Genomic Structure of the Amphioxus Calcium Vector Protein[1]

Hajime J. Yuasa,* Jos A. Cox,† and Takashi Takagi*,[2]

*Biological Institute, Graduate School of Science, Tohoku University, Sendai 980-8578; and †Department of Biochemistry, University of Geneva, CH-1211 Geneva 4, Switzerland

Calcium vector protein (CaVP) is an EF-hand Ca²⁺-binding protein, which is unique to the protochordate, amphioxus. CaVP is supposed to act as a Ca²⁺ signal transductor, but its exact function remains unknown. Not only its function but also its exact evolutionary relationship to other Ca²⁺-binding proteins is unclear. To investigate the evolution of CaVP, we have determined the complete sequences of *Ca VP* cDNAs from two amphioxus species, *Branchiostoma lanceolatum* and *B. floridae*, whose open reading frame cDNA and amino acid sequences show 96.5 and 98.2% identity, respectively. We have also elucidated the structure of the gene of *B. floridae Ca VP*, which is made up of seven exons and six introns. The positions of four of the six introns (introns 1, 2, 3, and 5) are identical with those of calmodulin, troponin C, and the Spec protein of the sea urchin. These latter proteins belong to the so-called troponin C superfamily (TnC superfamily) and thus CaVP likely also belongs to this family. Intron 6 is positioned in the 3′ noncoding region and is unique to *CaVP*, so it may represent a landmark of the CaVP lineage only. The position of intron 4 is not conserved in the genes of the TnC superfamily or CaVP, and seems to result from either intron sliding or the addition of an intron (randomly inserted into or close to domain III) to the genes of the TnC superfamily during their evolution.

Key words: amphioxus, calcium vector protein, cDNA sequences, EF-hand.

The EF-hand is a Ca²⁺-binding motif normally composed of 29 amino acids with a helix-loop-helix structure (*1*). Present-day proteins possess two to eight EF-hands and are classified as belonging to one protein family, called collectively the EF-hand family. The family comprises 39 subfamilies (*2*), and four of them, calmodulin (CaM), troponin C (TnC), myosin essential light chain (ELC), and myosin regulatory light chain (RLC), called the CTER subfamily, are evolutionarily closely related according to the results of analyses at both the amino acid (*3*) and cDNA levels (*4*). In fact, such a close relationship has been suggested from the earliest reports (*5, 6*), and the CTER together with parvalbumin (*7*) and Spec (*8–10*), a Ca²⁺-binding protein isolated from the sea urchin, have often been called the "TnC superfamily". Except for parvalbumin, they possess four EF-hand domains per molecule and may have evolved from a common four-domain ancestor. It is thought that parvalbumin lost the first domain during its evolution.

Calcium vector protein (CaVP) is a unique protein found in various tissues of amphioxus and is abundantly expressed in muscle (*11*). It forms a stable complex with a protein of 26 kDa, CaVP-target protein (CaVPT, *12–14*), and it is supposed to act as a Ca²⁺ signal transductor, although its exact function remains unknown. The evolution of CaVP is also uncertain. Similar to the members of the TnC superfamily, CaVP possesses four EF-hand domains, but binds only two Ca²⁺ per molecule. EF-hands I and II do not bind Ca²⁺, and in the native protein, a unique disulfide bridge is formed between these two domains. Kobayashi *et al.* (*11*) stressed that CaVP shows about 35% sequence identity with CaM and TnC. At the same time, however, they also suggested the possibility that CaVP is a chimera with a C-terminal half homologous to that of CaM and TnC, and an N-terminal half resembling those of parvalbumins and sarcoplasmic Ca²⁺ binding proteins. On the other hand, CaVP clusters with ELC on the dendrogram constructed by the maximum parsimony method (*15*), and on analysis of the individual EF-hand domains, those of the CaVP were found to cluster near those of S100, the small EF-hand protein which possess two domains per molecule (*3*).

In this study, we determined the complete cDNA sequences of *CaVPs* from two species, *Branchiostoma lanceolatum* and *B. floridae*, and elucidated the exon/intron organization of the latter and discussed the evolutionary relationship of CaVP with the TnC superfamily.

MATERIALS AND METHODS

Cloning of B. lanceolatum Ca VP cDNA—Total RNA of adult *B. lanceolatum* was prepared according to the acid guanidium thiocyanate method (*16*), and mRNA was

Abbreviations: CaM, calmodulin; CaVP, calcium vector protein; ELC, myosin essential light chain; nt, nucleotide; RLC, myosin regulatory light chain; TnC, troponin C.

