

## Members of a Dinoflagellate Luciferase Gene Family Differ in Synonymous Substitution Rates<sup>†</sup>

O. Keith Okamoto,<sup>‡</sup> Liyun Liu,<sup>‡</sup> Deborah L. Robertson,<sup>§</sup> and J. Woodland Hastings<sup>\*‡</sup>

Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138-2020, and  
Biology Department, Clark University, Worcester, Massachusetts 01610-1477

Received August 13, 2001; Revised Manuscript Received October 22, 2001

**ABSTRACT:** Regulation and evolution of dinoflagellate luciferases are of particular interest since the enzyme is structurally unique and bioluminescence is under circadian control. In this study, three new members of the dinoflagellate luciferase gene family were identified and characterized from *Pyrocystis lunula*. These genes, *lcfA*, *lcfB*, and *lcfC*, also exhibit the unusual structure and organization previously reported for the luciferase gene of a related dinoflagellate, *Lingulodinium polyedrum*: three repeated domains, each encoding an active catalytic site, multiple gene copies, and tandem organization. The histidine residues involved in the pH regulation of *L. polyedrum* luciferase activity, and implicated in the regulation of flashing, are also fully conserved in *P. lunula*. The interspecific conservation between the individual luciferase domains of *P. lunula* and *L. polyedrum* is higher than among domains intramolecularly, indicating that this unique gene structure arose through duplication events that occurred prior to the divergence of these dinoflagellates. However, *P. lunula* luciferase genes differ from *L. polyedrum* in several respects, notably, the occurrence of an intron in one gene (*lcfC*), a 2.25-kb intergenic region connecting *lcfA* and *lcfB*, and, of particular interest, an invariant rate of synonymous (silent) substitutions along the repeat domains, in contrast to *L. polyedrum* luciferase, where the occurrence of synonymous substitutions is practically absent in the central region of the domains.

Dinoflagellates are a diverse group of unicellular eukaryotes with distinct and unusual genomic characteristics, including a large genome (up to 60 times more DNA than a haploid human nucleus) packed in permanently condensed chromosomes lacking histones (1, 2). Although many dinoflagellate genes have been cloned and sequenced, genes involved in bioluminescence, namely, those for luciferase, which catalyzes the light-emitting reaction (3), and the luciferin (substrate)-binding protein (LBP)<sup>1</sup> (4), have been characterized only in *Lingulodinium polyedrum* (formerly *Gonyaulax polyedra*). When bound to LBP, luciferin (a novel tetrapyrrole) is prevented from reacting with luciferase; its release and the ensuing luminescence are attributed to a transient pH change (5). The expression of these two proteins is controlled by a circadian clock at the translational level, with synthesis and degradation occurring on a daily basis (6, 7).

The *L. polyedrum* luciferase is structurally unique, comprising a short N-terminus followed by three homologous domains (D1, D2, and D3), each enzymatically active (8).

In *L. polyedrum*, the *lcf* and *lbp* genes lack introns and are present in high copy number in the genome, the former being organized in tandem (9). Strikingly, these two genes have no similarities to luciferases from organisms in other phyla, or indeed to any sequences in the database (10), supporting the hypothesis of independent evolution of bioluminescence proteins in different phylogenetic groups (11).

Within dinoflagellates, bioluminescence has been described in about 6% of all genera and only in marine species (12). The bioluminescence system is similar among species with respect to cross-reactions between enzymes and substrates, pH-activity profiles, and its cellular localization to light emitting organelles called scintillons (13–15). However, there are clear differences in the *Pyrocystis lunula* system, where flashing but not glow has a circadian rhythm, and the control of flashing seems different, since LBP is absent. Also, the amounts of luciferin and luciferase are constant throughout day and night (16–18); thus, instead of daily de novo synthesis and destruction of all components, the rhythm appears to involve changes in their intracellular localization (19, 20).

To gain insight into evolution and molecular control of bioluminescence in this group, we examined the structure of other dinoflagellate luciferase genes and identified elements possibly involved in regulation. Three new genes from the species *P. lunula* were characterized, revealing that the primary structure of dinoflagellate luciferases is highly conserved. Sequence similarity and phylogenetic analyses of *P. lunula* and *L. polyedrum* luciferase domains provide evidence for the origin of this unique gene structure prior to

<sup>†</sup> This work was supported in part by grants from Fundação de Amparo a Pesquisa do Estado de São Paulo (00/01843-1), National Science Foundation (MCB 9982880), and Office of Naval Research (N00014-99-1-0412).

<sup>\*</sup> To whom correspondence should be addressed: J. Woodland Hastings, 16 Divinity Avenue, Cambridge, MA 02138-2020, Telephone: (617) 495-3714, Fax: (617) 496-8726, E-mail: <Hastings@fas.harvard.edu>.

<sup>‡</sup> Harvard University.

<sup>§</sup> Clark University.

<sup>1</sup> Abbreviations: Lcf, luciferase; LBP, luciferin-binding-protein; UTR, untranslated region; ORF, open reading frame.

the divergence of these genera during evolution. More striking is the finding that, while the three intramolecularly repeated domain organization is present in the luciferase genes of both dinoflagellates, there are substantial differences in the frequency of synonymous substitutions along their repeat domains.

