Identification and characterization of integrin-binding sialoprotein (IBSP) genes in reptile and amphibian

Author(s)
Shintani, S; Kamakura, N; Kobata, M; Toyosawa, S; Onishi, T; Sato, A; Kawasaki, K; Weiss, KM; Ooshima, T

Journal
Gene, 424(1-2): 11-17

URL
http://hdl.handle.net/10130/820
Identification and characterization of integrin-binding sialoprotein (IBSP) genes in reptile and amphibian

Key words: Xenopus laevis; African clawed toad; caiman; bone sialoprotein; evolution; cDNA sequence; genomic organization

Seikou Shintani*, a, Naofumi Kamakura a, Mitsuhiko Kobata a, Satoru Toyosawa b, Tomoyuki Onishi a, Akie Sato c, Kazuhiko Kawasaki d, Kenneth M. Weiss d, Takashi Ooshima a

Departments of a Pediatric Dentistry and b Oral Pathology, Osaka University Graduate School of Dentistry, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan

Tsurumi University School of Dental Medicine, 2-1-3 Tsurumi, Tsurumi-ku, Yokohama 230-8501, Japan

Department of Anthropology, Pennsylvania State University, University Park, Pennsylvania 16802, USA

*Corresponding author. Department of Pediatric Dentistry, Osaka University Graduate School of Dentistry, 1-8 Yamadaoka, Suita, Osaka 565-0871, Japan. tel: +81-6-6879-2963; fax: +81-6-6879-2965. E-mail address: shintani@dent.osaka-u.ac.jp

The nucleotide sequences of the caiman and African clawed toad IBSP genes have been deposited in GenBank under accession nos. EU007686, EU007687, and EU007688.

Abbreviations: IBSP, integrin-binding sialoprotein; SPP1, secreted phosphoprotein; DMP1, dentin matrix protein 1; DSPP, dentin sialophosphoprotein; MEPE, matrix extracellular phosphoglycoprotein; PCR, polymerase chain reaction; RT-PCR, reverse transcription PCR; GAPDH, glycerol 3-phosphate dehydrogenase; poly-A, polyadenylation; mRNA, messenger RNA; Arg, arginine; Asn, asparagine; Asp, asparaginic acid; Glu, glutamic acid; Gly, glycine; Ser, serine; Thr, threonine; Tyr, tyrosine; poly-E, stretch of poly-glutamic acid.
Abstract

Integrin-binding sialoprotein (IBSP) is a member of the small integrin-binding ligand N-linked glycoprotein (SIBLING) family; and the whole SIBLING family is further included in a larger secretory calcium-binding phosphoprotein (SCPP) family. SIBLING proteins are known to construct a part of the non-collagenous extracellular matrices of calcified tissues, and considered to have arisen by duplication and subsequent divergent evolution of a single ancient gene. To understand the alterations of SIBLING molecules associated with the evolution of calcified tissues in vertebrates, we initiated a search for lower vertebrate orthologs of SIBLING genes. In the present study, an IBSP ortholog from a reptile (caiman) and two distinct orthologs from an amphibian (African clawed toad) were identified and characterized. As expected, the toad IBSP genes were transcribed only in calcified tissue (jaw and tibia), as also seen in mammals. The caiman, toad, avian, and mammalian IBSPs share several unique features specific for IBSP and apparently have similar properties. Furthermore, analysis of the sequences suggested that the IBSP molecule might have gradually intensified its functions related to calcification during its evolutionary process through tetrapods.
1. Introduction