purified with Oligotex dT-30 Super (Japan Roche). The single-stranded cDNA was synthesized using a First-Strand cDNA Synthesis Kit (Pharmacia). The cDNA of *B. lanceolatum CaVP* was amplified by PCR using Ex *Taq* DNA polymerase (Takara). The redundant oligomer used for PCR was 5'-GA(AG)GA(AG)AA(AG)GA(CT)GA(AG)-TG(CT)ATGAA-3', which was designed based on the amino acid sequence, EEKDECMK (residues 12-19), of *B. lanceolatum* CaVP (*11*). The oligo-dT adaptor, 5'-GGGATCC-GAATTCT$_{17}$-3', was used as another primer. The 5' upstream of cDNA was determined as follows. The *Eco*RI-ended double-stranded cDNA was synthesized from mRNA using a TimeSaver cDNA Synthesis Kit (Pharmacia), and the *Eco*RI Cassette (Takara) was ligated at each end of cDNA. The 5' upstream region was amplified by PCR using the cassette-specific primer C1, 5'-GTACATATTGTCGT-TAGAACGCG-3', and 5'-CGAAGACAACACAGACTTT-ATTAC-3' (complementary to the sequence from nt 1057 to 1080).

Cloning of B. floridae CaVP cDNA—The λ ZAP II cDNA library of embryonic *B. floridae* was kindly provided by Dr. L.Z. Holland (Scripps Institute of Oceanography, University of California, San Diego). The *B. floridae CaVP* cDNA was also amplified by PCR using the cDNA library as a template. The primers used for 3'-half amplification were 5'-TGAGGGCGTTCAAGGTCTTC-3' (corresponding to the sequence from nt 275 to 294 of *B. lanceolatum CaVP* cDNA) and the T7 primer, 5'-TAATACGACTCACTATA-GGG-3', and those for 5'-half amplification were 5'-GAAT-CCTGTGGACTAACTTA-3' (complementary to the sequence from nt 487 to 506 of *B. lanceolatum CaVP* cDNA) and the T3 primer, 5'-ATTAACCCTCACTAAAGGGA-3'.

Screening of a Genomic Library—The λ FIX II genomic library of *B. floridae* was also provided by Dr. L.Z. Holland, and was screened using the open reading frame of *B. lanceolatum CaVP* cDNA as a probe, because the identity between these two species, *CaVP* cDNAs is very high (96.5%, see below). The probe was labeled with DIG-DNA Labeling Mixture (Boehringer Mannheim) by PCR using the following forward primer, 5'-CTCTACCAGTCTTTC-AGTCC-3' (corresponding to the sequence from nt −29 to −10) and the reverse primer, 5'-GAATCCTGTGGACTA-ACTTA-3' (complementary to the sequence from nt 487 to 506). Hybridization and washing were carried out according to the manufacturer's instructions (Boehringer Mann-heim), and positive clones were detected with a DIG Luminescent Detection Kit (Boehringer Mannheim).

DNA Sequencing—All PCR-amplified products and iso-lated genomic clones were subcloned to the pCR II plasmid vector (TA-cloning kit, Invitrogen) or pUC18 for sequenc-ing. The nucleotide sequences were determined by the dideoxy chain termination method with a Dye Primer Cycle Sequencing Kit (Applied Biosystems) using an automated DNA sequencer (Applied Biosystems 373A).

RESULTS AND DISCUSSION

CaVP cDNAs from Two Amphioxus Species—The cDNA of *B. lanceolatum CaVP* was amplified by PCR and the complete sequence of 1,124 nucleotides was established from two overlapping fragments (Fig. 1). The open reading frame is composed of 489 nucleotides and encodes a protein of 162 amino acid residues including the initial Met.

Compared to the previously reported amino acid sequence of *B. lanceolatum* CaVP (*11*), the sequence of positions 152 to 155 is Lys-Lys-Ser-Lys instead of Lys-Ser-Lys-Lys. This region is Lys-rich and therefore the precise sequence is difficult to deduce from the results of peptide analysis. This and the fact that the same sequence is also observed in *B. floridae* CaVP (Fig. 1) led us to conclude that the previous report probably contained a peptide sequence error.

The cDNA of *B. floridae CaVP* is composed of 1,080 nucleotides and the open reading frame is 489 nucleotides long, encoding a protein of 162 amino acid residues includ-ing the initial Met, *i.e.* it is of the same length as in *B. lanceolatum* (Fig. 1). Within the coding regions, there are 17 nucleotide substitutions (96.5% identical) reflecting three amino acid differences (98.1% identical) between the CaVPs of the two species. Two of the three substitutions occur in domain II and the other at the C-terminus. Thus, there is likely no difference in Ca^{2+} binding between the two CaVPs, *i.e.* both bind two Ca^{2+} per molecule at EF-hand domains III and IV.

The identity between the two species' CaVPs is as high as those of CaMs, TnCs and skeletal actins, which are known to be highly conserved proteins (see Table I). However, mitochondrial proteins, such as cytochrome *b* and NADH dehydrogenase, of the two amphioxus species show equally high identity, as compared to the lower level of identity between the human and mouse mitochondrial proteins. Therefore, the high similarity of the two species' CaVPs may not reflect the conservatism of CaVP, but may rather be due to relatively recent divergence of *B. lanceolatum* and *B. floridae*.