## EXPERIMENTAL PROCEDURES

**Culture and Strain Information.** *P. lunula* cells (CCMP 731) were obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton, and grown in f/2 medium as previously described (18). Cells were harvested by filtration with Whatman paper No. 541, ground in liquid nitrogen, and immediately used for nucleic acid isolation.

**Construction and Screening of *P. lunula* cDNA Library.** Total RNA was isolated in extraction buffer (100 mM Tris pH 9, 200 mM NaCl, 15 mM EDTA, 0.5% sarcosyl, 100 mM  $\beta$ -mercaptoethanol) from night-phase *P. lunula* cells by the phenol/chloroform based method (21), and mRNA purified using biotinylated oligo(dT<sub>20</sub>) primers (Promega). First strand cDNA was synthesized from 4  $\mu$ g of mRNA using Superscript II reverse transcriptase (Life Technologies) and oligo(dT<sub>20</sub>) primers. Second strand synthesis and subsequent cloning steps were performed following the manufacturer's protocol (ZAP-cDNA synthesis kit, Stratagene).

Approximately 10<sup>6</sup> phages were plated on 20 NZY agar plates and screened with [ $\alpha$ <sup>32</sup>P]dCTP-labeled luciferase cDNA from *L. polyedrum*. *Escherichia coli* cells (XL1 blue), transformed with pBluescript II phagemids containing luciferase sequences, were grown overnight at 37 °C on LB agar plates containing ampicillin (50  $\mu$ g/mL). After adding a solution of *P. lunula* luciferin, luciferase activity was detected in the colonies as light emission, using a CCD camera. Phagemids carrying the *P. lunula* luciferase cDNA were excised and further analyzed by restriction mapping and automated sequencing by the ABI dye-terminator method (Applied Biosystems).

**Estimation of the Abundance of Luciferase Copies in the Genome.** The total number of genomic luciferase sequences was estimated as previously described (4). Briefly, known amounts of *P. lunula* genomic DNA and luciferase molecules (*lcfA* cDNA) were blotted on a nylon membrane and hybridized with [ $\alpha$ <sup>32</sup>P]dCTP-labeled *lcfA* cDNA. After exposure to X-ray film, the intensities of hybridization signals of the genomic DNA and luciferase standards were compared by densitometry analysis with the NIH Image v1.54.

**PCR Amplifications and Sequence Analysis.** PCR reactions containing either phenol/chloroform extracted genomic DNA (21) or *P. lunula* luciferase cDNA as template were carried out using standard procedures (9). Specific primers used in this study were P1 (5'-CATCAGCCCCAAGTTCAC3'), P2 (5'-GAGACGTTACCAACCGAT3'), P3 (5'-ATCGGTGGTA-ACGTCTC3'), P4 (5'-TTAAAGCTTGTGGCCAC3'), Paf (5'-CTCAAGTAAGCTGCCAC 3'), PBf (5'-GCCAGCGG-TTGCGACCA3'), PIf (5'-ATAGTGGCCCCCGGTGG3'), PIr (5'-GAGCAAGACCTTAGTGGC3'), PA5r (5'-GTGGC-AGCTTACTTGAG3'), and PB3f (5'-CAAACCAACCCACA-CACTC3').

Amplified fragments were gel purified, cloned into pCR 2.1 vector (Invitrogen), and sequenced. Luciferase sequences

were identified by Blast searches of GenBank. Pairwise analysis was done with the DNA Strider v1.2, and alignments were performed with the ClustalW software. Nucleotide substitutions in aligned sequences were calculated using methods described by Nei & Gojobori (22) as implemented by Korber (23). Phylogenetic tree was constructed based on parsimony analysis, using PAUP v4.0b8.

## RESULTS AND DISCUSSION

***P. lunula* Has at Least Three Highly Conserved Luciferase Genes.** Screening of the cDNA library resulted in the isolation of many positive clones, 22 of which, verified to contain luciferase by the light emission of colonies on the plates following the addition of luciferin, were sequenced and analyzed. On the basis of restriction mapping, comparison of the 3' untranslated regions (UTRs), and PCR analysis, at least three members of a luciferase gene family could be distinguished, and two of them, *lcfA* and *lcfB*, were fully characterized. Partial sequence of a third member, *lcfC*, was obtained by genomic PCR (discussed below).<sup>2</sup>

The structures of *lcfA* and *lcfB* open reading frames (ORFs), but not the UTRs, are similar to the *L. polyedrum* *lcf* gene (8). Pairwise comparison of the *lcfA* and *lcfB* nucleotide sequences shows an identity of 96% in the region encoding the three repeat domains (1131 nt each). The ORF sequence at the 5' end (321 nt in *lcfA* and 324 nt in *lcfB*) is less conserved between the genes (87%), and the 5' and 3' UTR sequences are 72 and 80% identical, respectively. These features can be visualized in the dot-matrix plot (Figure 1A) in which the off-center diagonal lines reflect the similarities among domains. Although luciferase variants have not been found (nor sought) in *L. polyedrum*, a variant *lbp* cDNA clone encoding an LBP isoform (89% identical at the nucleotide level) has been reported (4). As with the *lcfA* and *lcfB* of *P. lunula*, the sequences of the *lbp* genes are more divergent in the UTRs.

