Available online at www.sciencedirect.com
SCIENCE DIRECT®

Biochimie 87 (2005) 411–420

BIOCHIMIE

www.elsevier.com/locate/biochi

Characterization of *Sparus aurata* osteonectin cDNA and *in silico* analysis of protein conserved features: Evidence for more than one osteonectin in Salmonidae

Vincent Laizé, António R. Pombinho, M. Leonor Cancela *

CCMAR, University of Algarve, 8005-139 Faro, Portugal

Received 29 September 2004; accepted 20 January 2005

Available online 02 March 2005

Abstract

Osteonectin is a matricellular protein involved in various cellular mechanisms but its exact function remains unclear despite numerous studies. We present here the cloning of *Sparus aurata* partial osteonectin cDNA and the reconstruction of 15 other sequences from both vertebrates and invertebrates, almost doubling the set of available sequences (a total of 35 sequences is now available). Taking advantage of the resulting large amount of data, we have created multiple sequence alignments and identified osteonectin putative conserved features (intra- and inter-disulfide bonds, collagen- and calcium-binding domains and phosphorylation sites) likely to be important for protein structure and function. This work also provides the first evidence for the presence of more than one osteonectin in some species. Finally, *S. aurata* osteonectin gene expression has been shown to initiate during larval development shortly after gastrulation, and to be high in bone-derived cell lines while down-regulated during extracellular matrix mineralization, further emphasizing the important role of osteonectin in skeletal development and bone formation.

Keywords: Osteonectin; *Sparus aurata* (teleostei); SPARC; Extracellular matrix; Fish cell line; Larval development

1. Introduction

Osteonectin, also known as SPARC or BM-40, is an acidic, noncollagenous glycoprotein (M_r ~ 40-44 kDa) associated with the extracellular matrix (ECM) and belonging to the matricellular protein family [1]. Osteonectin has been originally isolated from bovine bone matrix [2] but is also present in a wide variety of embryonic and adult vertebrate tissues containing actively proliferating and remodeling cells [3,4]. Given the binding properties of osteonectin with a number of different ECM components (collagens, vitronectin and thrombospondin 1), growth factors (PDGF, VEGF, bFGF and TGF-β), cations (calcium and cupper) and minerals (hydroxyapatite), and the phenotype abnormalities of osteonectin-deficient animals—embryonic lethality, developmental defects, cataractogenesis, osteopenia and increased adipogenesis [5,6]—it has been proposed that osteonectin might play an important role in mechanisms implicated in the interaction of the cell with the extracellular milieu. These have been reviewed by Yan and Sage and Bradshaw and Sage [3,4], and include (1) ECM organization, (2) cell proliferation, migration and differentiation, (3) cell shape and adhesive properties, (4) wound healing, (5) tissue invasion by tumor cells, and (6) angiogenesis. Surprisingly, no human disease has been associated yet with mutations in this gene. Osteonectin has been identified in a number of evolutionary different species and analysis of available sequences has revealed high sequence conservation between species: more than 70% amino acid identity among vertebrates and around 40% between vertebrates and invertebrates [3]. Osteonectin is a single polypeptide chain composed of 4 distinct domains, containing numerous intramolecular disulfide bonds and undergoing post-translational modification by N-linked glycosylation [7].

Abbreviations: ON, osteonectin; SPARC, secreted protein acidic and rich in cysteine; BM-40, basement membrane protein 40; ECM, extracellular matrix; aa, amino acid(s); bp, base pair(s); kDa, kilodalton(s); PCR, polymerase chain reaction; cDNA, DNA complementary to RNA; DNase, desoxyribonuclease; RNase, ribonuclease.

* Corresponding author. Tel.: 00 351 289 800971; fax: 00 351 289 818353.
E-mail address: lcancela@ualg.pt (M.L. Cancela).

0300-9084/$ - see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.biochi.2005.01.008

Calcium ions are bound by 2 distinct regions within the protein: an acidic N-terminal domain that binds 5–8 Ca^{2+} with low affinity and a C-terminal domain containing 2 EF-hand loops that bind 2 Ca^{2+} ions with high affinity [8]. The crystal structure of the collagen- and calcium-binding domains of ON has been determined [9–11] and revealed possible interactions between these domains.

Altogether, data accumulated from the past 23 years suggest that osteonectin has an important physiological role. However, the mechanisms by which this occurs and the protein domain(s) responsible for these actions are still poorly known. Clues to protein function are often found by identifying evolutionary conserved features, which are likely to be critical for function (substrate-binding or catalytic active sites) and/or structure. The present study aims at identifying new osteonectin sequences in order to generate an accurate multiple sequence alignment with available sequences, identifies the positions where amino acids are most conserved and predicts post-translational modifications using bioinformatics prediction tools.

2. Materials and methods

2.1. RT-PCR

Total RNA was extracted from confluent cultures of VSa16 cells [12] as described by Chomczynski and Sacchi [13]. A total of 2.5 µg of total RNA was treated with RNase-free DNase I for 1 h at 37 °C to remove possible DNA contaminations and reverse transcribed by Moloney-murine leukaemia virus (MMLV) reverse transcriptase (Invitrogen) according to manufacturer's instructions. ON cDNA fragment was then amplified by the polymerase chain reaction (PCR) in a GeneAmp 2400 thermal cycler (Perkin Elmer, Boston, USA) using *Taq* DNA polymerase (Promega, Madison, USA) and two degenerate primers SaON-01F (5'-CTGCAAGAAGGGMAARGTGTGTGAG-3') and SaON-02R (5'-CCGAACTGCCAGTGVACRGGGAA-3') designed according to available fish ON sequences. The PCR reactions were performed as follows: 2 min at 94 °C, 40 cycles of amplification (one cycle is 30 s at 94 °C, 30 s at 50 °C, 1 min at 72 °C) and a final elongation step of 10 min at 72 °C. The PCR products were size separated by electrophoresis and fragments of expected size were purified and cloned in pGEM-T Easy (Promega). Final identification was achieved by DNA sequence analysis.

