-free recombinant peptides of the D2 and D3 luciferase domains in *E. coli* by inserting a TEV protease site between the histidine tag and the luciferase and then removing the tag by protease treatment. We chose the TEV protease for two reasons. It retains about 80% of its maximum activity at 277 K, an advantage to us because low temperature favors the stability of the luciferase. Also, TEV protease can be supplied in a histidine-tagged form (Life Technology Company) and can thus be easily removed along with any undigested His-tagged luciferase by passing the reaction mixture through a nickel column. To incorporate the TEV protease site, we amplified the *lcf* insert using the modified forward primer, which contains an additional stretch of nucleotides at the 5'-terminus for coding the peptide sequence of the TEV site (ENLYFQG), and cloned it into pQE30.

Following this strategy, we produced tag-free D2 and D3 proteins, which after purification by gel filtration were >97% pure. Western analysis with anti-histidine monoclonal antibody demonstrated that the histidine tag was removed from the TEV-treated samples. Both tag-free peptide constructs resulted in crystals suitable for diffraction, as described below.

Selenomethionine derivatives of D2 and D3 were generated using a method similar to that described by Van Duyne *et al.* (1993). Briefly, *E. coli* BL21 cells carrying a plasmid

with luciferase D2 or D3 were incubated at 310 K in 25 ml LB containing 50 µg ampicillin and 15 µg kanamycin per millilitre of medium. After 10 min, the above 25 ml was added to 25 ml of fresh medium containing ampicillin and kanamycin and placed in a 310 K shaking incubator (250 rev min⁻¹) for about 10 h. The culture was then centrifuged at 3000g for 15 min. The pellet was re-suspended in 25 ml of M9 medium and divided among four 4 l flasks containing pre-warmed M9. At this point, 8 ml of 50% glucose, 1 ml thiamine (5 mg ml⁻¹), 1 ml of 1 M MgSO₄, 0.1 ml of 1 M CaCl₂ and fresh ampicillin and kanamycin were added. All solutions were either autoclaved or sterilized by filtration. The flasks were allowed to shake at 265 rev min⁻¹, 310 K until OD₆₀₀ reached 0.4. The temperature was then reduced to 295 K and 10 ml of an amino-acid cocktail was added to each flask. The cocktail consisted of 10 mg ml⁻¹ each of lysine, threonine and phenylalanine and 5 mg ml⁻¹ each of leucine, isoleucine and valine. At this time D,L-selenomethionine (2 ml of 25 mg ml⁻¹) was also added. After 15 min, the cells were induced with 0.1 M IPTG and shaken at 250 rev min⁻¹ at 295 K overnight. The cells were harvested by centrifugation for 25 min at 3000g, 277 K, resuspended in 300 mM NaCl, 50 mM phosphate buffer pH 8.0 and stored at 193 K. Purification was carried out as described for the non-selenomethionine construct.

3. Results and discussion

3.1. Crystallization of D2 and D3 constructs

D2-LCF crystals were grown by equilibrating protein (7 mg ml⁻¹, 20 mM Tris-HCl pH 8.0, 1 mM EDTA, 25 mM NaCl, 2 mM DTT) in sitting drops against 18% PEG 10 000, 0.1 M Bicine at pH 7.7. The 10 µl total volume of the sitting drop was composed of equal amounts of protein and reservoir solution. After four weeks, crystals of untagged D2-LCF were obtained which had dimensions of 0.4 × 0.08 × 0.05 mm (Fig. 2a). Crystals containing selenomethionine had the same habit, but grew at a slightly lower PEG concentration. Both native and selenomethionine crystals have essentially the same unit-cell parameters $a = 49.1$, $b = 104.7$, $c = 180.3$ Å, but the selenomethionine crystals displayed better diffraction properties.

Crystals of selenomethionine D2-LCF were transferred to a cryoprotectant solution (13% PEG 10 000, 0.075 M Bicine pH 7.7, 25% glycerol) 30 min prior to mounting. Using an R-AXIS IV++ image-plate