The transcript size of *P. lunula* luciferase was reported earlier to be ~4.3 kb (18), slightly larger than that of *L. polyedrum* (4.1 kb). This difference was attributed to a correspondingly larger ORF, since the native luciferase, estimated by Western blots, was ~145 kDa as compared to ~137 kDa for *L. polyedrum*. Our results agree with the transcript size previously reported but not with the molecular mass of the translation product. We found that the difference in transcript size between the two species is due to 3' UTR sequences being about 200 bp longer in *P. lunula*, and not to larger ORFs, which in *P. lunula* predict molecular masses around 137 kDa, the same as for *L. polyedrum*. The previous report of 145 kDa for *P. lunula* luciferase from cell extracts may be due to a posttranslational modification of the native protein, such as phosphorylation.

***lcfA* and *lcfB* Are Organized in Tandem.** The genomic organization of luciferase genes was investigated by a series of PCR reactions using combinations of outwardly directed primers targeted to the UTRs of *lcfA* and *lcfB* and genomic DNA as template. When a sense primer in the 3' UTR of

<sup>2</sup> Nucleotide sequences of *P. lunula* luciferase genes *lcfA*, *lcfB*, *lcfC*, and intergenic region of *lcfA/lcfB* have been deposited in the GenBank Database under accession numbers AF394059, AF394060, AF394061, and AF394062, respectively.

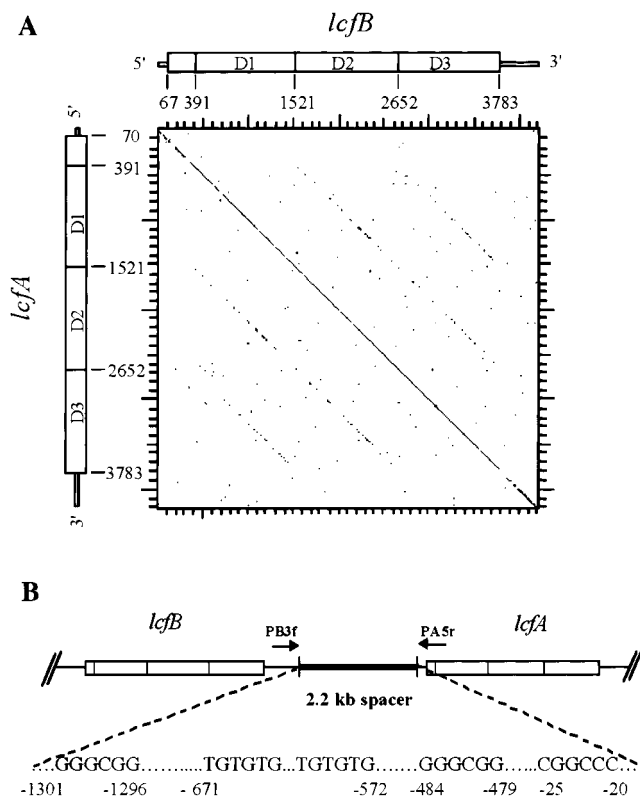


FIGURE 1: (A) Pairwise analysis of the nucleotide sequences of *lcfA* and *lcfB*, illustrated by a dot-matrix plot. The schematic structures of their respective cDNAs are also shown. (B) Diagram showing the tandem organization of *lcfB* and *lcfA* in the *P. lunula* genome. The 2.2 kb spacer region, flanked by the 3' UTR of *lcfB* and 5' UTR of *lcfA*, contains a 100-bp stretch of repeated TG motifs, consensus sequences for the GC box promoter element, and a GC-rich motif conserved in different dinoflagellate promoters. GenBank accession numbers for *lcfA*, *lcfB*, and their intergenic region are AF394059, AF394060, and AF394062, respectively.

*lcfB* (PB3f) was combined with an antisense primer in the 5' UTR of *lcfA* (PA5r), a single 2.6 kb product was amplified. Sequence analysis of this product showed that *lcfB* and *lcfA* are linked in tandem by a 2.25 kb long spacer region (Figure 1B). This does not exclude the possibility that *lcfA* and *lcfB* also occur in the genome as unlinked genes.

The intergenic sequence is larger than that described in a luciferase tandem repeat of *L. polyedrum* (1.46 kb long), and might contain promoter sequences, since the luciferase message in *P. lunula* is not polycistronic (18). As in other dinoflagellate genes, no TATA or CAAT boxes were found. However, a 100 bp stretch rich in TG motifs and the consensus sequence for the GC box promoter element were identified within the intergenic region (Figure 1B). Although shorter in length, a TG motif is also found in the intergenic region of *L. polyedrum lcf* (9), as well as in the 3' UTR of *lbp*, where a potential regulatory protein binds to the transcript in a circadian fashion (24). This might have a regulatory function in relation to the promoter.