Mineralized tissues in living vertebrates are mainly composed of calcium phosphates in the form of carbonate hydroxyapatite (HAp) and various types of collagen. Non-collagenous proteins (NCPs) are also dispersed in calcified tissues as extracellular matrices and exert great regulatory influence on calcified tissue formations. One category of NCPs has been termed as the small integrin-binding ligand N-linked glycoprotein (SIBLING) family, based on their common characteristics (Fisher et al., 2001). This family originally consisted of integrin-binding sialoprotein (IBSP, bone sialoprotein), secreted phosphoprotein 1 (SPP1, osteopontin), dentin matrix protein 1 (DMP1), dentin sialophosphoprotein (DSPP), and matrix extracellular phosphoglycoprotein (MEPE), which have been shown to be closely linked on human, mouse, rat (Huq et al., 2005), and chicken (Kawasaki and Weiss, 2006) chromosomes. Thus, it is considered that this entire family might be the result of duplication and subsequent divergent evolution of a single ancient gene (Fisher et al., 2001), and may be crucial to the evolution of vertebrate tissue mineralization (Kawasaki and Weiss, 2006). Furthermore, it has been suggested that SIBLING family proteins have diverged from the secretory calcium-binding phosphoprotein (SCPP) family by gene duplication as well as enamel matrix proteins, milk casein, and salivary proteins (Kawasaki and Weiss, 2006; Kawasaki et al., 2007).

As part of a systematic effort to understand the alteration of SIBLING molecules associated with the evolution of calcified tissues in vertebrates, we initiated a search for lower vertebrate orthologs of each SIBLING gene. To date, the DSPP and MEPE
genes have been only in mammalian genomes (Kawasaki and Weiss, 2006; Kawasaki et al., 2007), however, IBSP, SPP1, and DMP1 have functional similarities (Fisher et al., 2001), with IBSP and SPP1 especially noted to share some structural and characteristic features (Alford and Hankenson, 2006). These three proteins have been cloned from chickens (Ganss et al., 1999; Sodek et al., 2000; Toyosawa et al., 2000), while the DMP1 gene has been cloned from a crocodile (Toyosawa et al., 1999). In this study, we attempted to identify IBSP orthologs in lower tetrapods, since it is relatively specific for calcified tissues (Ganss et al., 1999), whereas the SPP1 gene is widely expressed in soft tissues such as epithelial cells, macrophages, and lymphocytes (Sodek et al., 2000). Herein, we present our findings of IBSP-encoding cDNAs isolated from a caiman and an African clawed toad, used as reptile and amphibian subjects, respectively.
2. Materials and Methods

2.1. Sources and isolation of genomic DNA

All of the experiments were reviewed and approved by the Osaka University Graduate School of Dentistry Intramural Animal Use and Care Committee prior to beginning the study. An African clawed toad (*Xenopus laevis*) was killed under anesthesia, after which tissue samples (upper jaw, tibia, heart, and liver) were excised and immediately frozen in liquid nitrogen at -70°C until use. Genomic DNA was isolated from muscle tissues from the thighs using a DNeasy Tissue Kit (Qiagen, Hilden, Germany). The cDNA libraries of smooth-fronted caimans (*Paleosuchus palpebrosus*) and African clawed toads constructed from total RNAs of their jaws, which have been described previously (Toyosawa et al., 1998), were used.

2.2. Isolation of RNA and cDNA synthesis

Frozen tissue samples were ground to a fine powder, from which total RNA was extracted using TRIZOL Reagent (Invitrogen, Carlsbad, CA, USA). First-strand cDNA was synthesized with oligo(dT) primer using SuperScript III RNaseH- reverse transcriptase (Invitrogen).

2.3. Isolation of IBSP cDNA clones

The caiman IBSP sequence was amplified from the caiman cDNA library by polymerase chain reaction (PCR) using the degenerate primer IBSPDG1 (antisense: 5’-CCTTGTAGTAGCTGTACTCATCYTCRTA-3’), based on the conserved regions
of IBSP genes, in combination with vector anchor primers to obtain the 5’-end part of the clone. A full-length clone was obtained by PCR performed with the caiman IBSP-specific primer IBSPC1 (sense: 5’-GGGTTATGGGTATGAAGAAAGTCCTA-3’) and vector anchor primers. As for the toad, IBSPDG1 and vector anchor primers were also used to amplify the 5’-end part of one of two IBSP sequences (toad-A), while the 3’-end part was amplified by a combination of IBSPX5 (sense: 5’-AAACGGAAAGCATAAGTGTAGAGTA-3’), a toad-A-specific primer, and vector anchor primers. The partial coding sequence of another IBSP transcript (toad-B) was screened by PCR using the toad-A-specific primers IBSPX2 (antisense: 5’-CTCATCCTCATACCGTGCATAATT-3’) and IBSPX15 (sense: 5’-ACAAGGCATAAATGGAAAAAGCAGAT-3’), then a full-length clone was obtained with the use of the toad-B-specific primers IBSPX19 (sense: 5’-GAGACTACACAAGCAAGCACAACCTG-3’) and IBSPX20 (antisense: 5’-CAGTTACCCACCACCATTGATT-3’), in combination with vector anchor primers.