2.2. Northern blot analysis

Total RNA was extracted as described above from Chomczynski and Sacchi [13] from (1) VSa13, VSa16, ABSa15 and CFSa1 fish cell lines cultured under normal conditions or treated (VSa13 and VSa16 only) with 50 µg/ml of L-ascorbic acid, 10 mM of β-glycerophosphate and 4 mM of $CaCl_2$ to induce extracellular matrix mineralization [12] and (2) *S. aurata* whole larvae collected at different developmental stages. Ten micrograms of total RNA were fractionated on 1% agarose-formaldehyde gels and transferred to a Hybond-XL nylon membrane (Amersham Biosciences, Carnaxide, Portugal) by capillary blotting with 10× standard saline citrate buffer (SSC; 1×SSC is 0.15 M NaCl and 15 mM sodium citrate, pH 7.0). The DNA probes for *S. aurata* osteonectin (480-bp fragment, GenBank accession number AY239014) and ribosomal protein L27a (467-bp fragment, GenBank accession number AY188520) were radiolabeled with [α-^{32}P]dCTP (3000 Ci/ml; Amersham Biosciences) using the random priming Rediprime II kit and purified from unincorporated nucleotides using MicroSpin G-50 columns (Amersham Biosciences). All hybridizations were performed overnight at 42 °C in ULTRAhyb buffer (Ambion, Austin, USA). Blots were washed 2 × 5 min in low stringency solution (2×SSC, 0.1% SDS) and 2 × 15 min in high stringency solution (0.1×SSC, 0.1% SDS) at 55 °C, and then autoradiographed. Relative levels of SaON mRNA were determined by densitometric methods using the Quantity One software (Bio-Rad) and normalized by comparison with L27a.

2.3. Sequence reconstruction

Expressed Sequence Tag (EST) database and WGS Trace archive (genomic raw sequences) from the GenBank (National Center for Biotechnology Information, NCBI) were extensively searched using BLAST facilities at www.ncbi.nlm.nih.gov for sequences showing similarities to known osteonectin transcripts or genes. Species-specific sequences were first clustered and elements of each cluster were assembled using CLUSTAL X software [14] to generate, after manual correction, highly accurate consensus sequences. Virtual transcripts and genes were deduced from joined consensus sequence using stringent overlap criteria. Virtual gene structure was predicted using comparative methods (homology to previously annotated genes) and electronic splicing as predicted by GENSCAN (at genes.mit.edu).

2.4. Sequence analysis

For each new gene, putative splice sites and potential coding regions were predicted using GENSCAN and confirmed manually. Putative signal peptides, *N*- and *O*-glycosylations and phosphorylation sites were identified using SIGNALP, NETNGLYC, NETOGLYC and NETPHOS facilities, respectively, at www.cbs.dtu.dk [15,16]. Conserved domains present in protein sequences were identified using InterProScan facilities at www.ebi.ac.uk [17].

2.5. Multiple sequence alignment and sequence logos

Alignment of ON sequences was created using T-COFFEE multiple sequence alignment software [18] with parameters set to default. Manual adjustments (insertion of gaps) were made in a few cases to improve alignments. Sequence logos were created from T-COFFEE multiple sequence alignments using WEBLOGO facilities at www.bio.cam.ac.uk [19].

3. Results and Discussion

3.1. Cloning of partial S. aurata osteonectin cDNA

A single DNA fragment of approximately 480 bp was obtained by PCR amplification using reverse-transcribed total RNA from VSa16 cells and SaON-01F/SaON-02R primer set (results not shown). This fragment was sequenced and identified by BLAST comparison as the partial cDNA of *S. aurata* osteonectin (SaON; GenBank accession number AY239014; Fig. 1), spanning only protein coding sequence. The peptide encoded by this cDNA fragment is 144-aa long, exhibits 20–91% identity to known osteonectins (Table 1) and contains part of the follistatin (FS) domain (Val^{1} to Val^{10}) and the complete kazal-like (Cys^{11} to Cys^{65}), Osteonectin_2 (Phe^{79} to Asn^{89}) and α-helix (Asn^{101} to Glu^{137}) domains previously identified in osteonectins. SaON contains 8 cysteine residues ($Cys^{9, 11, 17, 28, 39, 46, 65, 71}$) located in FS and kazal-like domains that have been shown previously to be highly conserved in osteonectins [20], and 1 putative site of *N*-linked glycosylation (Asn^{32}) located in the kazal-like domain.

3.2. Osteonectin gene expression in S. aurata cell lines and during larval development

Osteonectin gene expression was measured in 4 different cell lines derived from *S. aurata* vertebrae (VSa13 and VSa16), branchial arches (ABSa15) and caudal fin (CFSa1). Under normal growth conditions, SaON gene expression levels were found to be high in VSa13, VSa16 and CFSa1 cells, and low in ABSa15 cells (Fig. 2). Further investigation in bone-derived VSa13 and VSa16 cells showed a decrease of SaON gene expression levels during extracellular matrix mineralization (3 weeks of treatment) by 73% and 62%, respectively. These results indicate that SaON gene expression is strong in bone-derived cell lines and is regulated during in vitro mineralization. Osteonectin gene expression was also investigated during *S. aurata* larval development up to 6 days after hatching (DAH). Osteonectin transcript was first detected in 18 hour-old larvae indicating that it is not maternally inherited and consequently not needed for the first stages of development (Fig. 3). Expression levels were shown to progressively increase up to 2 DAH then to stabilize. Interestingly, gene expression of ribosomal protein L27a was undetectable during the first developmental stages, being switched-on only in 10 hour-old larvae undergoing gastrulation (Fig. 3). Results from northern analysis of osteonectin expression during development also suggest the existence of two transcripts, the longer transcript being also the weaker. Similar results were observed in human adult tissues [21].

Altogether, these results suggest an active role of osteonectin in fish skeletal development and bone formation, as already demonstrated in mammals [22–27].

3.3. Reconstruction of osteonectin sequences

Searching on-line public databases using BLAST revealed numerous ESTs or genomic clones of different origin with high similarity to osteonectin. Analysis of these sequences permitted to reconstruct 16 new osteonectin sequences

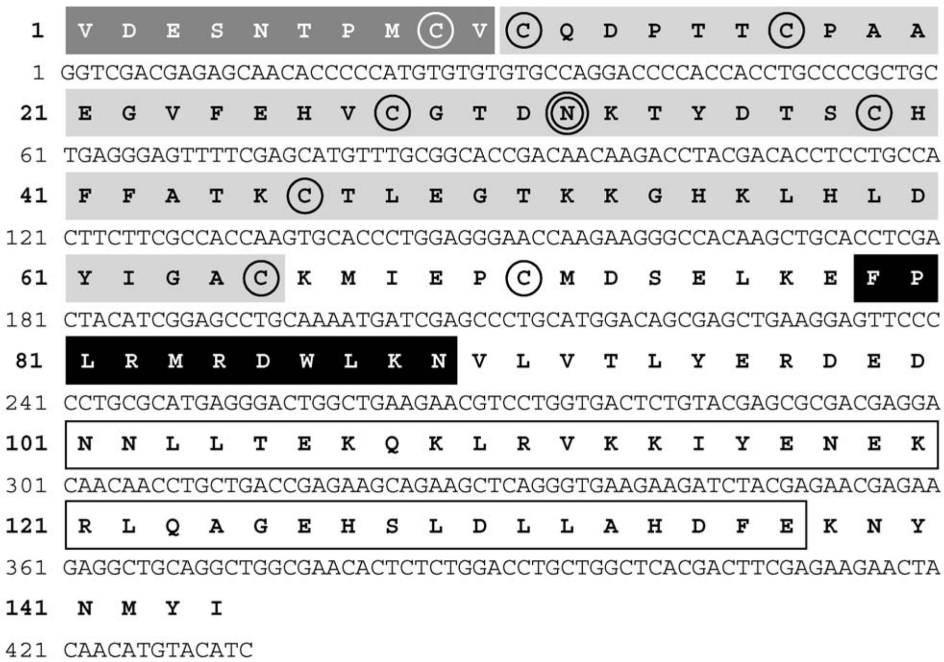


Fig. 1. Gilthead seabream [*S. aurata*] osteonectin cDNA sequence and deduced amino acid sequence. Conserved cysteine residues are circled; Putative *N*-linked glycosylation site is double-circled; Dark grey box indicates follistatin domain (incomplete); Light grey box indicates kazal-like domain; Black box indicates osteonectin_2 domain; White box indicates α-helix domain.