The GC box consensus sequence <GGGCGG> functions in both orientations and is usually present in multiple copies around position -110 (25). In *P. lunula*, these same consensus sequences are located further upstream from the putative start site, and whether they are involved in the expression of luciferase in this species is unknown. Furthermore, a GC rich motif <C(G/C)GCCC>, similar to that

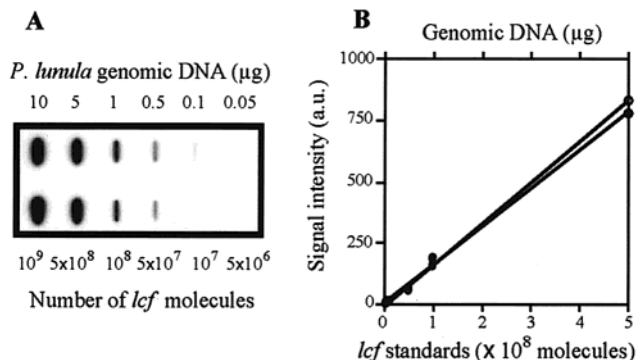


FIGURE 2: Estimation of the abundance of luciferase gene copies in *P. lunula*. (A) Known amounts of genomic DNA and luciferase standard (*lcfA*) were blotted on a nylon membrane and hybridized with [ $\alpha^{32}$ P]dCTP labeled-*lcfA* cDNA. (B) Densitometry of the respective blot indicating that 1  $\mu$ g of *P. lunula* genomic DNA (filled circles) has the same hybridization signal (in arbitrary units) as 10<sup>8</sup> molecules of luciferase standard (open circles).

reported in the ferredoxin gene of the dinoflagellate *P. bipes* (26), is found upstream in *P. lunula lcfA* and *L. polyedrum lcf* and *lbp* genes, the positions of which vary. The consensus sequence <CGTGAACGCAGTG> present in both *L. polyedrum lcf* (9) and *pcp* (27) intergenic regions was not found in *P. lunula*.

**Luciferase Sequences Are Abundant in the Genome and Some Contain an Intron.** Because of their high sequence similarities, it is difficult to determine the copy number of individual members of the luciferase gene family. However, the abundance of total luciferase sequences in *P. lunula* was estimated directly by slot blot hybridizations (Figure 2). Although the genome size of *P. lunula* has not been determined, dinoflagellates are known to have large genomes ranging from 2 to 200 pg/nucleus (1). Within these limits, 200 is the minimum number of copies of luciferase sequences in *P. lunula*, twice the number estimated for *L. polyedrum* (9). This estimate seems reasonable, since the amount of the substrate luciferin has been reported to be 100 times higher in *P. lunula* than in *L. polyedrum* (15).

An intron was detected in a luciferase gene of *P. lunula* by PCR analysis. Two pairs of primers, covering most of the luciferase ORF, were used in PCR reactions containing either genomic DNA or luciferase cDNAs as template. Primers P1, P2, P3, and P4 were designed on the basis of highly conserved sequences so as to hybridize to both *lcfA* and *lcfB*. With primer set P3/P4, amplification products from cDNA clones and genomic DNA had the same length, indicating an absence of introns in the region covered (Figure 3A, lanes 4–6). However, with primers P1/P2, an additional product was obtained from genomic DNA, 400 bp longer than the one from cDNA (lane 3). This genomic PCR product was cloned, and its sequence revealed a single 403 bp intron within the first repeat domain (D1), between bases 684 and 685 (Figure 3B). The intronic sequence has several short palindromic repeats and stop codons in all reading frames. When the intron is removed, the ORF is restored. The fact that the coding sequence of the genomic PCR product containing the intron is slightly different from those of *lcfA* and *lcfB* (Figure 3B) affirms the existence of a third member of the gene family, named *lcfC*. Whether the intron is present in functional copies or in pseudogenes, or both, cannot be specified at this point. However, PCR analysis has demon-