2.4. PCR amplification

PCR was performed as described previously (Shintani et al., 2007). To examine the mRNA expression of the toad IBSP genes, cDNA synthesized from the toad tissues was amplified with the toad IBSP-specific primers IBSPX17 (sense: 5’-CTCTTTTTGGTGTTTTGAGCAGTT-3’) and IBSPX14 (antisense: 5’-TTCTGGAGCTTCTTTGTTCTTTTCTTCT-3’) for toad-A, and IBSPX22 (sense:
5’-TCTTTTGTTGGTTTGAGCAGC-3’) and IBSPX28 (antisense: 5’-ACCTCATCTGAGCTATCTTCTTCTGAA-3’) for toad-B. The glycerol 3-phosphate dehydrogenase (GAPDH) gene was amplified by GAPDH1 and GAPDH2 as an internal positive control (Shintani et al., 2002, 2003).

2.5. Cloning and sequencing

Cloning and sequencing were carried out as described previously (Shintani et al., 2002). To determine the organization of the toad IBSP genes, primers were designed to amplify regions of the individual exon-intron boundaries (Table 1) and the genomic sequence data were compared with the cDNA sequences to identify the boundaries.

2.6. Southern blot analysis

For Southern blotting, 10 µg of genomic DNA was digested overnight with 100 units of the restriction enzymes BamHI, BglII, and EcoRV (NEW ENGLAND Biolabs, Beverly, MA). For the caiman, we used genomic DNA prepared previously (Shintani et al, 2002) from a spectacled caiman (Caiman crocodilus apaporiensis), which is a near relation of Paleosuchus palpebrosus, as caimans are rarely obtainable in Japan due to an animal protection law put into effect in 2005. The recovered DNA was loaded onto a 0.8% agarose gel, transferred to a Hybond-N⁺ nylon membrane (Amersham Biosciences, Buckinghamshire, UK), and then crosslinked to the filter by UV radiation. A 339-bp DNA fragment within the exon 7 sequence of the caiman IBSP and a 346-bp DNA fragment within the exon 7 sequence of toad-A were amplified.
using the caiman \textit{IBSP} gene-specific primer pair IBSPC3 (sense: 5’-GGCACTAGCACCAACTACTGAAAG-3’) and IBSPC4 (antisense: 5’-ATAATCTGTTGCTTTCATATCCGGT-3’), and the toad \textit{toad-A}-specific primer pair IBSPX2 and IBSPX30 (sense: 5’-TGAAACAGAGACAGAATGACGAAT-3’), respectively. Next, 100 ng of each DNA sample was labeled with digoxigenin (DIG) independently using a PCR DIG Probe Synthesis Kit (Roche Diagnostics GmbH, Basel, Switzerland) and the filters were hybridized with the probes. After being washed and blocked using a DIG Wash and Block Buffer Set (Roche Diagnostics GmbH), the bands were immunodetected with anti-DIG Fab fragments conjugated to alkaline phosphatase and visualized by chemiluminescence with a DIG Luminescent Detection Kit (Roche Diagnostics GmbH), and then exposed to XAR5 film (Kodak, Rochester, NY).

2.7. \textit{Data analysis}

The nucleotide and predicted protein sequences were aligned using SeqPup computer software (http://iubio.bio.indiana.edu/soft/molbio) and GENETYX-MAC version 13 genetic information processing software (SOFTWARE DEVELOPMENT Co., Tokyo, Japan). Evaluation of sequence similarities, drawing of phylogenetic trees, and other analyses were also performed using GENETYX-MAC.
3. Results and Discussion