Table 1
Pairwise percent identities among osteonectin protein sequences. *Light grey*, invertebrates; *dark grey*, vertebrates; * partial sequences.

	Ci	Bm	Dm	Dy*	Ce	Af	Ag	Mm	Rn	Cf	Mmu	Hs	Ss	Bt	Oc	Gg	Cc	Ts*	Cn*	Es*	Xt	Xl	Rc	Ga	Tr	Tn	Ol	Ip	Dr	Om1	Om2	Ssa1	Ssa2	Sa*	Ca*
Ci																																			
Bm	20																																		
Dm	22	44																																	
Dy*	20	31	96																																
Ce	22	29	33	31																															
Af	20	49	39	30	34																														
Ag	20	45	46	42	31	42																													
Mm	32	21	21	20	28	22	21																												
Rn	32	21	21	20	28	22	21	98																											
Cf	33	21	21	21	28	22	21	92	92																										
Mmu	33	21	21	20	28	22	21	92	91	97																									
Hs	33	21	21	20	28	22	21	92	92	97	100																								
Ss	33	21	21	20	28	22	21	92	91	96	96	96																							
Bt	33	22	21	21	28	22	21	92	92	98	98	99	97																						
Oc	33	22	21	21	28	22	21	90	90	94	94	94	93	95																					
Gg	32	20	21	19	27	21	20	83	84	86	85	85	87	85	84																				
Cc	33	21	21	19	28	22	21	84	85	88	87	87	88	88	86	97																			
Ts*	46	28	31	33	32	29	30	90	90	91	91	91	90	91	91	93	94																		
Cn*	45	28	29	31	31	28	30	90	90	91	91	91	91	91	90	96	96	93																	
Es*	46	29	29	31	30	28	30	89	89	90	90	90	89	89	88	91	91	90	93																
Xt	33	22	20	18	26	21	21	79	79	81	81	81	82	81	79	80	81	89	89	89															
Xl	32	22	20	19	27	22	21	79	79	81	81	81	82	81	80	80	82	89	88	88	96														
Rc	32	22	21	20	26	22	21	77	77	81	81	81	81	80	79	79	79	87	87	86	86	86													
Ga	33	21	20	18	26	20	21	75	76	76	75	75	75	75	74	73	73	83	83	83	77	78	76												
Tr	33	20	20	19	25	20	20	72	73	73	73	73	73	73	71	72	72	82	81	81	73	73	70	81											
Tn	33	21	20	19	26	20	20	74	75	75	74	74	74	75	73	74	74	83	81	81	75	75	73	81	93										
Ol	34	20	20	19	26	22	22	77	78	78	78	78	78	78	77	78	78	85	84	83	77	77	74	82	82	86									
Ip	33	21	22	21	27	23	22	74	74	74	74	74	75	74	73	74	75	83	82	80	75	76	72	78	76	77	82								
Dr	33	21	21	19	27	22	21	76	77	76	75	75	76	75	74	76	76	84	83	81	77	77	72	78	76	79	83	89							
Om1	32	22	21	19	27	20	22	74	75	75	74	74	75	74	73	73	73	82	82	81	74	74	70	78	76	77	79	78	80						
Om2	32	22	21	21	27	21	22	74	75	76	75	75	76	75	74	73	74	84	83	81	73	75	73	80	75	77	78	79	79	94					
Ssa1	32	22	21	20	27	21	22	74	74	74	74	74	74	74	74	73	74	84	83	81	73	74	72	78	76	77	79	78	79	98	94				
Ssa2	32	22	21	21	27	21	22	75	75	77	76	76	76	76	75	74	74	84	83	81	74	76	73	80	76	78	79	80	80	95	99	95			
Sa*	42	22	27	32	23	20	23	84	84	83	84	84	83	83	83	82	83	84	83	83	83	84	81	90	90	91	87	83	86	88	89	88	89		
Ca*	41	24	26	25	27	22	24	80	80	77	77	77	77	77	76	81	80	80	78	77	82	85	77	76	76	81	78	87	92	78	77	78	77	80	

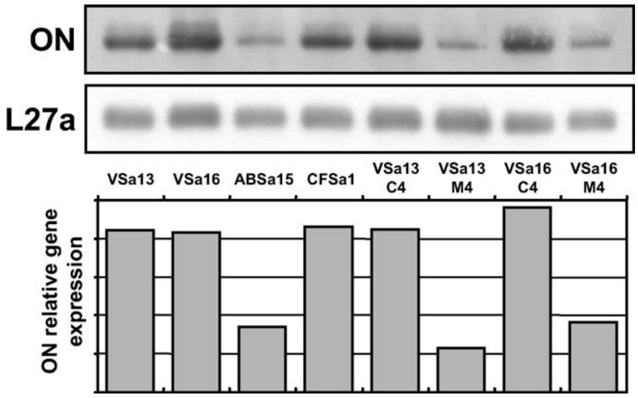


Fig. 2. Osteonectin gene expression in *S. aurata* cell lines. *Top panel* represents osteonectin and ribosomal protein L27a signals after autoradiography; *Bottom panel* represents ON relative gene expression normalized with L27a. *M4* indicates cells treated during 4 weeks for mineralization; *C4* indicates control cells left untreated.

(15 cDNAs and 1 gene; Fig. 4) from evolutionary distinct species. In most cases, only one homologue was found for each species and only the closely related fish species *Oncorhynchus mykiss* and *Salmo salar* exhibited two homologues, with 94% and 95% identity, respectively (Table 1). This finding was not unexpected since the common ancestor of all salmonids is believed to have undergone a recent genome duplication event [28]. Since that time, some genes have been lost or silenced, and others have retained their original function or evolved a new one. The two ON homologues found in trout and salmon are likely to have originated during tetraploidization of Salmonidae and whether both proteins retain their original function would have to be investigated in another study. Finally, the degree of sequence homology between vertebrates is remarkable (at least 70% sequence identity) suggesting an important function for osteonectin in bone as already proposed by [22,23,25].