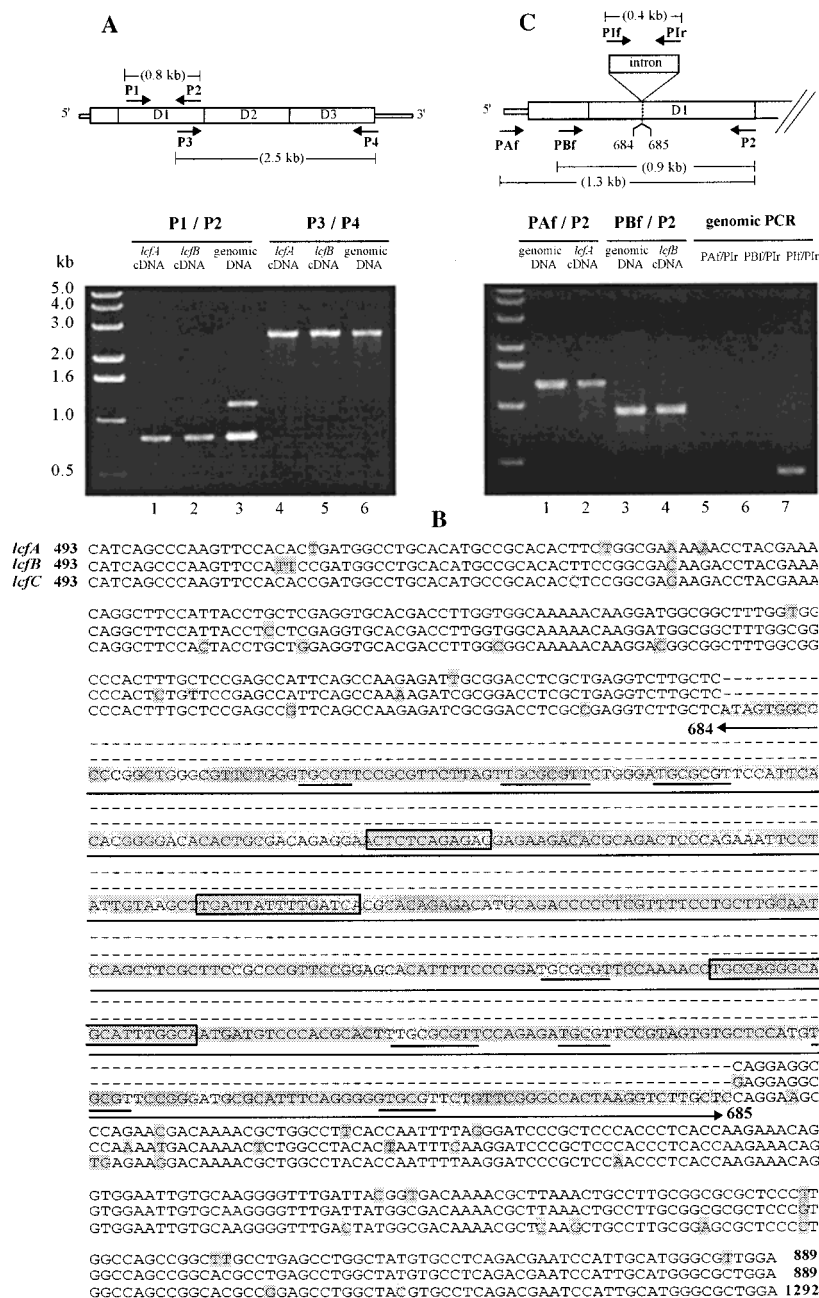


FIGURE 3: An intron within the first repeat domain is present in a *P. lunula* luciferase gene but not in *lcfA* or *lcfB*. (A) PCR analysis with *lcfA* and *lcfB* cDNAs or genomic DNA as template with the primers shown in the scheme. Primers P1/P2 amplified an extra 403 bp-longer fragment that was further cloned and sequenced. (B) Alignment with partial sequences corresponding to D1 of the cloned PCR product (named *lcfC*, AF394061), *lcfA* and *lcfB*; nucleotide differences are shaded. Arrow indicates the intron located between bases 684 and 685 in the cDNA sequences; short palindromic repeats are underlined, and hairpin structures are boxed. (C) PCR analysis using primers specific to the 5' UTR of *lcfA* (PAf) and N-terminus of *lcfB* (PBf). Primers to the 5' and 3' ends of the intron (PIf/PIr) were used in control reaction (lane 7).

strated that *lcfC* has the same three-domain structure as the other members of the *lcf* family (results not shown).

The presence of two specific PCR products from reactions with primer set P1/P2 and genomic DNA suggests that the intron is not present in all luciferase genes. Indeed, PCR amplification using forward primers specific to the 5' UTR of *lcfA* (PAf) and N-terminus of *lcfB* (PBf) and a reverse primer located at the 3' end of the first domain (P2) demonstrated that neither *lcfA* nor *lcfB* has the intron (Figure 3C, lanes 1–4). Genomic PCR reactions with the specific primers PAF or PBF in combination with a primer located at

the 3' end of the intronic sequence (PIr) also did not yield specific products (lanes 5 and 6), confirming this conclusion.

There are only two other reports of introns in dinoflagellate genes. In *Symbiodinium* sp., genes encoding ribulose 1,5-biphosphate carboxylase/oxidase contain two introns (28), and in *Cryptocodinium cohnii* four introns were described in the gene HCc1 encoding a DNA binding protein (29). As in these cases, the intron found within D1 of *P. lunula lcfC* has neither the conserved dinucleotides GT-AG found at the 5' and 3' ends of nuclear or mitochondrial group II introns (30) nor the short consensus sequences responsible for the

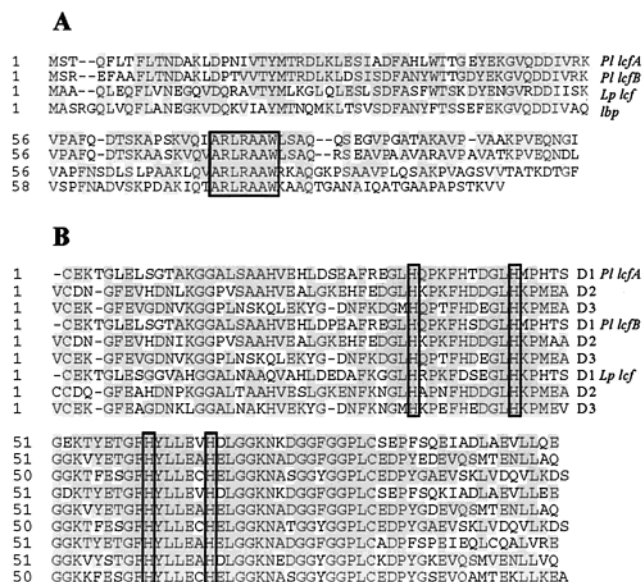


FIGURE 4: (A) Peptide alignment of the N-terminal sequences deduced from *P. lunula* (Pl) *lcfA* and *L. polyedrum* (Lp) *lcf* and *lbp*. The box indicates a stretch of seven residues that is fully conserved in all proteins. (B) Peptide alignment of the first 100 residues of each of the luciferase domains (D1, D2, and D3). The four intramolecularly conserved histidines involved in the pH regulation of *L. polyedrum* luciferase activity (boxed) are conserved in *P. lunula*. Residues matching the consensus sequence are shaded.

secondary structure of group I introns (31), although three short hairpin structures could be identified (Figure 3B, boxed). Also, these dinoflagellate introns, including the one described in this study, do not share common nucleotides at their putative splicing sites.

