

Identification of an Osteocalcin Isoform in Fish with a Large Acidic Prodomain*[§]

Received for publication, January 13, 2006, and in revised form, March 21, 2006. Published, JBC Papers in Press, March 24, 2006, DOI 10.1074/jbc.M600373200

Vincent Laizé^{†1}, Carla S. B. Viegas^{‡2}, Paul A. Price[§], and M. Leonor Cancela[‡]

From the [†]Centro de Ciências do Mar, Universidade do Algarve, 8005-139 Faro, Portugal and [§]Division of Biology, University of California–San Diego, La Jolla, California 92093-0368

Osteocalcin is a small, secreted bone protein whose gene consists of four exons. In the course of analyzing the structure of fish osteocalcin genes, we recently found that the spotted green pufferfish has two possible exon 2 structures, one of 15 bp and the other of 324 bp. Subsequent analysis of the pufferfish cDNA showed that only the transcript with a large exon 2 exists. Exon 2 codes for the osteocalcin propeptide, and exon 2 of pufferfish osteocalcin is ~3.4-fold larger than exon 2 previously found in other vertebrate species. We have termed this new pufferfish osteocalcin isoform OC2. Additional studies showed that the OC2 isoform is restricted to a unique fish taxonomic group, the Osteichthyes; OC2 is the only osteocalcin isoform found so far in six Osteichthyes species, whereas both OC1 and OC2 isoforms coexist in zebrafish and rainbow trout. The larger size of the OC2 propeptide is due to an acidic region that is likely to be highly phosphorylated and has no counterpart in the OC1 propeptide. We propose 1) that OC1 and OC2 are encoded by distinct genes that originated from a duplication event that probably occurred in the teleost fish lineage soon after divergence from tetrapods and 2) that the novel OC2 propeptide could be, if secreted, a phosphoprotein that participates in the regulation of biomineralization through its large acidic and phosphorylated propeptide.

Osteocalcin (OC),³ also known as bone Gla protein (or BGP), is a small secreted protein, specific for vertebrate-calcified tissues, that belongs to the vitamin K-dependent (VKD) protein family. VKD proteins are characterized by the presence of several Gla residues resulting from the post-translational vitamin K-dependent γ -carboxylation of specific glutamates and through which they can bind to a calcium-containing mineral such as hydroxyapatite (1, 2). Osteocalcin is synthesized by osteoblasts and odontoblasts as a pre-pro-protein (3) and contains 3–4 Gla residues located within a conserved domain in the central part of the mature protein (4, 5). Although a number of questions still remain concerning its mode of action at the molecular level, it has been

proposed that osteocalcin is central to the control of tissue mineralization mechanisms (6–8).

Nishimoto and colleagues (9) have recently reported osteocalcin gene and cDNA structures for two pufferfish (*Takifugu rubripes* and *Tetraodon nigroviridis*) based on computer analysis of the genomic sequences retrieved from public data bases. Surprisingly, the size of putative exon 2 in these species was much smaller than those observed for other species (including fish; Fig. 1A), raising the question of the correctness of the computer prediction, in particular because manual analysis of both genomic sequences revealed that exon 2 could be longer, predicting a far larger osteocalcin peptide (Fig. 1B). However, experimental data were missing to confirm or refute one prediction or the other. The present study brings experimental data and new insights into the structure of the osteocalcin gene from *T. nigroviridis* with the identification of a large exon 2 coding for a prodomain containing multiple possible phosphorylation sites. We also present *in silico* evidence for the presence of this osteocalcin isoform in more fish species but not in mammals, birds, or amphibians.

EXPERIMENTAL PROCEDURES

Fish Culture—Specimens of tropical fish *T. nigroviridis* (spotted green pufferfish) were obtained from a local pet shop (Picanço, Faro, Portugal) and kept until used at 22–24 °C in a closed recirculating system with water salinity at 8–12 parts/million and a natural photo period. Fish were fed once a day with frozen mollusks.

RNA Preparation and cDNA Amplification—Total RNA was extracted from the head and bones as described by Chomczynski and Sacchi (10). One microgram of total RNA was reverse-transcribed for 1 h at 37 °C using Moloney-murine leukemia virus (M-MLV) reverse transcriptase (Invitrogen), RNase Out (Invitrogen), and an oligo(dT) adapter (5'-ACGCGTCGACCTCGAGATCGATG(T)₁₃-3'). PCR amplifications of osteocalcin cDNA fragments were achieved in a PerkinElmer Life Sciences GeneAmp 2400 thermal cycler (Applied Biosystems) using 5 μ l of the reverse-transcribed RNA, *Taq* DNA polymerase (Invitrogen), and primers designed according to *T. rubripes* and *T. nigroviridis* putative OC exon 1 (PufOC-FW, 5'-CATGAAGACCCTGACTCTCTCT-3') and exon 4 (PufOC-RV, 5'-CTAGAAGGGTGGGGGGCCGTAGTA-3'). PCR reactions were performed as follows consisting of an initial denaturation step at 95 °C for 5 min, 30 cycles of amplification (one cycle is 30 s at 95 °C, 45 s at 55 °C, and 1 min at 72 °C) and a final elongation step of 10 min at 72 °C. All PCR products were size-separated on a 2% (w/v) agarose gel, and selected PCR fragments were purified using the GFX Gel Band Purification kit (Amersham Biosciences). Purified PCR products were subsequently cloned into pGEM-T Easy vector (Promega) and sequenced (Macrogen, Seoul, South Korea).