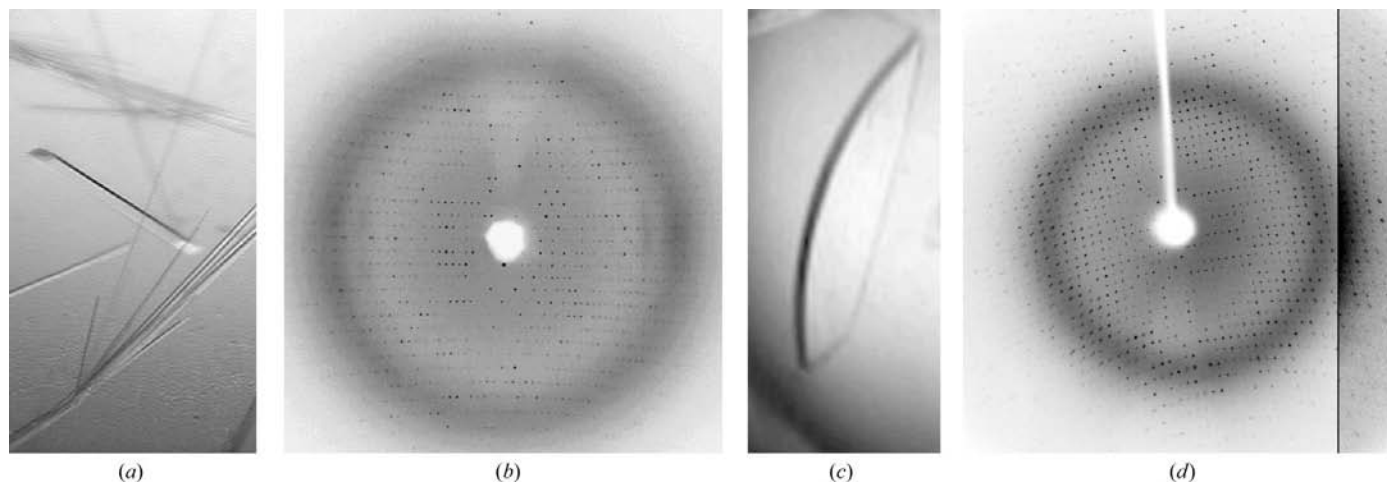


Figure 2

(a) Crystals of D2-LCF from *L. polyedrum*. The largest central crystal is of dimensions $0.4 \times 0.08 \times 0.05$ mm. (b) A 0.5° oscillation photograph of selenomethionine D2-LCF crystals taken as described within the text. The maximum resolution at the edge of the frame is 2.6 \AA . (c) Crystals of D3-LCF from *L. polyedrum*. The crystals have typical dimensions of $0.8 \times 0.1 \times 0.05$ mm. (d) A 2° oscillation photograph of selenomethionine D3-LCF crystals taken on an R-AXIS IV using X-rays generated at 50 kV and 92 mA from a copper rotating-anode generator. The maximum resolution at the corner of the frame is 2.1 \AA . The band at the right is at increased contrast to allow inspection of the higher resolution spots.

detector mounted on a MicroMax 007 operating at 40 kV, 20 mA, a data set to 2.9 \AA containing 238 frames of 0.5° per frame was collected at 93 K. Using *Crystal-Clear* (Rigaku/MS, 1999), these data were indexed and processed with statistics as shown in Table 1. Following analysis of the unit cell, diffraction pattern symmetry and systematic absences, we made a space-group assignment of $P2_12_12_1$.

We are anticipating two molecules of D2-LCF per asymmetric unit. Calculation of the Matthews coefficient for two D2-LCF molecules yields a value of $2.6 \text{ \AA}^3 \text{ Da}^{-1}$ and a solvent content of 53% (Matthews, 1968). Similar calculations for three molecules per asymmetric unit yielded values of $1.75 \text{ \AA}^3 \text{ Da}^{-1}$ and 29%, respectively.

For the tag-free D3 construct, initial crystallization conditions were determined from a 1536 high-throughput crystallization scheme (Luft *et al.*, 2001). Upon optimization of crystallization conditions, this protein produced plate-like crystals. Native data were collected to 2.1 \AA on beamline F1 at the Cornell High Energy Synchrotron Source (CHESS). However, these data were highly mosaic and could not be indexed acceptably. Subsequently, using the hanging-drop method, we crystallized the selenomethionine derivative, which yielded crystals that were superior to those of the native protein. The protein solution used for the crystallization was $8\text{--}10 \text{ mg ml}^{-1}$ selenomethionine D3 luciferase in 10% glycerol, 25 mM NaCl and 20 mM Tris pH 8.0. They were equilibrated against 18–20% methyl ether PEG 2000, 100 mM EPPS buffer pH 7.8–8.4 and have typical dimensions of

Table 1

Data-collection statistics for D2-LCF and D3-LCF.

Values in parentheses are for the last resolution bin: $3.0\text{--}2.9 \text{ \AA}$ for D2-LCF and $2.38\text{--}2.3 \text{ \AA}$ for D3-LCF.

	D2-LCF	D3-LCF
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters		
<i>a</i> (Å)	49.1	58.86
<i>b</i> (Å)	104.7	63.98
<i>c</i> (Å)	180.3	95.76
$\alpha = \beta = \gamma$ (°)	90.0	90.0
Resolution (Å)	2.9	2.3
<i>T</i> (K)	93	93
Observations	92737	334994
Unique reflections	20440	29595
Completeness (%)	95.5 (92.6)	95.0 (90.0)
R_{merge}^\dagger (%)	10.1 (33.3)	7.2 (21.5)
$\langle I/\sigma(I) \rangle$	15.6 (3.5)	24.6 (7.4)

$$\dagger R_{\text{merge}} = \sum |I_i - \langle I \rangle| / \sum \langle I \rangle.$$

$0.8 \times 0.1 \times 0.05$ mm (Fig. 2*b*). Crystals were cryopreserved in a solution of 20% glycerol, 18% methyl ether PEG 2000, 100 mM EPPS buffer pH 7.0, mounted in 1 mm cryoloops (Hampton) and frozen directly in the nitrogen cold stream (133 K).