3.1. Characteristics of caiman and toad IBSP cDNA sequences

The coding sequence of the caiman IBSP cDNA (GenBank accession no. EU007686) was deduced from two separate overlapping cDNA clones; the first of which was amplified by anchored PCR with IBSPDG1 from the caiman cDNA library and the second with IBSPC1 (Fig. 1). The caiman IBSP sequence was 1680 bp long and contained an ORF for 337 amino acid residues beginning with the putative translation start site at 69 bp and a 16-residue hydrophobic signal peptide sequence at sites 69-116. The stop codon at 1080 bp was followed by a 3’-untranslated region (UTR) of 598 bp with three putative polyadenylation (poly-A) signals (AATAAA) at sites 1319, 1473, and 1646 bp. To investigate the possibility that other IBSP-like loci may exist in the caiman, Southern blot analysis was carried out using a 339-bp long caiman cDNA fragment. The blot obtained following digestion with BamHI, BglII, and EcoRV restriction endonucleases revealed a single hybridizing band (Fig. 2A). This result indicated that there is likely only a single IBSP locus in the caiman genome.

As for the toad, two IBSP genes (toad-A and toad-B) could be amplified from the toad cDNA library. The complete coding sequence of toad-A cDNA (GenBank accession no. EU007687) was deduced from three separate overlapping cDNA clones; one of which was amplified by anchored PCR with IBSPDG1 and the other with IBSPX5 (Supplementary Figure S1). The partial fragment of the toad-B transcript (GenBank accession no. EU007688) was captured by PCR with IBSPX15 and IBSPX2, and the others were amplified by anchored PCR with IBSPX19 and
IBSPX20 (Supplementary Figure S2). The two clones differed by a number of substitutions, as well as some insertions and deletions. The result of Southern blot analysis showed that the blot obtained following digestion with BamHI, BglII, and EcoRV restriction endonucleases had two distinct hybridizing bands in all three blots (Fig. 2B). Thus, it could be established that the clones were derived from different genes, though the genes were 82% identical. The *toad*-A sequence was 1435 bp long and contained an ORF for 264 amino acid residues beginning with the putative translation start site at 15 bp and a 16-residue hydrophobic signal peptide sequence at sites 15-62. The stop codon at 807 bp was followed by a 3’UTR of 626 bp with two putative poly-A signals (ATTAAA and AATAAA) at the 1443-bp and 1823-bp sites. The longest clone of *toad*-B was found to be 931 bp long and contain an ORF for 248 amino acid residues. However, neither the translation start site (ATG) nor the putative poly-A signal was contained in the cDNA sequence, because it lacked both the 5’-end and the 3’-end parts. The stop codon was located at 745 bp and followed by a 3’UTR of 184 bp.

The *IBSP* gene was first isolated in rat tissue, after which orthologs have been subsequently cloned from several mammals, such as humans, pigs, cattle, mice, chickens, and hamsters (Ganss et al., 1999; Sasaguri et al., 2000). The sequence similarities of the caiman IBSP to the mammalian and avian proteins ranged from 47-44%, while those of the toad IBSPs ranged from 27-32%. The caiman IBSP, and toad-A and toad-B amino acid sequences, excluding the signal peptides, were found to be acidic proteins like other reported IBSPs (pI 3.96, 4.08, and 4.07) (Fisher et al.,
with calculated molecular weights of 35.8, 28.2, and 26.3 kDa, respectively. As in the case of mammalian and avian molecules, they were rich in Glu and Gly, which constituted 33-36% of the total amino acids (Ganss et al., 1999), while other amino acid percentages were also comparable with previously reported IBSP sequences. The hydrophobicity plots (Kyte and Doolittle, 1982) of the terapod sequences were similar and shared the following characteristics: the hydrophobic signal peptide of 16 amino acid residues was followed by internal hydrophilic segments divided by several short hydrophobic segments, and the C-terminal region was hydrophilic.

3.2. Tissue expression of toad IBSP genes

To determine the pattern of expression of the toad IBSP genes, cDNA obtained by reverse transcription from toad tissues (jaw, tibia, heart, and liver) were amplified by PCR with toad IBSP gene-specific primers (Fig. 3). The expected sizes of the amplification products with the primer pairs IBSPX17 and IBSPX14 for toad-A, and IBSPX22 and IBSPX28 for toad-B were 342 bp and 191 bp, respectively. The expected products obtained using toad IBSP gene-specific primers could only be determined in the jaw and tibia specimens, as no products were detected from the heart and liver. We concluded that the toad IBSP genes were expressed only in calcified tissues, as in the case of mammalian and avian IBSP genes (Ganss et al., 1999).