A total of 35 osteonectin sequences (30 complete and 5 partial sequences) have been collected for this study (cloned, reconstructed and previously annotated sequences; Fig. 4) representing most classes of vertebrates (28 sequences)—mammals (8), bony fish (12), amphibians (3), birds (2) and reptiles (3)—and invertebrates (7 sequences)—crustaceans (1), insects (4), nematodes (1) and tunicates (1) (Fig. 4).

3.4. Domain signatures and conserved residues

Each osteonectin sequence was analyzed for the presence of known domains using SIGNALP and INTERPROSCAN facilities. All sequences exhibited the following signature sequence elements (listed from N- to C-terminal end; Fig. 5): (1) a cleavable transmembrane signal peptide (17–29 aa long) for protein secretion; (2) an acidic domain rich in glutamate, aspartate and valine residues previously shown to bind up to 8 calcium ions with low affinity [8], as well as hydroxyapatite [29]; (3) a follistatin (FS) domain possibly involved in growth factor-binding and in regulation of cell proliferation [30]; (4) a kazal-like domain. Kazal domains often occur in tandem array but only one copy has been found in osteonectins. Its presence is usually indicative of serine protease inhibitors, but it has also been reported in non-protease inhibitors; (5) Osteonectin_2 domain (PROSITE entry PS00613) involved in collagen-binding epitope [9]; (6) a C-terminal calcium-binding domain containing 2 EF-hand motifs (helix-loop-helix structure of 28–30 aa; [11]) that can bind 2 Ca^{2+} ions (one per loop) with high affinity [8]. EF-hand motifs are shared by a large number of intracellular calcium-binding proteins and are often found in single or multiple pairs. Osteonectin is so far the only example of an extracellular EF-hand protein. In addition to the 6 domains identified through domain database search, an α-helix domain is located between kazal-like and EF-hand domains. The KGHK copper-binding site identified by Lane et al. [31] and located within the kazal-like domain was found to be totally conserved among all vertebrates but absent in all invertebrate sequences. This observation is in total agreement with the proposed role of this domain in chicken angiogenesis [31], a process occurring in vertebrates but not in invertebrates.

In order to identify osteonectin conserved residues and domains, the complete sequence data set was aligned using T-COFFEE multiple sequence alignment software and a consensus sequence (made of sequence logos) was generated

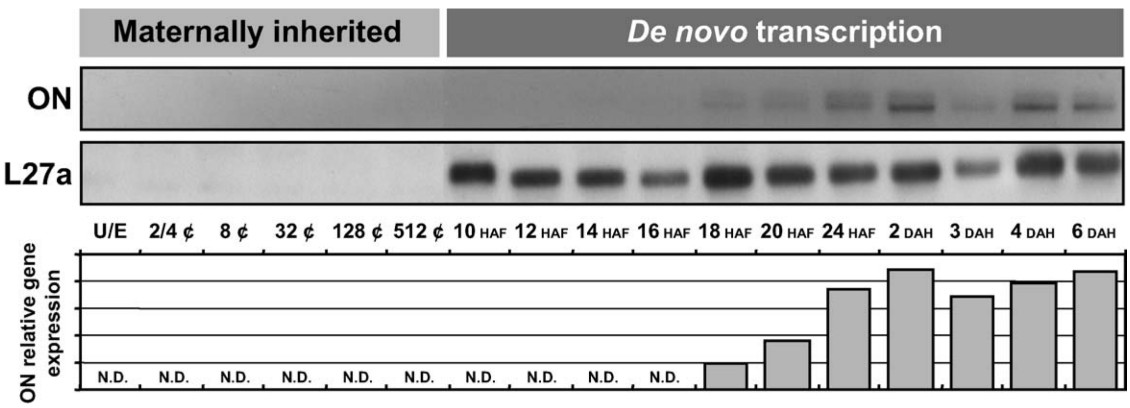
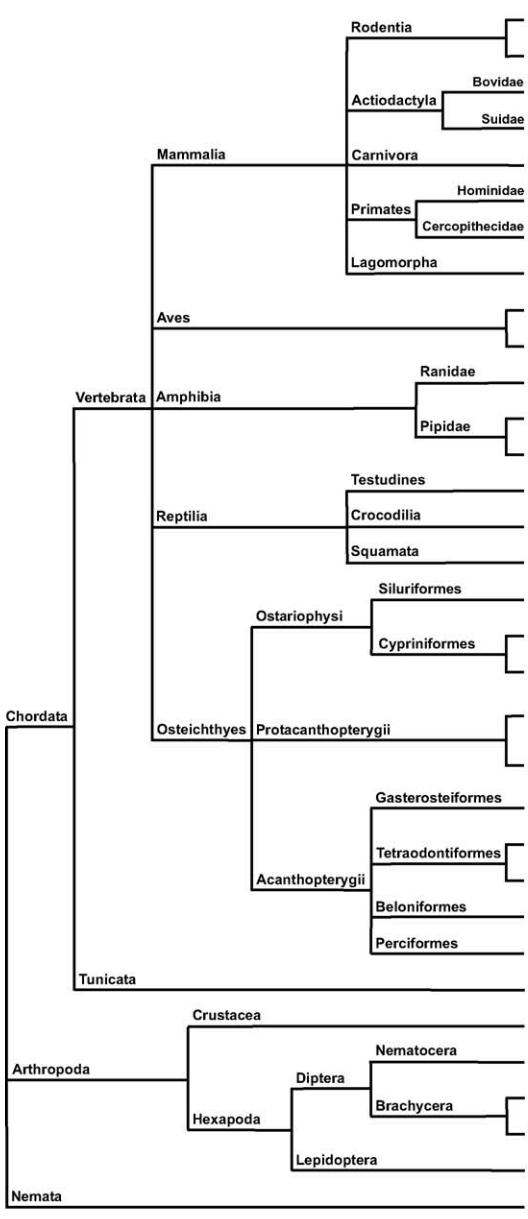


Fig. 3. Osteonectin gene expression during *S. aurata* larval development. *Top panel* represents osteonectin and ribosomal protein L27a signals after autoradiography; *Bottom panel* represents ON relative gene expression normalized with L27a; N.D., not detected; U/E, unfertilized egg; ¢, cell; HAF, hours after fertilization; DAH, days after hatching.



