Amino acid sequence of the Ca^{2+}-triggered luciferin binding protein of *Renilla reniformis*

Santosh Kumar[1], Maria Harrylock[1], Kenneth A. Walsh[1], Milton J. Cormier[2] and Harry Charbonneau[1]

[1]*Department of Biochemistry, University of Washington, Seattle, WA 98195 and* [2]*Department of Biochemistry, University of Georgia, Athens, GA 30602, USA*

Received 30 May 1990; revised version received 12 June 1990

The complete amino acid sequence of the Ca^{2+}-triggered luciferin binding protein (LBP) of *Renilla reniformis* has been determined. The apoprotein has an unblocked amino terminus and contains 184 residues with a calculated M_r of 20 541. LBP is a member of the EF-hand superfamily of Ca^{2+}-binding proteins and bears three predicted EF-hand domains. The sequence and organization of EF-hand domains are similar to those of the Ca^{2+}-dependent photoprotein, aequorin.

Amino acid sequence; Ca^{2+}-binding protein; Coelenterate bioluminescence

1. INTRODUCTION

The bioluminescent reaction of marine coelenterates is initiated by increases in the intracellular Ca^{2+} concentration [1–3]. In the anthozoan, *Renilla reniformis*, Ca^{2+}-induced bioluminescence involves two distinct proteins; the Ca^{2+}-triggered luciferin binding protein (LBP) and the enzyme, luciferase [1,3]. Luciferase catalyzes the reaction between coelenterate luciferin and O_2 to produce CO_2 and the excited state of oxyluciferin, the emitter. LBP confers Ca^{2+}-sensitivity, since it has at least two high affinity Ca^{2+}-binding sites (K_d = 0.2 μM) and contains one mol of tightly bound luciferin [3]. The luciferin of LBP is capable of reacting with luciferase and O_2 only when Ca^{2+} is bound.

In contrast, the reaction of coelenterates of the class hydrozoa is produced by Ca^{2+}-dependent photoproteins of which aequorin is the most thoroughly studied [2,4–6]. Light emission from photoproteins, such as aequorin, occurs through an intramolecular reaction requiring exogenous Ca^{2+} but neither molecular oxygen nor other cofactors. Aequorin contains equimolar quantities of tightly bound coelenterate luciferin and oxygen. In this report, we present the amino acid sequence of the LBP of *Renilla reniformis* and compare its structural features to those of aequorin and other high affinity Ca^{2+}-binding proteins.

Correspondence address: K.A. Walsh, Department of Biochemistry, SJ-70, University of Washington, Seattle, WA 98195, USA

Abbreviations: LBP, Ca^{2+}-triggered luciferin-binding protein; SCP, sarcoplasmic Ca^{2+}-binding protein

2. MATERIALS AND METHODS

LBP was purified from *Renilla reniformis* using the procedure described by Charbonneau and Cormier [3]. Before cleavage, the protein was reduced and *S*-carboxymethylated with iodoacetic acid as described [7]. S-CM-protein was cleaved at lysines in the presence of 2 M urea [5] using *Achromobacter* protease I (kindly provided by Dr T. Masaki, Ibaraki, Japan). Methionyl residues were cleaved with cyanogen bromide [7] and aspartyl residues with 2% formic acid for 4.5 h at 110°C under vacuum [8]. Peptide K14 was subdigested by cleavage at glutamyl bonds using staphylococcal V8 protease as described [5]. Peptides were purified by reverse-phase HPLC chromatography using linear 0.1% trifluoroacetic acid/acetonitrile gradients generated on a Hewlett-Packard 1090 M chromatograph equipped with a diode array detector. For this purpose, a narrow bore RP-300 (2.1 × 100 mm) column from Brownlee was eluted at a flow rate of 300 μl/min.

Amino acid analyses were performed by separating phenylthiocarbamoyl derivatives using reverse-phase HPLC chromatography (Waters Picotag system) [9]. Automated Edman degradations were carried out on an Applied Biosystems Model 470A gas-phase sequencer equipped with a Model 120A on-line Pth analyzer.