**Primary Structure of Dinoflagellate Luciferase Is Conserved.** Using *P. lunula lcfA* for comparisons (*lcfB* is similar), the deduced amino acid sequence is 79% identical to the *L. polyedrum* luciferase. The region embracing the three domains, which is most of the molecule, is 83% identical, while the short N-terminal region that precedes the three repeat domains is only about 45% identical between the two species. The function of this latter region remains an enigma, especially because it is homologous to the N-terminus of LBP (45–50%) (Figure 4A).

Although both luciferase and LBP bind luciferin, the N-terminus is not required for luciferase activity (3), and different functions for this conserved region have been postulated: (i) protein–protein interaction between luciferase and LBP, (ii) signal for protein degradation, since both luciferase and LBP are synthesized and destroyed in a circadian fashion in *L. polyedrum*, and (iii) protein interaction with the vacuolar membrane to form scintillons, the bioluminescent organelles (8). The first two suggestions appear to be ruled out, since LBP is absent in *P. lunula*, and its luciferase is not degraded on a daily basis (15, 18).

It is worth noting that a seven amino acid sequence is conserved in the N-termini of *P. lunula* and *L. polyedrum* luciferases and also in LBP (boxed in Figure 4A) and is likely to have functional importance. Exon recombination was suggested as an explanation for the presence of homologous N-termini in these two *L. polyedrum* proteins, even though their genes lack introns (9, 32). The presence of an intron in a luciferase gene of *P. lunula* suggests that this could indeed have been the mechanism.

As in *L. polyedrum*, the three domains of both *P. lunula* luciferases are similar to one another, especially within a 145 amino acid stretch in their central regions, where the intramolecular identity between domains is 84% at the amino acid level, in contrast to 43 and 77% in their N- and C-boundary flanking regions, respectively. The higher conservation in the center of the luciferase domains observed in both species suggests that this region contains the catalytic active sites. This is supported by the fact that, in *P. lunula*, individual luciferase domains, as well as full-length molecules, are enzymatically active when expressed in *E. coli*. In *L. polyedrum*, the corresponding activities were reported to be maximum at pH 6.3, dropping to near zero at pH 8, and four intramolecularly conserved histidines are involved in the loss of activity at high pH (33). These histidine residues are also conserved in *P. lunula* (Figure 4B).

**A Common Origin for the Domains of Dinoflagellate Luciferases.** The occurrence of three intramolecularly homologous domains in both *P. lunula* and *L. polyedrum* luciferases indicates that the configuration has a common origin, but it does not explain how the repeat domain structure evolved. If the domain structure evolved prior to the speciation of *L. polyedrum* and *P. lunula*, the similarity between corresponding luciferase domains of different dinoflagellate species (intermolecular) would be greater than that among the three domains in the same species (intramolecular). Conversely, if the domains originated independently through recent duplication events or by mechanisms such as gene conversion, then the three intramolecular domains should be more similar to each other than to the luciferase domains of other dinoflagellate species.

Intra- and intermolecular pairwise comparisons and phylogenetic analysis of the domain sequences showed that D1 of *P. lunula* luciferases were more similar to D1 of *L. polyedrum* luciferase, at both the amino acid and nucleotide levels, than with D2 or D3 of any of the luciferases (Figure 5). The same pattern was observed in analyses with D2 and D3, supporting the hypothesis that luciferase domains originated through duplication events that occurred prior to the divergence of these dinoflagellate genera.

**Synonymous Substitution Rates Differ in the Luciferase Genes of *P. lunula* and *L. polyedrum*.** A higher intramolecular conservation exists in the center of the luciferase domains of these two dinoflagellate species. This can be visualized by the distribution of nonsynonymous substitutions along the aligned domains, which shows a lower rate of substitutions in the middle region, as well as in the region where the conserved histidines are located (Figure 6A).

However, one highly unusual and altogether unexplained feature of the *L. polyedrum lcf* gene is that the frequency of synonymous (silent) substitutions is also significantly lower in the central region of the domains, as compared to the flanks, and this difference is not due to codon bias (8). In contrast, the frequency of synonymous substitutions in *P. lunula* is uniform along the domains (Figure 6B). All three domains of each of the three luciferases were aligned, and their codons were numbered sequentially on the abscissa. On the ordinate, the cumulative number of synonymous substitutions are given corresponding to the codons. As can be seen, the rate of synonymous substitution for the first 126 codons is the same for all three sequences, but it decreases in the