Genomic Structure of the CaVP Gene—Using *B. lan-ceolatum CaVP* cDNA as a probe, we initially isolated three positive clones (clones 1, 2, and 3) from the genomic library of *B. floridae*. As clones 1 and 2 showed similar digestion patterns with several restriction enzymes, only clones 1 and 3 were used for further investigation. The nucleotide sequence of clone 1 was determined: it comprises 6,729 bp and spans the 5'-half of the *CaVP* gene up to Ile-90 (Fig. 2A). In this region, the *CaVP* gene consists of four exons divided by three introns (introns 1 to 3), and the nucleotide sequences of the exons exactly match those of the cDNA. All the introns start with GT and end with AG, and, according to the nomenclature[3] of Kretsinger and Naka-yama (*17*), the introns positions are −14/0, 1.01/1, and 2.13/1, respectively.

Clone 3 comprises 4,613 bp and spans the 3'-half of the *CaVP* gene starting from Ile-105 (Fig. 2A). In this region, the gene consists of three exons divided by two introns (introns 5 and 6); the position of the former intron is 4.21/1, and the last intron is inserted within the 3'-untranslated

---

[3] The positions of introns are indicated according to the nomenclature of Kretsinger and Nakayama (*17*). The first number corresponds to the EF-hand domain numbered sequentially from N to C. The second number (following the period) represents the residue number of the intron insertion in the EF-hand domain, which as a rule consists of 29 residues. The last number (following the slash) is the phase: 0 means that the intron lies between triplet codons, 1 between first and second nucleotides of the codon, and 2 between the second and third ones. For instance, 4.21/1 means insertion in domain IV, 21st residue and phase 1. −14/0 means phase 0, 14 residues before the beginning of domain I. 3+01/1 means phase 1, 1 residue beyond domain III, within the region between domain III and IV.
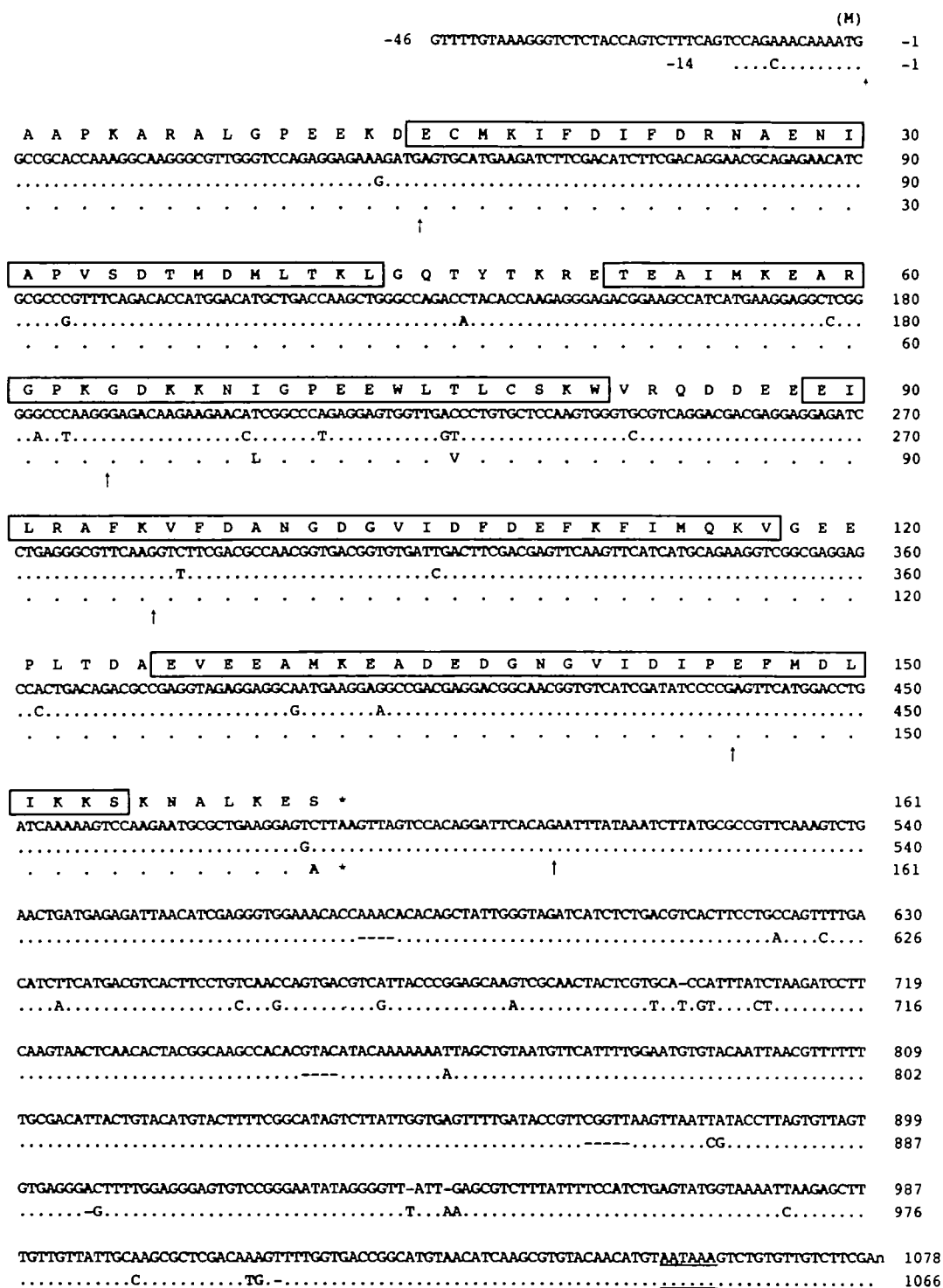
```
                                                                    (M)
                          -46   GTTTTGTAAAGGGTCTCTACCAGTCTTTCAGTCCAGAAACAAAATG   -1
                                -14      ....C.........              -1
                                                                           ↑

A  A  P  K  A  R  A  L  G  P  E  E  K  D│E  C  M  K  I  F  D  I  F  D  R  N  A  E  N  I│     30
GCCGCACCAAAGGCAAGGGCGTTGGGTCCAGAGGAGAAAGATGAGTGCATGAAGATCTTCGACATCTTCGACAGGAACGCAGAGAACATC   90
...........................................G.............................................   90
.  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .     30
                                            ↑

│A  P  V  S  D  T  M  D  M  L  T  K  L│G  Q  T  Y  T  K  R  E│T  E  A  I  M  K  E  A  R│     60
GCGCCCGTTTCAGACACCATGGACATGCTGACCAAGCTGGGCCAGACCTACACCAAGAGGGAGACGGAAGCCATCATGAAGGAGGCTCGG  180