The GENEPRO program (version 4.2, Riverside Scientific, Seattle) was used to identify homologous sequences from the PIR protein database (Release 21.0, National Biomedical Research Foundation) and GenBank (Release 61.0) nucleotide database. Sequence alignments were performed using the ALIGN program [10] from the National Biomedical Research Foundation. Alignment scores are expressed as units of standard deviation as defined by Dayhoff et al. [10].

3. RESULTS

Nearly 90% of the sequence was obtained by analysis of the intact S-CM-protein and peptides generated by cleavage at lysine and methionine. The remainder of the sequence and two key overlaps were provided by analyses of two peptides generated from acid cleavage

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

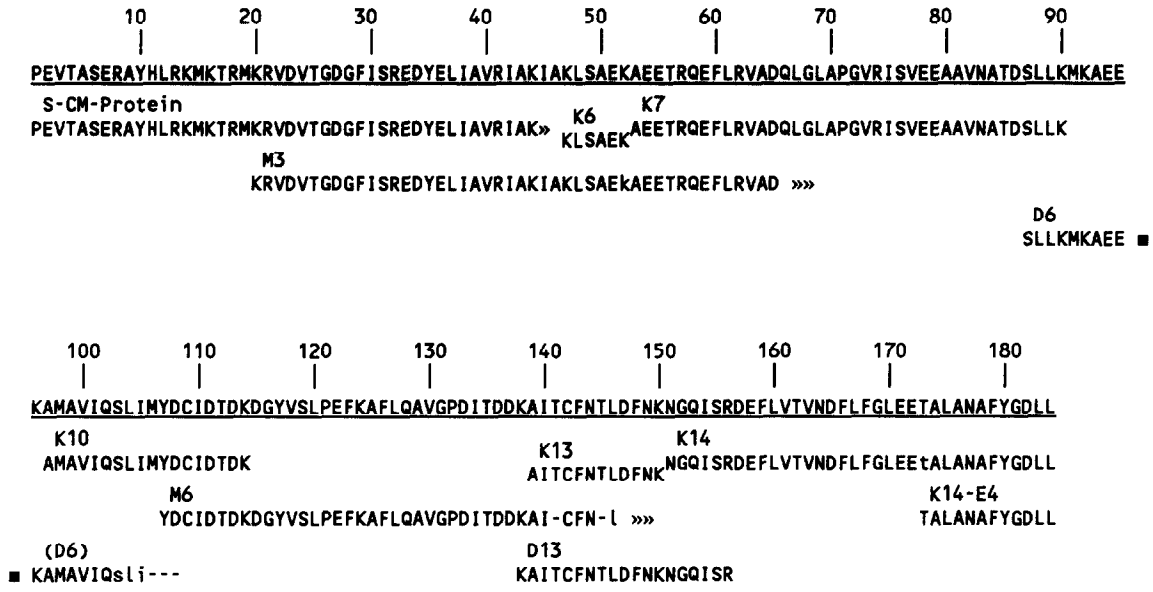

Fig. 1. Summary of the proof of sequence of LBP. The deduced summary sequence is given in single-letter code and is underlined. Results from the Edman degradation of specific peptides are displayed below the summary sequence; the prefixes M, K, D and E denote fragments generated by cleavage at methionyl, lysyl, aspartyl and glutamyl bonds, respectively. Lower case letters designate tentative identifications. Residues not identified are indicated by dashes and by the symbol, » », which indicates that much of the C-terminal sequence of the peptide was not determined.

at aspartyl residues. The results of the Edman degradations and the detailed proof of sequence of the 184-residue LBP are given in Fig. 1.