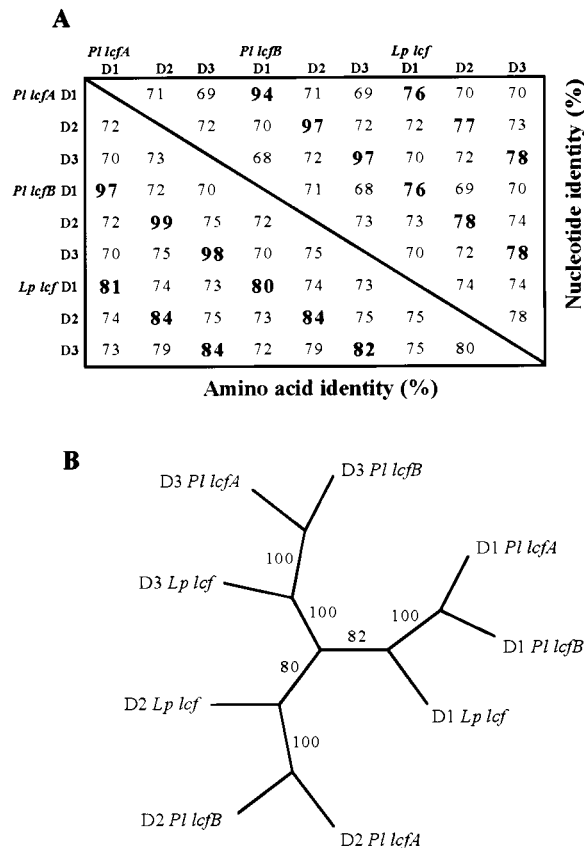


FIGURE 5: Luciferase domain similarities between and within species. (A) Sequence identity based on pairwise alignments at the amino acid and nucleotide levels. Identities between corresponding domains of different molecules (intermolecular) are boldfaced. (B) Molecular phylogeny of dinoflagellate luciferase domains based on amino acid sequence alignments. The 50% majority consensus tree is shown. Numbers indicate bootstrap percentage values (100 replicates) from maximum parsimony analysis using heuristic searches and random addition of sequences. From a total of 377 characters resampled in each replicate, 355 were parsimony-informative, six were parsimony-uninformative, and 18 were invariant.

next 145 codons of *L. polyedrum* (center), increasing somewhat for the final 106 codons.

Given that the luciferase domains of *P. lunula* and *L. polyedrum* evolved from a common ancestor molecule, what is the explanation for the difference in synonymous substitution rates? While the high degree of amino acid sequence conservation in the center of the luciferase domains may be attributed to a conserved catalytic function in both species, this does not explain the conservation at the nucleotide level in *L. polyedrum*. Gene conversion has been suggested as the mechanism responsible for the synonymous substitution constraint in *L. polyedrum*, supported by the occurrence of multiple gene copies and tandem arrangement (8). However, those same features are present in *P. lunula*, making this explanation less plausible. Alternatively, the nucleotide conservation could be associated with a regulatory function. One possibility is that there is a functionally important protein that binds to this region in *L. polyedrum* DNA or RNA, thus accounting for its conserved sequence, and that this does not occur in *P. lunula*. In the former case, this might mediate the circadian control of luminescence, which involves the de novo synthesis and later degradation of the protein each day, regulated at the level of translation. A daily rhythm in

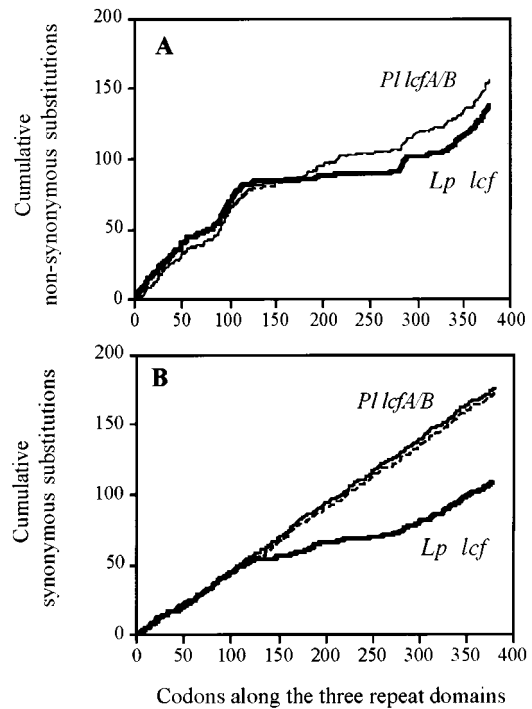


FIGURE 6: Frequency of nonsynonymous (A) and synonymous substitutions (B) along the domains of *P. lunula lcfA* (solid line), *lcfB* (dashed line), and *L. polyedrum lcf* (boldfaced line) genes. The cumulative number of nucleotide substitutions was determined as related to the deduced amino acid sequences of the three repeat domains (D1, D2, and D3) of each luciferase gene. Nonsynonymous substitutions in *lcfA* and *lcfB* superimpose in A.

the amount of LBP is similarly controlled, and in that case a putative regulatory protein has been found that binds to a 27 nt long UG repeat in the 3' UTR of the *lbp* transcript (24). A similar region is not present in the *L. polyedrum* luciferase molecule, but an analogous mechanism might be involved in its regulation. The circadian control of luminescence in *P. lunula* is altogether different in that it does not involve the synthesis and degradation of the luciferase, and thus would not involve such a mechanism. Whether by the proposed or some other mechanism, the strikingly conserved nucleotides in the repeat domains of the *L. polyedrum* may function in the translational control of luciferase synthesis.

In summary, we found that the unique luciferase structure, composed of a short N-terminus followed by three homologous and individually active domains, is conserved in a second dinoflagellate species, suggesting that the structure evolved early in the evolution of this lineage. Intramolecularly conserved histidines responsible for a characteristic pH-activity profile in *L. polyedrum* are also present in *P. lunula* luciferases. Despite the similarities in gene structure and organization, important differences are evident, notably: (i) presence of an intron in one but not all *P. lunula* luciferase genes; (ii) different sizes of untranslated regions in *P. lunula* and *L. polyedrum* luciferase genes, sharing no significant sequence similarities; and (iii) striking differences in the frequency of synonymous substitutions in the center of the repeat luciferase domains between these two species. How these distinct molecular features affect the expression of these dinoflagellate luciferases and whether they are involved in the circadian rhythm of bioluminescence are a matter for future investigation. Such molecular studies may provide



additional insight into gene regulation and evolution in this unusual group of eukaryotes.