.....G................................................A....................................C...  180
.  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .     60

│G  P  K  G  D  K  K  N  I  G  P  E  E  W  L  T  L  C  S  K  W│V  R  Q  D  D  E  E│E  I│    90
GGGCCCAAGGGAGACAAGAAGAACATCGGCCCAGAGGAGTGGTTGACCCTGTGCTCCAAGTGGGTGCGTCAGGACGACGAGGAGGAGATC  270
..A..T...................C.......T...........GT.....................C......................  270
.  .  .  .  .  .  .  .  L  .  .  .  .  .  V  .  .  .  .  .  .  .  .  .  .  .  .  .  .     90
                        ↑

│L  R  A  F  K  V  F  D  A  N  G  D  G  V  I  D  F  D  E  F  K  F  I  M  Q  K  V│G  E  E   120
CTGAGGGCGTTCAAGGTCTTCGACGCCAACGGTGACGGTGTGATTGACTTCGACGAGTTCAAGTTCATCATGCAGAAGGTCGGCGAGGAG  360
.................T.......................C................................................  360
.  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .   120
                  ↑

P  L  T  D  A│E  V  E  E  A  M  K  E  A  D  E  D  G  N  G  V  I  D  I  P  E  F  M  D  L│   150
CCACTGACAGACGCCGAGGTAGAGGAGGCAATGAAGGAGGCCGACGAGGACGGCAACGGTGTCATCGATATCCCCGAGTTCATGGACCTG  450
..C..................................G.......A............................................  450
.  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .   150
                                                                        ↑

│I  K  K  S│K  N  A  L  K  E  S  *                                                         161
ATCAAAAAGTCCAAGAATGCGCTGAAGGAGTCTTAAGTTAGTCCACAGGATTCACAGAATTTATAAATCTTATGCGCCGTTCAAAGTCTG  540
.....................................G....................................................  540
.  .  .  .  .  .  .  .  .  .  A  *              ↑                                           161

AACTGATGAGAGATTAACATCGAGGGTGGAAACACCAAACACACAGCTATTGGGTAGATCATCTCTGACGTCACTTCCTGCCAGTTTTGA   630
.............................................----.................................A....C....   626

CATCTTCATGACGTCACTTCCTGTCAACCAGTGACGTCATTACCCGGAGCAAGTCGCAACTACTCGTGCA-CCATTTATCTAAGATCCTT   719
....A.................C...G.........G.............A................T..T.GT....CT..........   716

CAAGTAACTCAACACTACGGCAAGCCACACGTACATACAAAAAAATTAGCTGTAATGTTCATTTTGGAATGTGTACAATTAACGTTTTTT   809
...........................----..........A...............................................   802

TGCGACATTACTGTACATGTACTTTTCGGCATAGTCTTATTGGTGAGTTTTGATACCGTTCGGTTAAGTTAATTATACCTTAGTGTTAGT   899
.......................................................-----........CG...............   887

GTGAGGGACTTTTGGAGGGAGTGTCCGGGAATATAGGGGTT-ATT-GAGCGTCTTTATTTTCCATCTGAGTATGGTAAAATTAAGAGCTT   987
........-G.........................................T...AA.................................C........   976

TGTTGTTATTGCAAGCGCTCGACAAAGTTTTGGTGACCGGCATGTAACATCAAGCGTGTACAACATGTAATAAAGTCTGTGTTGTCTTCGAn  1078
............C..........TG.-...........................................................   1066
```