Acronym	Scientific name	(common name)	Accession
RnON	*Rattus norvegicus*	(Norway rat)	NM_012656
MmON	*Mus musculus*	(house mouse)	NM_009242
BtON	*Bos taurus*	(domestic cattle)	NM_174464
SsON	*Sus scrofa*	(pig)	BK005111 [a]
CfON	*Canis familiaris*	(domestic dog)	BK005116 [a]
HsON	*Homo sapiens*	(human)	NM_003118
MmuON	*Macaca mulatta*	(rhesus monkey)	BK005110 [a]
OcON	*Oryctolagus cuniculus*	(European rabbit)	BK005112 [a]
GgON	*Gallus gallus*	(chicken)	L24906
CcON	*Coturnix coturnix*	(common quail)	AF077327
RcON	*Rana catesbeiana*	(bullfrog)	AB116365
XlON	*Xenopus laevis*	(African clawed frog)	BC045013
XtON	*Xenopus tropicalis*	(western clawed frog)	BC064259
TsON	*Trachemys scripta*	(red-eared slider)	AJ243134 [b]
CnON	*Crocodylus niloticus*	(Nile crocodile)	AJ011394 [b]
EsON	*Elaphe sp.*	(snake)	AJ286863 [b]
IpON	*Ictalurus punctatus*	(channel catfish)	BK005107 [a]
DrON	*Danio rerio*	(zebrafish)	BK005113 [a]
CaON	*Carassius auratus*	(goldfish)	[39]
OmON1 OmON2	*Oncorhynchus mykiss*	(rainbow trout)	BK005114 [a] U25721
SsaON1 SsaON2	*Salmo salar*	(Atlantic salmon)	BK005117 [a] BK005118 [a]
GaON	*Gasterosteus aculeatus*	(three spined stickleback)	BK005115 [a]
TrON	*Takifugu rubripes*	(torafugu)	BK005108 [a]
TnON	*Tetraodon nigroviridis*	(green pufferfish)	BK005136 [a]
OlON	*Oryzias latipes*	(Japanese medaka)	BK005109 [a]
SaON	*Sparus aurata*	(gilthead seabream)	AY239014 [a,b]
CiON	*Ciona intestinalis*	(sea squirt)	BK005150 [a]
AfON	*Artemia franciscana*	(brine shrimp)	AB052961
AgON	*Anopheles gambiae*	(African malaria mosquito)	BK005145 [a]
DmON	*Drosophila melanogaster*	(fruit fly)	NP_651509
DyON	*Drosophila yakuba*	(fruit fly)	AY232231 [b]
BmON	*Bombyx mori*	(domestic silkworm)	BK005149 [a]
CeON	*Caenorhabditis elegans*	(round worm)	NP_500039

Fig. 4. Osteonectin sequences used in this study and taxonomy of represented species. a, this study; b, partial sequence. Taxonomic data were retrieved [September, 14, 2004] from the Integrated Taxonomic Information System on-line database, http://www.itis.usda.gov.

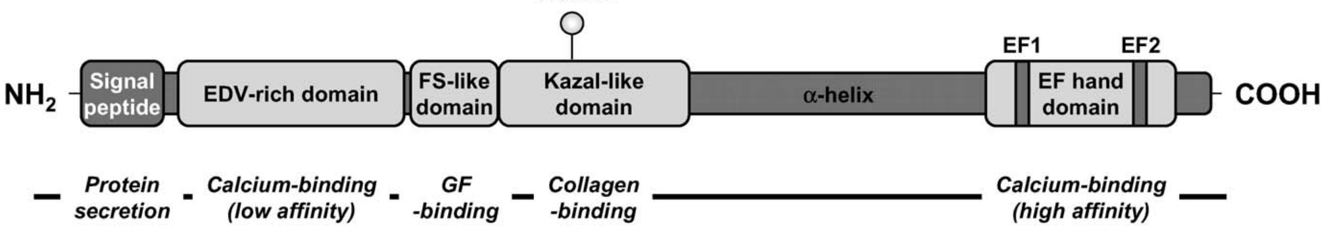


Fig. 5. Domain organization of osteonectins. GlcNAc is for *N*-linked glycosylation; FS is for follistatin; GF, is for growth factor; E, D and V are for glutamate, aspartate and valine, respectively.

using WEB LOGO facilities. The sequence logos are graphical displays where the height of each letter is made proportional to its frequency. This shows the conserved residues as larger characters (Fig. 6). As expected from previous studies of this family, osteonectins revealed strong overall conservation, dropping somewhat in the N-terminal regions corresponding to signal peptide and glutamate/aspartate rich domain. Most of the 53 highly conserved residues identified in Fig. 6 (black letters) cluster in the previously identified domains. Among these highly conserved residues, we found 13 cysteine residues corresponding to the well-known invariant cysteines located within follistatin, Kazal-like and EF-hand domains and forming intramolecular disulfide bonds probably to maintain the structural integrity of surface loops [20]. The Cys^{234}, one of the invariant cysteine residues, did not belong to a known domain and might also not be involved in any intramolecular disulfide bound. Zhou et al. [32] have reported the possible crosslinking of osteonectin with another Cys-rich protein through disulfide bonds. We suggest that Cys^{234} could be involved in this intermolecular disulfide bond.

Interestingly, osteonectin_2 motif was found to be mostly composed of highly conserved residues (9 out of 11) and is therefore likely to be a structurally or functionally important segment of osteonectin. However, despite its remarkable conservation, this domain, which is present in all osteonectins and not in any other protein, has still no assigned role.

A closer look at the organization of the two osteonectin EF-hand motifs identified an extra residue in EF-hand 1 (EF1) loop when compared to EF-hand 2 (EF2) loop (Fig. 7 and [11,33]. There are 2 main types of EF-hands: the canonical EF-hand with a 12 residue loop and the pseudo EF-hand with a 13–14 residue loop (for review see [34]). Osteonectin EF1 loop consists of 13 residues with the pattern X**Y*Z*-Y*-X**-Z and EF2 loop of 12 residues with the pattern X*Y*Z*-Y*-X**-Z. The residues X, Y, Z, -Y, -X, -Z participate in binding Ca^{2+} and the intervening residues are represented by asterisks. According to previous studies [34,35], Asp or Asn is usually found at X and Y positions; Asp, Asn, or Ser at Z; a variety of residues at -Y; usually Asp, Asn, or Ser at -X; and usually Glu at -Z. Residue between Z and -Y is