Highly redundant data from the D3-LCF SeMet crystals were collected on an R-AXIS IV using X-rays generated at 50 kV and 92 mA from a copper rotating-anode generator (Fig. 2*d*). Data were measured using $2.5\text{--}3^\circ$ oscillations at 10 min deg^{-1} for a total of 720° . The crystal-to-detector distance was 194 mm. Data were processed using *DENZO* and *SCALEPACK* with statistics as shown in Table 1 (Otwinowski & Minor, 1997). Notably, the data collected were highly redundant, with 75% of the reflections having >9 observations. Analysis of the diffraction pattern symmetry and

systematic absences supports a space-group assignment of $P2_12_12_1$.

Since no homologous proteins are present in the Protein Data Bank, solution of the phase problem will require use of a selenium-based anomalous phasing technique. The structural solution for these two constructs should yield closely related structures that may contain a novel fold. Following solution of the D2 and D3 constructs, crystals will be sought for other related constructs such as the D1-D2-D3 protein, the holoenzyme and the holoenzyme plus luciferin-binding protein. Key questions to be answered involve the description of the luciferin-binding region, the mechanism of pH-triggered luminescence and the nature of the interactions between each individual domain and its two related neighbors, as well as with the luciferin-binding protein.

We would like to acknowledge Dr Catherine Klein and Dr Kris Tesh for the use of their data-collection facilities at Rigaku MSC, The Woodlands, Texas, USA. We also thank the Institute of Molecular Design, University of Houston and Molecular Simulations Inc. for the use of computational resources. This publication was supported in part by grants from the Robert A. Welch Foundation (KLK) and the Office of Naval Research (JWH – N000149910412 and LWS – N000140210188).

References

- Desjardins, M. & Morse, D. (1993). *Biochem. Cell Biol.* **71**, 176–182.

- Fritz, L., Morse, D. & Hastings, J. W. (1990). *J. Cell Sci.* **95**, 321–328.
- Hastings, J. W. (2001). *Handbook of Behavioral Neurobiology*, Vol. 12, *Circadian Clocks*, edited by J. Takahashi, F. Turek & R. Y. Moore, pp. 321–334. New York: Plenum.
- Hastings, J. W. & Dunlap, J. C. (1986). *Methods Enzymol.* **133**, 307–323.
- Hastings, J. W. & Wood, K. V. (2001). *Photobiology for the 21st Century*, edited by D. Valenzano & T. Coohill, pp. 199–210. Overland Park, KS, USA: Valdenmar.
- Knaust, R., Urbig, T., Li, L., Taylor, W. & Hastings, J. W. (1998). *J. Phycol.* **34**, 167–172.
- Krieger, N. & Hastings, J. W. (1968). *Science*, **161**, 586–589.
- Le, Q. H., Markovic, P., Hastings, J. W., Jovine, R. V. M. & Morse, D. (1997). *Mol. Gen. Genet.* **255**, 595–604.
- Lee, D.-H., Mittag, M., Sczekan, S., Morse, D. & Hastings, J. W. (1993). *J. Biol. Chem.* **268**, 8842–8850.
- Li, L. & Hastings, J. W. (1998). *Plant Mol. Biol.* **36**, 275–284.
- Li, L., Hong, R. & Hastings, J. W. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 8954–8958.
- Li, L., Liu, L., Hong, R., Robertson, D. & Hastings, J. W. (2001). *Biochemistry*, **40**, 1844–1849.
- Luft, J. R., Wolfley, J., Jurisica, I., Glasgow, J., Fortier, S. & DeTitta, G. T. (2001). *J. Cryst. Growth*, **232**, 591–595.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Morse, D., Pappenheimer, A. M. & Hastings, J. W. (1989). *J. Biol. Chem.* **264**, 11822–11826.
- Nakamura, H., Kishi, Y., Shimomura, O., Morse, D. & Hastings, J. W. (1989). *J. Am. Chem. Soc.* **111**, 7607–7611.
- Nicolas, M.-T., Morse, D., Bassot, J.-M. & Hastings, J. W. (1991). *Protoplasma*, **160**, 159–166.
- Nicolas, M.-T., Nicolas, G., Johnson, C. H., Bassot, J.-M. & Hastings, J. W. (1987). *J. Cell Biol.* **105**, 723–735.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Rigaku/MSD (1999). *CrystalClear*. Molecular Structure Corporation, The Woodlands, Texas, USA.
- Schmitter, R., Njus, D., Sulzman, F. M., Gooch, V. & Hastings, J. W. (1976). *J. Cell. Physiol.* **87**, 123–34.
- Van Duyne, G., Standaert, R., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.