3.3. Phylogenetic relationship

A phylogenetic tree based on the amino acid sequences of the tetrapod IBSP
proteins is shown in Figure 4. The relationships among the sequences were estimated from the taxonomical positions of the species from which they were derived, giving distinct clades for mammals, as well as two branches containing the caiman and the avian sequences independently, and a clade of two toad sequences that may represent polyploidyzation. The toad-B sequence had a slightly longer terminal branch length than the toad-A sequence, showing that toad-B may be tending toward pseudogene status.

3.4. Genomic organization of toad IBSP genes

The products obtained by PCR amplification of toad genomic DNA with the primers shown in Table 1 were partially sequenced, and then the boundaries between the exons and introns were determined by comparisons with the cDNA sequences. There were no polymorphisms in the coding regions of the toad IBSP genes between the cDNA and genomic DNA sequences. As with the human IBSP gene (Kerr et al., 1993), the toad IBSP genes have 6 encoding exons (exon 2-7) separated by 5 introns (Fig. 5). All of the determined intron phases were type 0, with all donor sites containing ‘GT’ and all acceptor sites containing ‘AG’ consensus splice sequences (Table 2). The border between exon 6 and 7 might shift to the upstream region by reorganizations of the exon-intron boundaries in the amniote lineage after the divergence from amphibians (Fig. 5, 6). Other toad exons have lengths similar to their mammalian and avian counterparts, with the exception of the first encoding exon of the avian gene. These findings suggest that the IBSP gene usually consists of 6
encoding exons and the avian gene is unusual. The first encoding exon of the avian gene is considered to be a united exon derived from exons 2 and 3 (Yang and Gerstenfeld, 1997), because the genome size of birds was reduced due to the requirement of a high metabolic rate for flight (Kawasaki and Weiss, 2006).

3.5. Consideration of IBSP evolution

Results of the caiman and toad IBSPs showed that tetrapod IBSPs share several unique features. The most characteristic structures of IBSP molecules are the stretches of several Glu residues (poly-Es), with poly-E believed to be involved in binding Ca$^{2+}$ ions in the dimensions of HAp. The middle regions of mammalian IBSPs possess two (the first and second) poly-Es and the sauropsid molecules have an extra (the third), whereas a single poly-E exists in the toad IBSPs. Poly-Es have 8-12 contiguous Glu residues in at least one of two Poly-Es (mammals) and the extra (sauropsids). The poly-Es of toad-A and toad-B sequences were each shown to possess only 6 or 7 contiguous Glu residues. It was previously indicated that a sequence of at least 8 contiguous Glu residues was required for active nucleation of HAp by IBSP (Tye et al., 2003). Hence, the poly-Es of both toad IBSP proteins did not meet the requirement. The unique feature of poly-E originated in the repetitive DNA sequences GAG and GAA. These repeats, which are sometimes found in unrelated genes of other species, were probably derived from a microsatellite (Shintani et al., 2000a, 2000b). It seems reasonable to propose that IBSP molecules lengthened their contiguous Glu chains of poly-E through incorporation of microsatellites and a mistake of DNA replication.
during the evolutionary process. The extra poly-E (third) might have arisen in the common ancestor of sauropsid lineages after divergence from the amphibian lineage. In addition, the phosphoserines combined with poly-E, which were present at the conserved position 185 (Fig. 6), presumably play a role in the interaction with Ca\(^{2+}\) ions and HAp (Saad et al., 2005). Viewed in this light, the IBSP molecule might have gradually intensified its functions related to calcification by extending the contiguous Glu chains of a poly-E and growing the number of poly-Es itself during its evolutionary process through tetrapods.