Fig. 1. **Comparison of the cDNA and deduced amino acid sequences of CaVPs of two species of amphioxus.** Upper lanes: *B. lanceolatum* CaVP cDNA and deduced amino acid sequence; lower lanes: *B. floridae* CaVP cDNA and deduced amino acid sequence. Identical nucleotides and amino acids to those in *B. lanceolatum* CaVP are indicated by dots (.), and gaps are inserted for maximal similarity and shown by bars (-). The typical polyadenylation signals (AATAAA) are underlined, and the four EF-hand Ca^{2+} binding domains are boxed. The initiator methionine (residue -1, parenthesized) likely is removed after translation. The upward arrows (↑) indicate the positions of introns in the *B. floridae* CaVP genes.

region. Both introns are delineated in conformity to the GT/AG rule. Compared to the cDNA sequence, the genomic nucleotide sequence exhibits some differences: in the open reading frame the GAT codon encoding Asp-143 changes to GAC in the genome and the 3'-noncoding region contains one substitution and three nucleotide insertions

(data not shown).

Since there is no overlapping region between clones 1 and 3, the region encoding from Leu-91 to Val-104 can not be adequately described. Therefore we performed further screening to obtain another clone containing the 91/104 region and succeeded in isolating clone 4. Clone 4 contains all the *CaVP* gene except exon 1 (see Fig. 2A). Of this clone only the exons and regions flanking the splice junctions were sequenced basically. Within the Leu-91/Val-104

TABLE I. **The identity (%) of several molecules between two species of amphioxus, and between human and mouse.**

	B. floridae vs. B. lanceolatum		Human vs. Mouse	
	cDNA (ORF)	Amino acid	cDNA (ORF)	Amino acid
CaVP	96.5	98.2		
CaM	97.1	100.0	91.8	100.0
TnC	95.4	97.6	90.3	98.8
Skeletal actin	95.5	97.1	90.8	100.0
Cytochrome b	97.7	99.2	73.5	76.7
NADH 5[a]	98.2	99.3	65.2	63.0

[a]NADH 5, NADH dehydrogenase subunit 5. The accession numbers (EMBL/DDBJ/GenBank) of the sequences used are; Y09863 (*B. floridae* CaM), Y09880 (*B. lanceolatum* CaM), M19311 (human CaM *II*), M27844 (mouse CaM *II*), D88977 (*B. floridae* TnC), D88976 (*B. lanceolatum* TnC), X07898 (human fast skeletal TnC), M57590 (mouse fast skeletal TnC), D87407 (*B. floridae* skeletal actin), Y13662 (*B. lanceolatum* skeletal actin), M20543 (human skeletal alpha-actin), M12347 (mouse skeletal alpha-actin), AF-035173 (*B. floridae* cytochrome b), AF035172 (*B. floridae* NADH dehydrogenase subunit 5), Y16474 (*B. lanceolatum* mitochondrial DNA), D38112 (human mitochondrial DNA), and J01420 (mouse mitochondrial DNA).
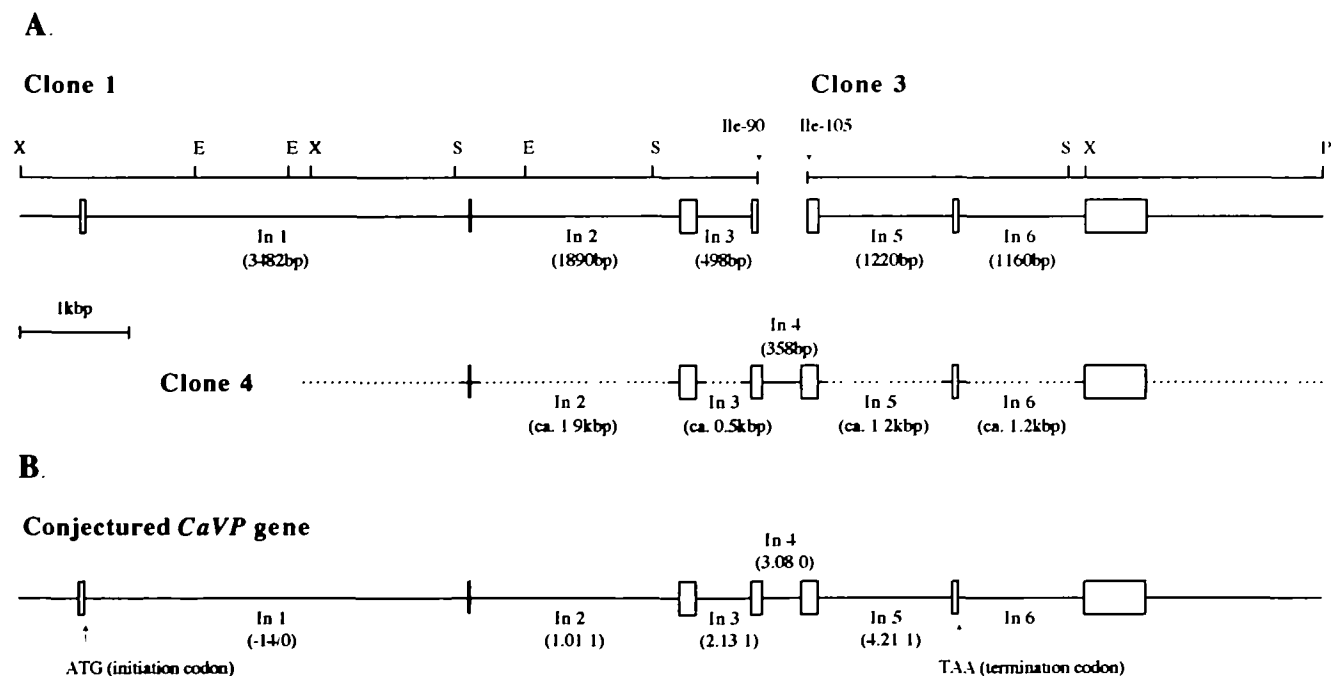
region an intron (intron 4) was observed at position 3.08/0, whereas the positions of the other introns are identical to those in clones 1 and 3. The fourth intron was identified in clone 4 only, but in the *B. lanceolatum CaVP* gene an intron located at exactly the same position, as determined on sequencing of a partially amplified genomic PCR fragment (data not shown). To summarize, the conjectural organization of the *B. floridae CaVP* gene is shown in Fig. 2B, and it may consist of 7 exons and 6 introns (*ca.* 11.5 kbp).