LBP is not N-terminally blocked and the Edman degradation of the S-CM-protein yielded the unambiguous sequence of residues 1–44. The peptide M3 overlaps the sequence of the intact protein and that of the peptide K7 to provide residues 1–90. The three overlapping peptides, K10, M6 and K13, together establish the sequence of residues 97–150. The gap at residues 91–96 was filled by the analysis of D6, a peptide generated by cleavage at aspartates. D6 clearly overlaps both K7 and K10, proving the sequence of residues 1–150.

Peptide K14 lacked Lys, indicating that it was derived from the C-terminus. Residues 151–184 were identified by the analysis of K14. The tentative identification of Thr-173 was confirmed by the sequence of K14-E4, a subdigestion product generated by cleavage at glutamyl residues. The sequence of D13 provides extensive overlaps between K13 and K14 to establish the complete sequence.

The absence of Glu in peptide K14-E4 (Table I) demonstrates that it is located at the C-terminus of the protein. The amino acid composition of K14-E4 agrees with sequence analysis except for an additional 0.5 and 0.7 mol/mol peptide of Ser and Gly, respectively. The detection of nearly equimolar quantities of these amino acids indicates that the C-terminal sequence could extend two residues beyond Leu. However, Edman degradations of C-terminal peptides, K14 and K14-E4, showed no evidence of additional Gly or Ser residues after Leu-184. Nevertheless, the assignment of the C-terminal residue must be viewed as tentative.

Table I

Amino acid compositions[a]

	LBP	K14-E4
Asx (D/N)	20.4 (22)	2.2 (2)
Glx (E/Q)	20.5 (21)	
Cys (C)	2.0 (2)	
Ser (S)	8.8 (8)	0.5
Gly (G)	11.1 (14)	1.7 (1)
His (H)	1.2 (1)	
Arg (R)	9.8 (10)	
Thr (T)	9.9 (11)	0.9 (1)
Ala (A)	20.0 (21)	2.8 (3)
Pro (P)	3.6 (4)	
Tyr (Y)	4.8 (5)	0.8 (1)
Val (V)	12.8 (13)	
Met (M)	3.2 (5)	
Ile (I)	10.9 (11)	
Leu (L)	18.0 (18)	3.0 (3)
Phe (F)	10.5 (10)	1.0 (1)
Lys (K)	14.7 (13)	
Trp (W)	(0)	
Total	(184)	(12)

[a] Residues/molecule by amino acid analysis or from sequence (in parentheses)

4. DISCUSSION

LBP is a single polypeptide chain of 184 residues with a calculated molecular weight for the apoprotein of 20541. Since it is unblocked, LBP, like aequorin

			1	2	3	4	5	* 6	7	* 8	9	10	11	12
LBP	Domain I	(23-34)	D	V	T	G	D	G	F	I	S	R	E	D
LBP	Domain III	(111-122)	D	T	D	K	D	G	Y	V	S	L	P	E
LBP	Domain IV	(147-158)	D	F	N	K	N	G	Q	I	S	R	D	E
CaM	Domain I	(20-31)	D	K	D	G	D	G	T	I	T	T	K	E
TnC	Domain III	(106-117)	D	K	N	A	D	G	F	I	D	I	E	E

Fig. 2. Predicted Ca^{2+}-binding loops of LBP are aligned with analogous EF-hand loops from two proteins of known crystal structure [16,17], mammalian calmodulin, CaM and chicken troponin C, TnC. The positions of the five oxygen-containing, Ca^{2+}-binding ligands are underlined; asterisks designate the invariant glycine and hydrophobic residues typical of EF-hands.

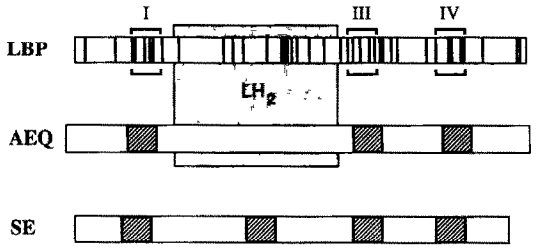

Fig. 3. Structural relationship between LBP and aequorin. The Ca^{2+}-binding loops of EF-hand domains are indicated by cross-hatched boxes for aequorin, AEQ, and by horizontal lines for LBP. The vertical lines of the LBP diagram, denote the position of residues identical to those of aequorin. The putative luciferin-binding sites of LBP and aequorin are designated by the shaded area labeled LH_2. The closely related four domain, Ca^{2+}-binding protein from the bacterium [15] *Streptomyces erythraeus*, SE, is aligned for comparison.