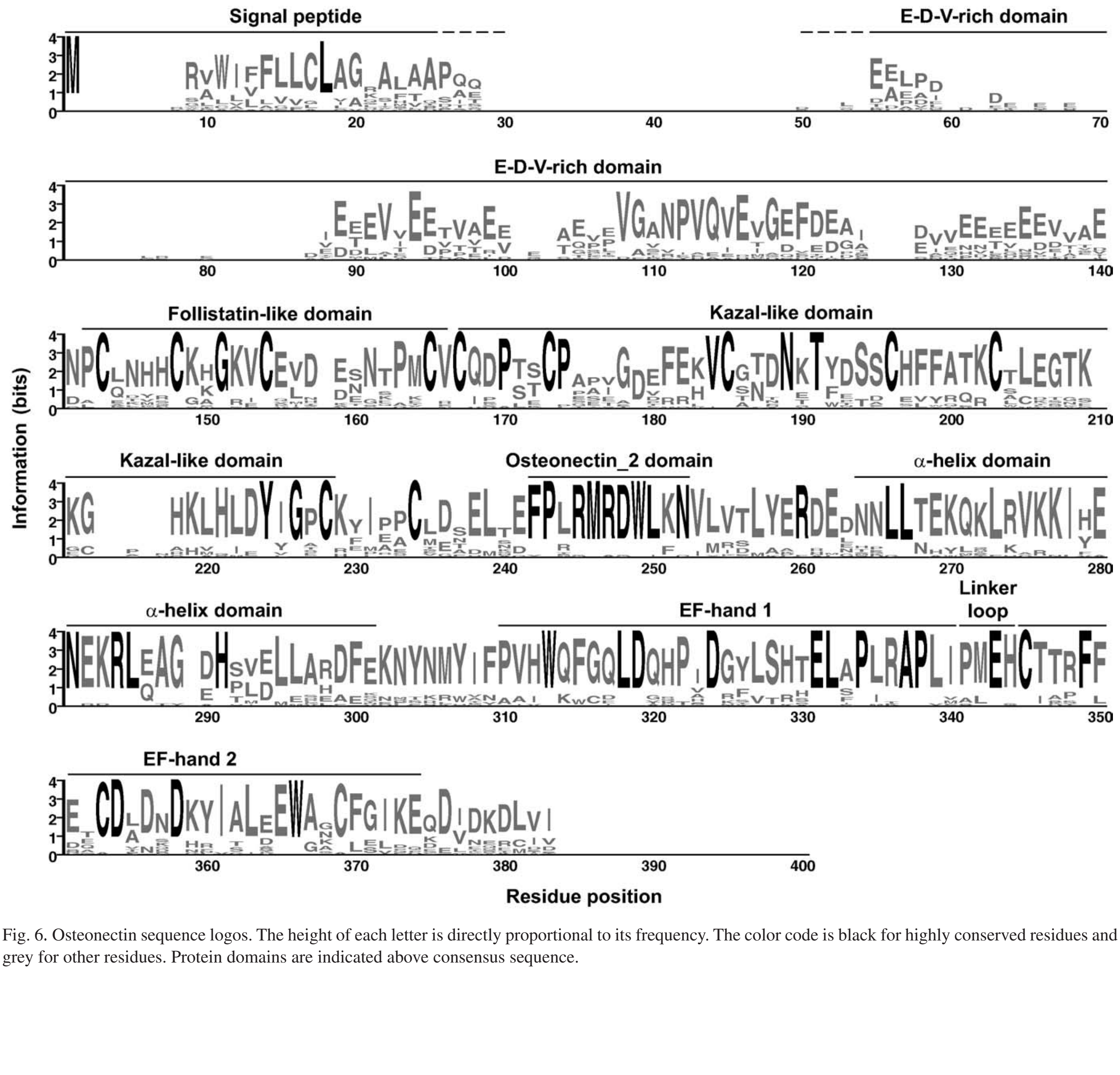


Fig. 6. Osteonectin sequence logos. The height of each letter is directly proportional to its frequency. The color code is black for highly conserved residues and grey for other residues. Protein domains are indicated above consensus sequence.

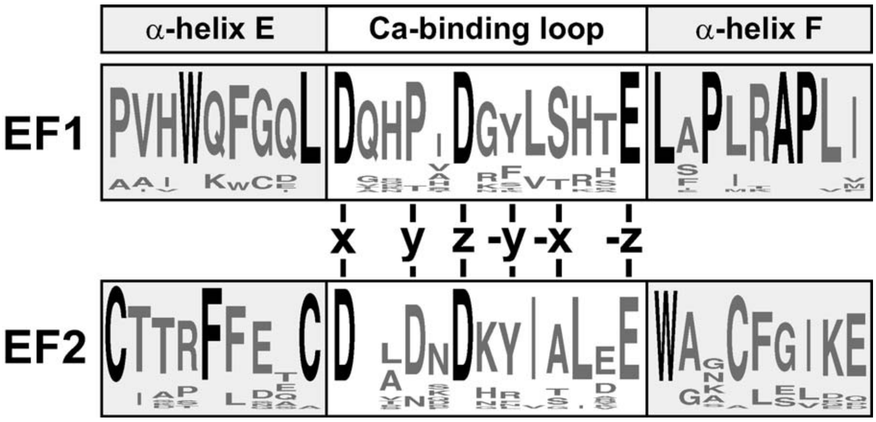


Fig. 7. EF-hand motifs of osteonectins. Putative calcium binding sites (x, y, z, -x, -y and -z) are indicated between amino acid sequences. Predicted EF-hand motif secondary structure is indicated above the sequence logos. EF1 and EF2 are for EF-hand motifs 1 and 2, respectively.

often Gly, as is Ile between -Y and -X. This residue is almost always a glutamate, but occasionally can be an aspartate. The loop linking the two EF-hands in a pair, termed the linker loop, is one of the most variable regions of EF-hand calcium-binding proteins. It varies both in composition and in length.

3.5. Conserved post-translational modifications

Searching osteonectin sequences for *N*- and *O*-glycosylations using NETNGLYC and NETOGLYC facilities identified a single putative and conserved *N*-linked glycosylation site involving the highly conserved residue Asn[189] located in Kazal-like domain (Figs. 6 and 8). No conserved site for *O*-linked glycosylation was predicted. Glycosylation of human osteonectin has been already demonstrated at Asn[99] (position equivalent to Asn[189] in our consensus sequence) and shown to be involved in collagen binding [36,37]. Our prediction made on 35 sequences obtained from evolutionary distinct species further confirms that osteonectin *N*-linked glycosylation exists and occurs at a unique and conserved site. It also suggests that this post-translational modification is essential for the function and/or structure of the protein.