In bone, it appears that IBSP contributes to remodeling, the formation of the initial HAp crystal, cell attachment by osteoclasts and osteoblasts (Ganss et al., 1999; Rapuano et al., 2004), and the pathological effects on bony tissues (Ganss et al., 1999; Alford and Hankenson, 2006). Also, the conserved unique Arg-Gly-Asp (RGD) motif at position 348-350 (Fig. 6) plays an important role in those processes via its reaction with the cell-surface integrin. The caiman and toad molecules also contained motifs in consistent regions, except for toad-B, in which Arg was replaced by Lys.

Posttranslational modifications are believed to be necessary for such functions as crystal growth and attachment activity (Ganss et al., 1999; Qin et al., 2004). IBSP is a highly glycosylated and sulfated phosphoprotein, and an N-linked glycosylation site (Asn at position 213 in Fig. 6) with the consensus motif ‘Asn-X-Thr (NXT)’ and four phosphorylation sites for casein kinase II (Ser at position 68, 69, 78 and 185 in Fig. 6) with the motif ‘Ser-X\(_{1-2}\)-Glu (SXE or SXXE)’ were found to be highly conserved among tetrapod IBSPs (‘X’ denotes any residue). Contrary to our expectations, the
SXE sequence position 32, which has been widely conserved in the N-terminus of SCPP molecules, disappeared from both toad IBSPs. The site might have been lost in the toad, since even fish SCPP genes have the motif (Kawasaki and Weiss, 2006).

On the other hand, tetrapod IBSPs have a number of Thr residues for O-linked glycosylation sites (Zaia et al., 2001: Salih et al., 2004) in the segments encoded by exon 7 (Thr-rich region in Fig. 6). Glycosylated oligosaccharides may sometimes be associated with sulfation. In addition, it is evident that sulfation also occurs against Tyr residues, which are rich in the flanking region of the RGD sequence in mammalian IBSPs (Ganss et al., 1999; Zaia et al., 2001). Sulfation and high glycosylation also confer mucin-like properties to the IBSP molecule (Ganss et al., 1999). The caiman and toad molecules also contain Tyr-rich segments in the analogous position, and there are 7 highly conserved Tyr residues at positions 352, 355, 359, 361, 367, 370, and 367 (Fig. 6).

Taken together, it can be seen that tetrapod IBSPs share several common functional features, with the exception of poly-E, which have been maintained in their respective lineages.

3.6. Conclusion

The IBSP molecule might have gradually intensified its functions related to calcification during its evolutionary process through tetrapods, while other features have been maintained in their respective lineages over the course of about 350 million years of evolution.
References


characterization. Genomics 17, 408-415.


318, 125-136.


**FIGURE LEGENDS**

**Fig. 1**  Nucleotide and translated amino acid sequences of caiman *IBSP* cDNA clone. Amino acid residues are shown with the IUPAC-IUB single letter code. The signal peptide is underlined with a dotted line. The stop codon is indicated by an asterisk (*) and polyadenylation signals are underlined with a solid line. Primer positions and orientations are indicated by double lines with arrowheads above their boxed locations in the sequence. The sequences can be accessed in the GenBank database under accession No. **EU007686**.

**Fig. 2**  Southern blot analysis of the *IBSP* genes. Genomic DNA was digested with *Bam*HI (Bam), *Bgl*II (Bgl), and *Eco*RV (Eco) restriction enzymes, then hybridized with the *IBSP* cDNA probe.

(A) Caiman IBSP gene; genomic DNA derived from a spectacled caiman (*Caiman crocodilus apaporiensis*), which is a near relation of *Paleosuchus palpebrosus*, was used.

(B) Toad IBSP genes.

**Fig. 3**  RT-PCR analysis of toad tissues.

(A) cDNA was amplified from jaw, tibia, heart, and liver specimens using the *IBSP* cDNA-specific primer sets IBSPX17 and IBSPX14 for *toad-A* (Lane *a*), and IBSPX22 and IBSPX28 for *toad-B* (lane *b*).
(B) cDNA was amplified with primers GAPDH1 and GAPDH2, and used as a positive control.

**Fig. 4** Neighbor-joining tree of amino acid sequences of IBSP proteins. The scale bar represents the distance based on the proportion of amino acid differences. Gapped sites were removed from all sequences before distance estimates were made. Numbers on the nodes indicate the percentage recovery of that node in 1000 bootstrap replications.