When the intronic sequences of clones 1, 3, and 4 are compared, some substitutions and insertions/deletions are observed: the identity between each pair of splice junction flanking regions is *ca.* 90% on average. Moreover, in the exons four substitutions were observed, all of them being silent substitutions (data not shown). Thus, clone 4 may be derived from a gene that is different from in the case of clones 1 and 3. It is possible that clones 1 and 3 also come from distinct genes. There are two possible explanations for the observed discrepancies: (i) amphioxus *CaVP* may be multi-copy gene, or (ii) the different clones may be derived from distinct animal specimens, since the genomic library was not constructed from a single amphioxus specimen (*18*). Although the three clones are derived from different genes, it is supposed that the positions of the introns are identical in all genes.

Comparison of Intron Positions among TnC Superfamily Genes—The distribution of introns in the genes of *CaVP* and members of the TnC superfamily is shown in Table II. In all genes the first intron is basically inserted just after the initiation codon, ATG (shown as M in Table II). Furthermore, the positions of the second, third, and fifth

A.

B.

Fig. 2. **Structural organization of the B. floridae CaVP genomic clones.** A: Restriction and exon/intron maps of clones 1 and 3. *Eco*RI (E), *Pst*I (P), *Sal*I (S), and *Xba*I (X) restriction sites are indicated. Exons are shown by boxes, the 5'- and 3'-flanking regions and introns (In) are shown by bars, and the length of each intron is indicated. As for clone 4, the regions whose sequences were not determined are shown by dotted lines, and the length of each intron was estimated by PCR. B: The conjectural organization of the gene of *B. floridae CaVP*. It may be composed of 7 exons and 6 introns, and the position of each intron is indicated in parenthesis. The positions of the initiation codon (ATG) and the termination codon (TAA) are also indicated.

TABLE II.   The intron positions of TnC superfamily genes (based largely on Refs. *17* and *20*).