[4–6], may be synthesized as a longer molecule that is clipped at the N-terminus either artifactually during isolation or as part of a physiologically relevant processing step. However, proteolytic processing of LBP seems unlikely since Xaa-Pro bonds are resistant to cleavage. A potential N-linked glycosylation site exists at Asn-83, but no evidence of glycosylation was found in this study or in a previous chemical analysis of LBP [3]. In contrast to aequorin [5,6], structural analysis of LBP revealed no evidence of sequence microheterogeneity. However, ion-exchange chromatography performed during purification resolved at least one minor LBP isoform from the major species used in these studies (Charbonneau and Cormier, unpublished results) suggesting that LBP isoforms may also exist.

LBP is clearly a member of the EF-hand superfamily of Ca^{2+}-binding proteins [11,12]. A search of sequence databases revealed no proteins with significant similarity to LBP other than those of the EF-hand superfamily. Those Ca^{2+}-binding proteins displaying the greatest similarity to LBP are listed below with the percent sequence identity and Dayhoff alignment score [10] given in parentheses: sarcoplasmic Ca^{2+}-binding protein, SCP, from the mollusc [13] *Patinopectin yessoensis* (26%, 7.2 SD); SCP, from the annelid [14] *Perinereis vancaurica tetradentata* (25%, 11.5 SD); Ca^{2+}-binding protein from the bacterium [15], *Streptomyces erythraeus* (31%, 7.5); and aequorin (21%, 7.8). The degree of similarity to proteins of other EF-hand subfamilies (e.g., parvalbumin, calmodulin or troponin C) was much lower and restricted to those sequences comprising the helix-loop-helix structures of the EF-hand motif.

As illustrated in Fig. 2, LBP has three regions that are predicted to be functional Ca^{2+}-binding EF-hand domains, since they meet well-established criteria [18,19], including the presence within the loop structure of an invariant Gly at position 6, a hydrophobic side-chain at position 8, and oxygen-containing side chains at positions 1, 3, 5, 9 and 12. Direct Ca^{2+}-binding measurements have demonstrated the presence of at least two Ca^{2+}-binding sites on LBP [3]. If these data are corrected by using the molecular weight calculated from the sequence reported here, then the stoichiometry becomes 2.7 mol Ca^{2+} per mol protein, indicating that all three EF-hand domains of LBP may be capable of binding Ca^{2+}.

As shown in Fig. 3, the structural organization of LBP closely resembles that of aequorin [4,5]. Several groups [2,4–6] have suggested that the region between Ca^{2+}-binding domains I and III of aequorin may be involved in forming a portion of a luciferin-binding site (Fig. 3). In this model, the putative luciferin-binding region is postulated to have diverged from or replaced what was originally domain II of a four domain progenitor. Since LBP has a nearly identical arrangement of domains, it is likely that its luciferin-binding site is located in an analogous position and that it evolved by similar mechanisms. Even though LBP has no catalytic function, it is somewhat surprising that LBP and aequorin do not display greater similarity in the putative luciferin-binding regions (Fig. 4) because they do bind identical luciferin molecules. Nor is there significant similarity between the sequence predicted from a cDNA for *Renilla* luciferase [20] and those of either LBP or aequorin. Luciferase from the crustacean [21] *Vargula hilgendorfii* displays no significant sequence similarity

```
LBP(41-104): RIAKIAKLSAE--KAEETRQEFLRVAD-QLGLAPGVRISVEEAAVNATDSLLKMKAEEKAMAVIQSL
AEQ(44-110): VINNLGATPEQAKRHKDAVEAFFGGAGMKYGVETDWPAYIEGWKKLATDELEKYAKNEPTLIRIWGD
```