Searching osteonectin sequences for tyrosine, threonine and serine phosphorylations using NETPHOS facilities identified a candidate phosphorylation site at position 328 in the

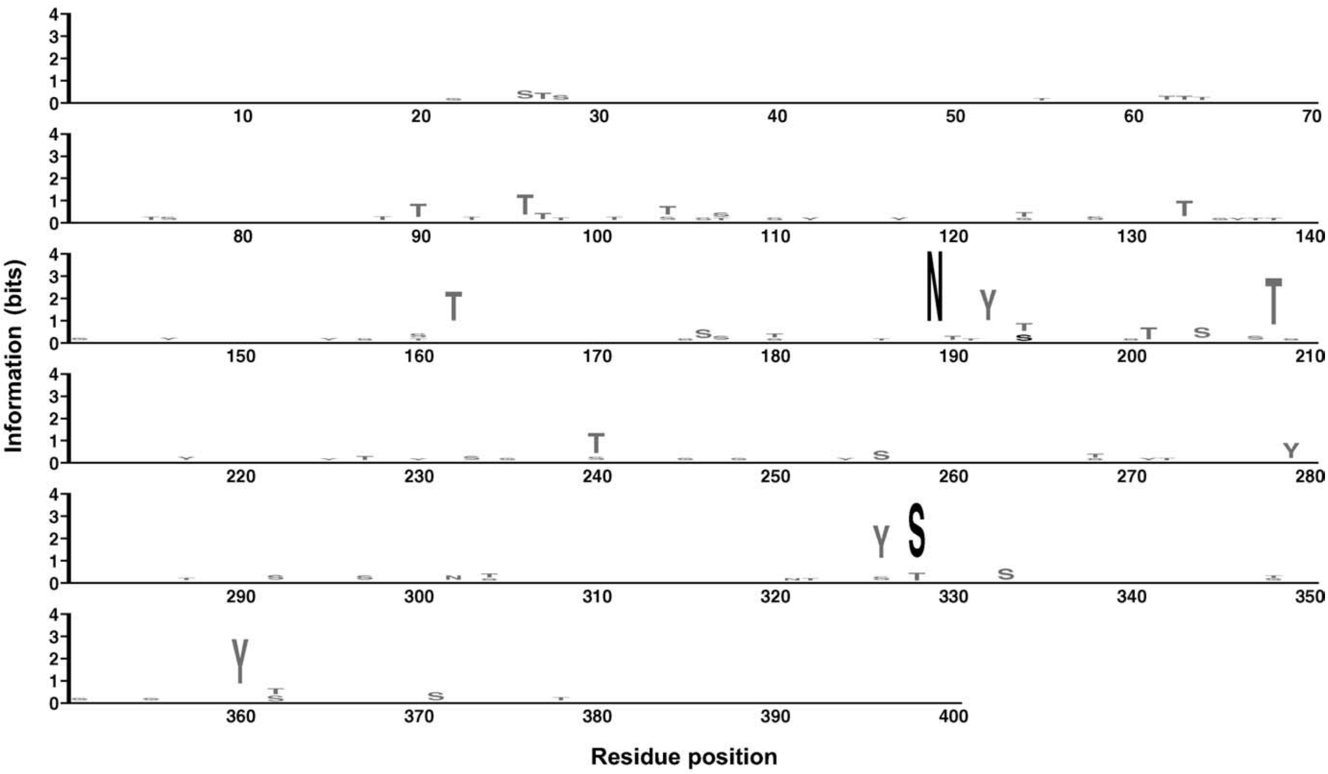


Fig. 8. Sequence logos for osteonectin post-translational modifications. Predictions were done on all 35 sequences using NETPHOS and NETNGLYC facilities and reported on ON multiple sequence alignment created using T-COFFEE facilities. The color code is black for highly conserved residues and grey for other residues. S is for serine, Y for tyrosine, T for threonine and N for asparagine. [39].

loop of EF-hand 1 domain (Figs. 6 and 8). AA[328] is a serine in 28 sequences out of 32 (not included 3 incomplete sequences) and a threonine in the other 4 sequences (GaON, TrON, TnON and OlON). Interestingly, the threonine-containing sequences are from species belonging to the same evolutionary taxon, the bony fish Acanthopterygii subgroup (Fig. 4), suggesting that a mutation event has occurred in the osteonectin gene of these fishes common ancestor. Phosphorylation of EF-hand proteins has been observed in vivo in a number of cases [38], but never within the EF-hand domain itself. Since no experimental data have confirmed so far that osteonectin could be phosphorylated, our prediction needs to be further investigated.

4. Conclusions

This work identified 16 new osteonectins (through cDNA cloning or reconstruction), almost doubling the set of available sequences (a total of 35 sequences are now available) and providing the first evidence for the presence of more than one osteonectin in some species. Taking advantage of the resulting large amount of data, including sequences from vertebrates and invertebrates, we have created multiple sequence alignments and identified highly conserved features likely to be important for protein structure and function. The overall sequence conservation is high, especially among vertebrates. Numerous conserved residues have been identified, most of them clustering into known protein domains. These conserved residues are implicated in protein structure (intra- and inter-molecular disulfide bonds), function (collagen- and calcium-binding sites) and regulation (phosphorylation site). This work has also provided evidences for osteonectin gene expression starting shortly after gastrulation and being high in bone-derived cell lines while down-regulated during extracellular matrix mineralization, further emphasizing the important role of osteonectin in development and bone formation.

Acknowledgements

This work was partially funded by the grant POCTI/BCI/48748/2002 from the Portuguese Science and Technology Foundation (FCT). V.L. and A.R.P. were the recipients of a FCT postdoctoral fellowship (BPD/1607/2000) and a CCMAR/University of Algarve fellowship, respectively. The help of Dr. E. Laizé in the identification of TnON gene is gratefully acknowledged.

References

[1] P. Bornstein, Diversity of function is inherent in matricellular proteins: an appraisal of thrombospondin 1, J. Cell Biol. 130 (1995) 503–506.

[2] J.D. Termine, H.K. Kleinman, S.W. Whitson, K.M. Conn, M.L. McGarvey, G.R. Martin, Osteonectin, a bone-specific protein linking mineral to collagen, Cell 26 (1981) 99–105.

[3] Q. Yan, E.H. Sage, SPARC, a matricellular glycoprotein with important biological functions, J. Histochem. Cytochem. 47 (1999) 1495–1505.

[4] A.D. Bradshaw, E.H. Sage, SPARC, a matricellular protein that functions in cellular differentiation and tissue response to injury, J. Clin. Invest. 107 (2001) 1049–1054.

[5] D.T. Gilmour, G.J. Lyon, M.B.L. Carlton, J.R. Sanes, J.M. Cunningham, J.R. Anderson, et al., Mice deficient for the secreted glycoprotein SPARC/osteonectin/BM40 develop normally but show severe age-onset cataract formation and disruption of the lens, EMBO J. 17 (1998) 1860–1870.

[6] P. Bornstein, E. Sage, Matricellular proteins: extracellular modulators of cell function, Curr. Opin. Cell Biol. 14 (2002) 608–616.

[7] M.E. Bolander, M.F. Young, L.W. Fisher, Y. Yamada, J.D. Termine, Osteonectin cDNA sequence reveals potential binding regions for calcium and hydroxyapatite and shows homologies with both a basement membrane protein (SPARC) and a serine proteinase inhibitor (ovomucoid), Proc. Natl. Acad. Sci. USA 85 (1988) 2919–2923.

[8] P. Maurer, U. Mayer, M. Bruch, P. Jeno, K. Mann, R. Landwehr, et al., High-affinity and low-affinity calcium binding and stability of the multidomain extracellular 40-kDa basement membrane glycoprotein (BM-40/SPARC/osteonectin), Eur. J. Biochem. 205 (1992) 233–240.