Subfamily species (type)	Positions of introns					
CaVP						
Amphioxus	−14/0 (M)	1.01/1	2.13/1	3.08/0	4.21/1	3′-n.r.[a]
TnC						
Human (f)	−17/0 (M)	1.01/1	2.13/1	3.11/2	4.21/1	
Human (s/c)	−10/0	1.01/1	2.13/1	3.11/2	4.21/1	
Amphioxus	−17/0 (M)	1.01/1	2.13/1	3.24/0	4.21/1	
Ascidian	−10/0 (M)	1.01/1	2.13/1	3.24/0	4.21/1	
CaM						
Human (III)	−10/0 (M)	1.01/1	2.13/1	3.12/0	4.21/1	
Rat (I,III)	−10/0 (M)	1.01/1	2.13/1	3.12/0	4.21/1	
Rat (II)	−10/0 (M)	1.01/1	−	3.12/0	4.21/1	
Chicken	−10/0 (M)	1.01/1	2.13/1	3.12/0	4.21/1	
Aplysia	−10/0 (M)	1.01/1	2.13/1	−	4.21/1	
Drosophila	−10/0 (M)	−	2.13/1	−	4.21/1	
Ascidian	−10/0 (M)	1.01/1	2.04/0	3.12/0	4.21/1	
Spec						
Sea urchin	−12/0 (M)	1.01/1	2.13/1	3.18/2	4.21/1	
ELC						
Human	−09/0 (M)	1.01/1	2.12/1	3+01/0	4.21/1	
Rat (L1)	−09/0 (M)	1.01/1	2.12/1	3+01/0	4.21/1	
Rat (L4)	−08/0 (M)	1.01/1	2.12/1	3+01/0	4.21/1	
Chicken (L1)	−09/0 (M)	1.01/1	2.12/1	3+01/0	4.21/1	
Chicken (L3)	−08/0 (M)	1.01/1	2.12/1	3+01/0	4.21/1	
Parvalbumin						
Human	−	−	2.11/1	3.23/2	4.21/1	
Rat	−	−	2.11/1	3.23/2	4.21/1	
RLC						
Rat	−27/0 (M)	1.05/0	1+01/1	2+03/1	3.21/2	4.02/0

(M) shows the intron is inserted just after the initiation codon, ATG. (−) shows the absence of intron. [a]The last intron of *CaVP* is inserted within the 3′-noncoding region.

introns (1.01/1, 2.13/1, and 4.21/1, respectively) of *CaVP* are identical with those of TnCs, CaMs, and Spec. Thus, the gene structure of amphioxus *CaVP* is typical of the TnC superfamily and *CaVP* is clearly a member of the TnC superfamily. Within the family CaVP seems more closely related to the following three members: TnCs, CaMs, and Spec (Table II). The sixth intron, which is observed in the 3′-noncoding region, is unique to *CaVP*, and may be a characteristic feature, *i.e.* a landmark of the *CaVP* lineage.

Except for in RLC,[a] the exon/intron organizations are well conserved among all members of the TnC superfamily (including *CaVP*). Basically they all possess five introns with the following characteristics: intron 1 is inserted just after the initiation codon (phase 0), and in a few exceptions, it is also inserted before domain I and the phase 0 mode is conserved; intron 2 is inserted at position 1.01/1 in all members without any exception; intron 3 is inserted at position 2.13/1 in *CaM*, *TnC*, *Spec*, and *CaVP*, whereas in the other members the insertion point is slightly different, but phase 1 is conserved (with the notable exception of intron 3 in the *CaMs* of the ascidian, *Halocynthia roretzi* with 2.04/0, Ref. *19*); finally intron 5 is inserted at position 4.21/1 in all members without any exception.

Only the positions of intron 4 is not conserved at all in the TnC superfamily. Two explanations can account for this diversity: either intron 4 acquired different positions in the TnC superfamily genes through intron sliding, or the members of the TnC superfamily may have gained a fourth

intron independently during their evolution. According to the first hypothesis the fourth introns of the TnC super-family are derived from a common ancestral intron, but its sequence changed so rapidly that the homology quickly disappeared. Indeed these introns are different in length and sequence; there is no positive proof that they are homologous. In general the evidence of intron sliding itself is weak (*20*), and besides, except in one instance, between protochordate/vertebrate *TnCs* (*21*), no sliding has occurred within each lineage of subfamilies. The second hypothesis implies that the ancestor gene possessed only four introns and the present-day members gained a fourth intron independently during their evolution. Surprisingly, this intron 4 was gained within or near domain III for all genes. Among the members of the TnC superfamily, the most conservative molecule is CaM, thus the common ancestor of the family is supposed to have been a CaM-like protein or, perhaps, CaM itself. Interestingly, in inverte-brate (*Drosophila* and *Aplysia*) *CaM* genes (*19, 22, 23*) the fourth intron is absent, and this may reflect the absence of the fourth intron in the ancestor gene. The discovery of new members of the TnC superfamily may explain the evolu-tionary rules that govern the position of intron 4.

In this study, we elucidated the gene structure of a new Ca^{2+}-binding protein and show that, although structurally and functionally it is a unique protein, it belongs to the TnC superfamily. The gene structure may be used as a tool to assign a protein to a given superfamily. This gene structure sheds more light on the evolution of constant and variable elements in the genes of the TnC superfamily.

---

[a] The positions of introns in the RLC are different from those of the other members of the TnC superfamily, but for conciseness this particular diversity is not analyzed in this study.

phy, University of California San Diego, for providing us with the cDNA and genomic library of *B. floridae* larvae.