## ACKNOWLEDGMENT

We thank Drs. T. Wilson and T. Fagan for critical comments.

## REFERENCES

1. Spector, D. L. (1984) *Dinoflagellates*, Academic Press, Orlando, p 545.
2. Taylor, F. J. R. (1987) in *The Biology of Dinoflagellates* (Taylor, F. J. R., Ed) pp 1–23, Blackwell Scientific Publishers, Boston, MA.
3. Bae, Y. M., and Hastings, J. W. (1994) *Biochim. Biophys. Acta* 1219, 449–456.
4. Lee, D.-H., Mittag, M., Sczekan, S., Morse, D., and Hastings, J. W. (1993) *J. Biol. Chem.* 268, 8842–8850.
5. Morse, D. M., Pappenheimer, A. M., and Hastings, J. W. (1989) *J. Biol. Chem.* 264, 11822–11826.
6. Morse, D., Milos, P. M., Roux, E., and Hastings, J. W. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86, 172–176.
7. Mittag, M., Li, L., and Hastings, J. W. (1997) *Chronobiol. Int.* 15, 93–98.
8. Li, L., Hong, R., and Hastings, J. W. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94, 8954–8958.
9. Li, L., and Hastings, J. W. (1998) *Plant Mol. Biol.* 36, 275–284.
10. Hastings, J. W., and Wood, K. V. (2001) in *Photobiology for the 21st Century* (Valzeno, D. P., and Coohill, T. P., Eds.) pp 199–210, Valdenmar Publishing Co., Overland Park, KS.
11. Hastings, J. W. (1983) *J. Mol. Evol.* 19, 309–321.
12. Hastings, J. W., and Morin, J. G. (1991) in *Neural and Integrative Animal Physiology* (Prosser, C. L., Ed.) pp 131–170, Wiley-Interscience, New York.
13. Nicolas, M.-T., Nicolas, G., Johnson, C. H., Bassot, J.-M., and Hastings, J. W. (1987) *J. Cell Biol.* 105, 723–735.
14. Nicolas, M.-T., Sweeney, B. M., and Hastings, J. W. (1987) *J. Cell Sci.* 87, 189–196.
15. Schmitter, R. E., Njus, D., Sulzman, F. M., Gooch, V. D., and Hastings, J. W. (1976) *J. Cell. Physiol.* 87, 123–134.
16. Njus, D. L. (1975) in *The Control of Bioluminescence in Gonyaulax polyedra*, Ph.D. Thesis, Harvard University, Cambridge, MA.
17. Colepicolo, P., Roenneberg, T., Morse, D. M., Taylor, W. R., and Hastings, J. W. (1993) *J. Phycol.* 29, 173–79.
18. Knaust, R., Urbig, T., Li, L., Taylor, W., and Hastings, J. W. (1998) *J. Phycol.* 34, 167–172.
19. Widder, E. A., and Case, J. F. (1982) *Biol. Bull.* 162, 423–448.
20. Seo, K. S., and Fritz, L. (2000) *J. Phycol.* 36, 351–358.
21. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor.
22. Nei, M., and Gojobori, T. (1986) *Mol. Biol. Evol.* 3, 418–26.
23. Korber, B. (2001) in *Computational and Evolutionary Analysis of HIV Molecular Sequences* (Rodrigo, A. G., and Learn, G. H., Eds.) pp 55–72, Kluwer Academic Publishers, Boston, MA.
24. Mittag, M., Lee, D.-H., and Hastings, J. W. (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91, 5257–5261.
25. Everett, R. D., Baty, D., and Chambon, P. (1983) *Nucleic Acids Res.* 11, 2447–2464.
26. Yoshikawa, T., Takishita, K., Ishida, Y., and Uchida, A. (1997) *Fish. Sci.* 63, 692–700.
27. Le, Q. H., Markovic, P., Hastings, J. W., Jovine, R. V. M., and Morse, D. (1997) *Mol. Gen. Genet.* 255, 595–604.
28. Rowan, R., Whitney, S. M., Fowler, A., and Yellowlees, D. (1996) *Plant Cell* 8, 539–553.
29. Yoshikawa, T., Uchida, A., and Ishida, Y. (1996) *Fish. Sci. (Tokyo)* 62, 204–209.
30. Breathnach, R., Benoist, C., O'hare, K., Gannon, F., and Chambon, P. (1978) *Proc. Natl. Acad. Sci. U.S.A.* 75, 4853–5857.
31. Cech, T. R. (1990) *Annu. Rev. Biochem.* 59, 543–568.
32. Li, L., Hong, R., and Hastings, J. W. (1997) in *Bioluminescence and Chemiluminescence: Molecular Reporting with Photons* (Hastings, J. W., Kricka, L., and Stanley, P., Eds.) pp 74–77, John Wiley and Sons Ltd., Chichester, U.K.
33. Li, L., Liu, L., Hong, R., Robertson, R. L., and Hastings, J. W. (2001) *Biochemistry* 40, 1844–1849.

BI011651Q