[9] T. Sasaki, E. Hohenester, W. Gohring, R. Timpl, Crystal structure and mapping by site-directed mutagenesis of the collagen-binding epitope of an activated form of BM-40/SPARC/osteonectin, EMBO J. 17 (1998) 1625–1634.

[10] E. Hohenester, P. Maurer, R. Timpl, Crystal structure of a pair of follistatin-like and EF-hand calcium-binding domains in BM-40, EMBO J. 16 (1997) 3778–3786.

[11] E. Hohenester, P. Maurer, C. Hohenadl, R. Timpl, J. Jansonius, J. Engel, Structure of a novel extracellular Ca(2+)-binding module in BM-40, Nat. Struct. Biol. 3 (1996) 67–73.

[12] A.R. Pombinho, V. Laize, D.M. Molha, S.M.P. Marques, M.L. Cancela, Development of two bone-derived cell lines from the marine teleost *Sparus aurata*; evidence for extracellular matrix mineralization and cell-type-specific expression of matrix Gla protein and osteocalcin, Cell Tissue Res. 315 (2004) 393–406.

[13] P. Chomczynski, N. Sacchi, Single-step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction, Anal. Biochem. 162 (1987) 156–159.

[14] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.

[15] N. Blom, S. Gammeltoft, S. Brunak, Sequence and structure-based prediction of eukaryotic protein phosphorylation sites, J. Mol. Biol. 294 (1999) 1351–1362.

[16] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites, Protein Eng. 10 (1997) 1–6.

[17] N.J. Mulder, R. Apweiler, T.K. Attwood, A. Bairoch, D. Barrell, A. Bateman, et al., The InterPro Database, 2003 brings increased coverage and new features, Nucleic Acids Res. 31 (2003) 315–318.

[18] C. Notredame, D.G. Higgins, J. Heringa, T-Coffee: a novel method for fast and accurate multiple sequence alignment, J. Mol. Biol. 302 (2000) 205–217.

[19] T.D. Schneider, R.M. Stephens, Sequence logos—a new way to display consensus sequences, Nucleic Acids Res. 18 (1990) 6097–6100.

[20] J. Engel, W. Taylor, M. Paulsson, H. Sage, B. Hogan, Calcium binding domains and calcium-induced conformational transition of SPARC/BM-40/Osteonectin, an extracellular glycoprotein expressed in mineralized and nonmineralized tissues, Biochemistry 26 (1987) 6958–6965.

[21] M. Young, A. Day, P. Dominquez, C. McQuillan, L. Fisher, J. Termine, Structure and expression of osteonectin mRNA in human tissue, Connect. Tissue Res. 24 (1990) 17–28.

[22] J. Termine, Cellular activity, matrix proteins, and aging bone, Exp. Gerontol. 25 (1990) 217–221.

[23] A.M. Delany, M. Amling, M. Priemel, C. Howe, R. Baron, E. Canalis, Osteopenia and decreased bone formation in osteonectin-deficient mice, J. Clin. Invest. 105 (2000) 915–923.

[24] R. Ishigaki, M. Takagi, M. Igarashi, K. Ito, Gene expression and immunohistochemical localization of osteonectin in association with early bone formation in the developing mandible, Histochem. J. 34 (2002) 57–66.

[25] A.L. Boskey, D.J. Moore, M. Amling, E. Canalis, A.M. Delany, Infrared analysis of the mineral and matrix in bones of osteonectin-null mice and their wildtype controls, J. Bone Miner. Res. 18 (2003) 1005–1011.

[26] A.M. Delany, I. Kalajzic, A.D. Bradshaw, E.H. Sage, E. Canalis, Osteonectin-null mutation compromises osteoblast formation, maturation, and survival, Endocrinology 144 (2003) 2588–2596.

[27] A.D. Bradshaw, D.C. Graves, K. Motamed, E.H. Sage, SPARC-null mice exhibit increased adiposity without significant differences in overall body weight, Proc. Natl. Acad. Sci. USA 100 (2003) 6045–6050.

[28] F.W. Allendorf, G.H. Thorgaard, Tetraploidy and the evolution of salmonid fishes, in: B.J. Turner (Ed.), Evolutionary Genetics of Fishes, Plenum Press, New York, 1984, pp. 1–46.

[29] R. Romberg, P. Werness, P. Lollar, B. Riggs, K. Mann, Isolation and characterization of native adult osteonectin, J. Biol. Chem. 260 (1985) 2728–2736.

[30] L. Patthy, K. Nikolics, Functions of agrin and agrin-related proteins, Trends Neurosci. 16 (1993) 76–81.

[31] T. Lane, M. Iruela-Arispe, R. Johnson, E. Sage, SPARC is a source of copper-binding peptides that stimulate angiogenesis, J. Cell Biol. 125 (1994) 929–943.

[32] H.Y. Zhou, E. Salih, M.J. Glimcher, Isolation of a novel bone glycosylated phosphoprotein with disulphide cross-links to osteonectin, Biochem. J. 330 (1998) 1423–1431.

[33] E. Busch, E. Hohenester, R. Timpl, M. Paulsson, P. Maurer, Calcium affinity, cooperativity, and domain interactions of extracellular EF-hands present in BM-40, J. Biol. Chem. 275 (2000) 25508–25515.

[34] M. Nelson, W. Chazin, Structures of EF-hand Ca(2+)-binding proteins: diversity in the organization, packing and response to Ca2+ binding, Biometals 11 (1998) 297–318.

[35] N.D. Moncrief, R.H. Kretsinger, M. Goodman, Evolution of EF-hand calcium-modulated proteins. 1. Relationships based on amino-acid sequences, J. Mol. Evol. 30 (1990) 522–562.

[36] R.L. Xie, G.L. Long, Role of *N*-linked glycosylation in human osteonectin: effect of carbohydrate removal by *N*-glycanase and site-directed mutagenesis on structure and binding of type V collagen, J. Biol. Chem. 270 (1995) 23212–23217.

[37] R.J. Kelm, K.G. Mann, The collagen binding specificity of bone and platelet osteonectin is related to differences in glycosylation, J. Biol. Chem. 266 (1991) 9632–9639.

[38] K. Kim, M. Son, J.B. Peterson, D.L. Nelson, Ca2+-binding proteins of cilia and infraciliary lattice of *Paramecium tetraurelia*: their phosphorylation by purified endogenous Ca2+-dependent protein kinases, J. Cell Sci. 115 (2002) 1973–1984.

[39] D.B. Lehane, N. McKie, R.G.G. Russell, I.W. Henderson, Cloning of a fragment of the osteonectin gene from goldfish, *Carassius auratus*: its expression and potential regulation by estrogen, Gen. Comp. Endocrinol. 114 (1999) 80–87.