REFERENCES

1. Kretsinger, R.H. and Nockolds, C.E. (1973) Carp muscle calcium-binding protein. II. Structure determination and general description. *J. Biol. Chem.* 248, 3313-3326

2. Kawasaki, H. and Kretsinger, R.H. (1995) Calcium-binding proteins 1: EF-hands. *Protein Profile* 2, 305-490

3. Nakayama, S., Moncrief, N.F., and Kretsinger, R.H. (1992) Evolution of EF-hand calcium-modulated proteins. II. Domains of several subfamilies have diverse evolutionary histories. *J. Mol. Evol.* 34, 416-448

4. Nakayama, S. and Kretsinger, R.H. (1993) Evolution of EF-hand calcium-modulated proteins. III. Exon sequences confirm most dendrograms based on protein sequences: calmodulin dendrograms show significant lack of parallelism. *J. Mol. Evol.* 36, 458-476

5. Collins, J.H. (1974) Homology of myosin light chains, troponin-C and parvalbumins deduced from comparison of their amino acid sequences. *Biochem. Biophys. Res. Commun.* 58, 301-308

6. Weeds, A.G. and McLachlan, A.D. (1974) Structural homology of myosin alkali light chains, troponin C and carp calcium binding protein. *Nature* 252, 646-649

7. Dayhoff, M.O. (1978) *Atlas of Protein Sequence and Structure*, Vol. 5, supplement 3, pp. 273-283, National Biomedical Foundation, Washington, D.C.

8. Carpenter, C.D., Bruskin, A.M., Hardin, P.E., Keast, M.J., Anstrom, J., Tyner, A.L., Brandhorst, B.P., and Klein, W.H. (1984) Novel proteins belonging to the troponin C superfamily are encoded by a set of mRNA in sea urchin embryos. *Cell* 36, 663-671

9. Hardin, S.H., Carpenter, C.D., Hardin, P.E., Bruskin, A.M., and Klein, W.H. (1985) Structure of the gene encoding a major calcium-binding protein in the embryonic ectoderm of the sea urchin, *Strongylocentrotus purpuratus*. *J. Mol. Biol.* 186, 243-255

10. Hardin, P.E., Angerer, L.M., Hardin S.H., Angerer, R.C., and Klein, W.H. (1988) Spec2 genes of *Strongylocentrotus purpuratus*. Structure and differential expression in embryonic aboral ectoderm cells. *J. Mol. Biol.* 202, 417-431

11. Kobayashi, T., Takagi, T., Konishi, K., and Cox, J.A. (1987) The primary structure of a new M_r 18000 calcium vector protein from amphioxus. *J. Biol. Chem.* 262, 2613-2623

12. Takagi, T. and Cox, J.A. (1990) Primary structure of the target of calcium vector protein of amphioxus. *J. Biol. Chem.* 265, 19721-19727

13. Petrova, T.V., Comte, M., Takagi, T., and Cox, J.A. (1995) Thermodynamic and molecular properties of the interaction between amphioxus calcium vector protein and its 26 kDa target. *Biochemistry* 34, 312-318

14. Petrova, T.V., Takagi, T., and Cox, J.A. (1996) Phosphorylation of the IQ domain regulates the interaction between Ca^{2+}-vector protein and its target in amphioxus. *J. Biol. Chem.* 271, 26646-26652

15. Moncrief, N.D., Kretsinger, R.H., and Goodman, M. (1990) Evolution of EF-hand calcium-binding proteins. I. Relationships based on amino acid sequences. *J. Mol. Evol.* 30, 522-562

16. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156-159

17. Kretsinger, R.H. and Nakayama, S. (1993) Evolution of EF-hand calcium-modulated proteins. IV. Exon shuffling did not determine the domain compositions of EF-hand proteins. *J. Mol. Evol.* 36, 477-488

18. Holland, P.W.H., Holland, L.Z., Williams, N.A., and Holland, N.D. (1992) An amphioxus homeobox gene: sequence conservation, spatial expression during development and insights into vertebrate evolution. *Development* 116, 653-661

19. Yuasa, H.J., Yamamoto, H., and Takagi, T. (1999) The structural organization of the ascidian, *Halocynthia roretzi*, calmodulin genes. The vicissitude of introns during the evolution of calmodulin genes. *Gene* 229, 163-169

20. Stoltzfus, A., Logsdon, J.M., Jr., Palmer, J.D., and Doolittle, W.F. (1997) Intron "sliding" and the diversity of intron positions. *Proc. Natl. Acad. Sci. USA* 94, 10739-10744

21. Yuasa, H.J., Sato, S., Yamamoto, H., and Takagi, T. (1997) Structure of the ascidian, *Halocynthia roretzi*, troponin C gene. *J. Biochem.* 121, 671-676

22. Smith, V.L., Doyle, K.E., Maune, J.F., Munjaal, R.P., and Beckingham, K. (1987) Structure and sequence of the *Drosophila melanogaster* calmodulin gene. *J. Mol. Biol.* 196, 471-485

23. Swanson, M.E., Sturner, S.F., and Schwartz, J.H. (1990) Structure and expression of the *Aplysia calfornica* calmodulin gene. *J. Mol. Biol.* 216, 545-553