**Fig. 5** Structural organization of human, avian, and two toad IBSP genes. Exons are numbered according to the human IBSP gene. Coding lengths of exons with the number of amino acids are shown beneath the boxes. Open and shaded boxes indicate untranslated and translated exon regions, respectively. Exon 1 of the avian IBSP is illustrated by a box with dotted lines, because it was determined to be absent. The amino acid numbers of exon 1 of the toad genes are expressed as question marks, because it is unclear whether they are present. The letters ‘Glu’ and ‘Thr’ enclosed in the open boxes with arrows show the locations of the poly-Es and Thr-rich regions, respectively.

**Fig. 6** Amino acid sequence alignment of IBSP genes from human, bovine, rat, mouse, chicken, caiman, and toad specimens. Amino acid residues are shown with the IUPAC-IUB single letter code. Identity with the simple majority consensus sequence at the top is indicated by a dash (-) and deletions introduced for optimal alignment by asterisks (*). Vertical lines
indicate exon borders of the human, chicken, and toad IBSP genes, and numbers on both sides of the line denote exon numbers. Dotted vertical line shows the cleavage site for eliminating signal peptides. The poly-Es are surrounded by grizzled boxes. The conserved signature for the phosphoserine motifs and the Thr-rich regions are indicated by open boxes with solid lines and dotted lines, respectively. The white letters on shaded backgrounds represent the conserved RGD sequences.

Supplementary Fig. S1  Nucleotide and translated amino acid sequences of toad-A cDNA clone. Amino acid residues are shown with the IUPAC-IUB single letter code. The signal peptide is underlined with a dotted line. The stop codon is indicated by an asterisk (*) and the polyadenylation signals are underlined with a solid line. Primer positions and orientations are indicated by double lines with arrowheads above their boxed locations in the sequence. Vertical lines show the positions of the exon boundaries. The sequences can be accessed in the GenBank database under accession No. EU007687.

Supplementary Fig. S2  Nucleotide and translated amino acid sequences of toad-B cDNA clone. The indicators are the same as used in Fig.
S1. The sequences can be accessed in the GenBank database under accession No. **EU007688**.
Fig. 1
Fig. 2

(A) Bam  Bgl  Eco

7.2kb  →  7.5kb
5.7kb

(B) 6.6kb  5.9kb  5.7kb

6.8kb  6.5kb  6.1kb
Fig. 3
Fig. 4

Genetic distance

Toad-A
Toad-B

Cattle
Human
Mouse
Rat
Caiman
Chicken

100
100
100
45

Genetic distance

0 0.1
Fig. 5
Fig. 6
Supplementary Figure S1
Table 1. Oligonucleotide primers used for determination of exon-intron boundaries