Fig. 4. Alignment of sequences from the putative luciferin binding sites of LBP and aequorin [5]. The region shown here corresponds to the shaded area in Fig. 3 and is taken from the computer-generated alignment (see section 2) of the complete sequences. Sequence identities are designated with vertical bars (—) and hyphens indicate gaps required for optimum alignment. The percent sequence identity for these segments is 19%.

to either LBP or aequorin despite utilizing a luciferin molecule that is a closely related analog of that used by the coelenterates [22]. Thus, it appears that bioluminescent marine organisms have evolved several different strategies for binding coelenterate-type luciferins.

Acknowledgements: This work was supported by a National Institutes of Health grant, GM 15731, to K.A.W. and a grant to M.J.C. from the University of Georgia Research Foundation. This is contribution 654 from the University of Georgia Marine Institute, Sapelo Island, GA 31327.

REFERENCES

[1] Cormier, M.J. (1978) in: Bioluminescence in Action (Herring, P. ed.) pp. 78–108, Academic Press, London.
[2] Cormier, M.J., Prasher, D.C., Longiaru, M. and McCann, R.O. (1989) Photochemistry and Photobiology 49, 509–512.
[3] Charbonneau, H. and Cormier, M.J. (1979) J. Biol. Chem. 254, 769–780.
[4] 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.
[5] Charbonneau, H., Walsh, K.A., McCann, R.O., Prendergast, F.G., Cormier, M.J. and Vanaman, T.C. (1985) Biochemistry 24, 6762–6771.
[6] Prasher, D.C., McCann, R.O., Longiaru, M. and Cormier, M.J. (1987) Biochemistry 26, 1326–1332.
[7] Le Trong, H., Newlands, G.F.J., Miller, H.R.P., Charbonneau, H., Neurath, H. and Woodbury, R.G. (1989) Biochemistry 28, 391–395.
[8] Inglis, A.S. (1983) Methods Enzymol. 91, 324–332.
[9] Bidlingmeyer, B.A., Cohen, S.A. and Tarvin, T.L. (1984) J. Chromatogr. 336, 93–104.
[10] Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) Methods Enzymol. 91, 524–545.
[11] Persechini, A., Moncrief, N.D. and Kretsinger, R.H. (1989) Trends Neurosci. 12, 462–467.
[12] Kretsinger, R.H., Moncrief, N.D., Goodman, M. and Czelusniak, J. (1988) in: The Calcium Channel: Structure, Function, and Implications (Morad, M., Nayler, W., Kazda, S. and Schramm, M. eds) pp. 16–35, Springer, Berlin.
[13] Takagi, T., Kobayashi, T. and Konishi, K. (1984) Biochim. Biophys. Acta 787, 252–257.
[14] Kobayashi, T., Takasaki, Y., Takagi, T. and Konishi, K. (1984) Eur. J. Biochem. 144, 401–408.
[15] Swan, D.G., Hale, R.S., Dhillon, N. and Leadlay, P.F. (1987) Nature 329, 84–85.
[16] Herzberg, O. and James, M.N.G. (1985) Nature 313, 653–659.
[17] Babu, Y.S., Sack, J.S., Greenhough, T.J., Bugg, C.E., Means, A.R. and Cook, W.J. (1985) Nature 315, 37–40.
[18] Gariepy, J. and Hodges, R.S. (1983) FEBS Lett. 160, 1–6.
[19] Kretsinger, R.H. (1980) CRC Crit. Rev. Biochem. 8, 119–174.
[20] Lorenz, W. and Cormier, M.J. (1990) manuscript in preparation.
[21] Thompson, E.M., Nagata, S. and Tsuji, F.I. (1989) Proc. Natl. Acad. Sci. USA 86, 6567–6571.
[22] Hart, R.C. and Cormier, M.J. (1979) Photochemistry and Photobiology 29, 209–215.