<table>
<thead>
<tr>
<th>Designation</th>
<th>Sequence</th>
<th>Location</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toad A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBSPX2</td>
<td>CTCATCCTCATACCGTGCATAATT</td>
<td>E7</td>
<td>A</td>
</tr>
<tr>
<td>IBSPX9</td>
<td>CAAAGGAAATGGAAAGAGGATGAAG</td>
<td>E3</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX10</td>
<td>TTGATGTATTTGTCTATAACGTTGTTCCA</td>
<td>E4</td>
<td>A</td>
</tr>
<tr>
<td>IBSPX11</td>
<td>TGGAAACCAGTTATAGACAATACATCAA</td>
<td>E4</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX12</td>
<td>CTCATCTGAGCTATCTTCATTTTTCTTCTG</td>
<td>E5</td>
<td>A</td>
</tr>
<tr>
<td>IBSPX13</td>
<td>TCAGAAGAAAATGAAGATAGCTCAGATGA</td>
<td>E5</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX14</td>
<td>TTCTGGAGCTTTTCTTCTTCTTCTCCT</td>
<td>E6</td>
<td>A</td>
</tr>
<tr>
<td>IBSPX15</td>
<td>ACAAGGCATAAAATGGAAAGCAGAT</td>
<td>E6</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX17</td>
<td>CTCTTTTTGTGTTTTGAGCGTTT</td>
<td>E2</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX34</td>
<td>ATAGGAAGATAGCATATGAAACGAG</td>
<td>I2</td>
<td>A</td>
</tr>
<tr>
<td>Toad B</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IBSPX2</td>
<td>CTCATCCTCATACCGTGCATAATT</td>
<td>E7</td>
<td>A</td>
</tr>
<tr>
<td>IBSPX15</td>
<td>ACAAGGCATAAAATGGAAAGCAGAT</td>
<td>E6</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX22</td>
<td>TCTTTTTGTTGTTGAGGACAGC</td>
<td>E2</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX23</td>
<td>TCCTTTTCTTCATCATCTTTTCCAT</td>
<td>E3</td>
<td>A</td>
</tr>
<tr>
<td>IBSPX24</td>
<td>AAAGGAAATGGAAAGATGATGAAAGGA</td>
<td>E3</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX25</td>
<td>ATGTATAAGGCGCCCGTATCCATATT</td>
<td>E4</td>
<td>A</td>
</tr>
<tr>
<td>IBSPX26</td>
<td>CTATACATGGACACGATTTAGACC</td>
<td>E4</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX28</td>
<td>ACCTCATCTGAGCTATCTTCTTCTGAA</td>
<td>E5</td>
<td>A</td>
</tr>
<tr>
<td>IBSPX32</td>
<td>TTGTGTGAGTAGCTTTTTCTAAATCT</td>
<td>I4</td>
<td>S</td>
</tr>
<tr>
<td>IBSPX33</td>
<td>AGGCTATGAAAGTCTATTAGAGAACA</td>
<td>I6</td>
<td>A</td>
</tr>
</tbody>
</table>

A, antisense; S, sense; E, exon; I, intron
Table 2. Exon-intron organization of the toad *IBSP* genes

<table>
<thead>
<tr>
<th>Exon Size</th>
<th>Exon Intron Size</th>
<th>Sequence of exon-intron junctions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toad A</td>
<td></td>
<td>Splice Donor Splice Acceptor</td>
</tr>
<tr>
<td></td>
<td>(bp)</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>?</td>
<td>?······?</td>
</tr>
<tr>
<td>2</td>
<td>&gt;68</td>
<td>1.5 TCA TTT TAT gtaagtaatg ·····aacctttcag ATC AAT CGA</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>3.5 GGA AAA GGA gtaggtatattt ·····tgtttttccag ATA GCA AGG</td>
</tr>
<tr>
<td>4</td>
<td>84</td>
<td>0.6 TAT GCT AAG gtaggaacat ·····ttttcatattag GGT GGC AGC</td>
</tr>
<tr>
<td>5</td>
<td>42</td>
<td>2.0 TCA GAT GAG gtaagaatc ·····tatgtatag AGT GAT TCA</td>
</tr>
<tr>
<td>6</td>
<td>138</td>
<td>2.5 GCT CCA GAA gtaathttgct ·····tcttaatag GTG GAA GAA</td>
</tr>
<tr>
<td>7</td>
<td>&gt;1051</td>
<td></td>
</tr>
<tr>
<td>Toad B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>?</td>
<td>?······?</td>
</tr>
<tr>
<td>2</td>
<td>&gt;51</td>
<td>0.8 TCA TTT TAT gtaagtaatg ·····aacctttcag ATT AAT CGA</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>2.5 GGA AAA GGA gtaggtatattt ·····ttrttttccag ATA CCA AGG</td>
</tr>
<tr>
<td>4</td>
<td>75</td>
<td>0.8 TAT GCT AAG gtaggtacat ·····ttttctgtatag GGT GGC AGC</td>
</tr>
<tr>
<td>5</td>
<td>36</td>
<td>1.6 TCA GAT GAG gtaagaatc ·····tatctttatag AGT GAT TCA</td>
</tr>
<tr>
<td>6</td>
<td>135</td>
<td>1.5 GCT CCA GAA gtaathttgct ·····ttaataggtg GAA GAA AAT</td>
</tr>
<tr>
<td>7</td>
<td>&gt;586</td>
<td></td>
</tr>
</tbody>
</table>

Intron and exon sequences are indicated in lower-case and upper-case letters, respectively.