Genomic DNA Preparation and Gene Amplification—Genomic DNA was extracted from a mixture of tissues using the DNeasy Tissue kit (Qiagen). Osteocalcin gene fragments were amplified from 50 ng of

* This work was partially funded with grants from the Portuguese Science and Technology Foundation PRAXIS/BIA/11159/98, POCTI/34668/Fis/2000, and POCTI/BCI/48748/2002, the latter two including funds from Fundo Europeu de Desenvolvimento Regional and Orçamento do Estado. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[§] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. 1–3.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY840551 and AY663813.

¹ Recipient of a postdoctoral fellowship (BPD/1607/2000) from the Portuguese Science and Technology Foundation. To whom correspondence should be addressed: Centro de Ciências do Mar (CCMAR), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal. Tel: 351-289800971; Fax: 351-289818353; E-mail: vlaize@ualg.pt

² Recipient of a doctoral fellowship (BD/9077/2002) from the Portuguese Science and Technology Foundation.

³ The abbreviations used are: OC, osteocalcin; BLAST, basic local alignment search tool; MGP, matrix Gla protein.

Second Osteocalcin Isoform in Fish

genomic DNA by PCR using Advantage DNA polymerase mix (Clontech) and primers designed according to *T. rubripes* and *T. nigroviridis* putative OC exons 1 and 4 (see above). PCR reactions were performed as follows consisting of an initial denaturation step at 95 °C for 5 min, 30 cycles of amplification (one cycle is 30 s at 95 °C, 45 s at 60 °C, and 4 min at 72 °C) and a final elongation step of 10 min at 72 °C. PCR products were processed as described in the previous section.

Sequence Reconstruction—The GenBankTM expressed sequence tag data base and Trace archive were extensively searched using BLAST facilities at NCBI (www.ncbi.nlm.nih.gov) for sequences showing similarities to known osteocalcin transcripts or genes. Species-specific sequences were first clustered, and elements of each cluster were assembled using ClustalX (11) to generate, after manual correction, highly accurate consensus sequences. Virtual transcripts and genes were deduced from the joined consensus sequence using stringent overlap criteria. Virtual gene structure and splicing sites were predicted using comparative methods (homology to previously annotated genes) and GenScan facilities at the Massachusetts Institute of Technology.

Sequence Alignment and Analysis—Separate alignments for sequences of OC1 and OC2 mature protein were created using T-Coffee multiple sequence alignment software (12) with parameters set to the default. Manual adjustments were made in a few cases to improve alignments. An alignment of the whole set of sequences was produced using a similar procedure. Sequence logos presented in Fig. 6 were created from T-Coffee multiple sequence alignments using WebLogo (13). The sequence logos are presented as graphical displays, where the height of each letter is made proportional to its frequency. This shows the conserved residues as larger characters. Pairwise sequence identity values were computed as the percent of identical residues over the total number of aligned residues using alignments generated with T-Coffee and the Sequence Manipulation Suite (14).

Phylogenetic Analysis—The neighbor-joining tree was built from T-Coffee alignments of fish osteocalcin mature peptides using MEGA version 3.0 (15). The PAM (percent accepted mutation) mutation data matrix was chosen, and the rate of change was taken as site-independent (the use of a γ -distributed variable rate of change among sites was tried and produced worse results in all cases). The phylogenetic tree was generated using MEGA, where the internal branch labels, which are an estimate of branch assignment reliability, are branch support values and must not be confused with the bootstrap values produced by other programs, e.g. Phylip.

RESULTS

Cloning of Spotted Green Pufferfish Osteocalcin cDNA and Gene—PCRs performed on head/bone reverse-transcribed RNA or genomic DNA using the PufOC-FW/RV primer set amplified a cDNA fragment of 579 bp and a gene fragment of 833 bp. The sequence of both fragments was determined from three independent clones, compared with annotated sequences available in GenBankTM using the BLAST facilities at NCBI and identified as the *T. nigroviridis* osteocalcin partial cDNA and gene. The comparison of cDNA and gene sequences using the Spidey mRNA-to-genomic alignment tool at NCBI identified four exons and three introns (supplemental Fig. 1), as in other osteocalcin genes. Analysis of exon size revealed an exon 2 in the *T. nigroviridis* osteocalcin gene ~280 bp longer than those of other osteocalcin genes and 300 bp longer than the recent predictions of Nishimoto and colleagues (9) but in total agreement with the alternative prediction proposed in Fig. 1. The phase of intron insertion, as defined by Patthy (16), was found to be identical to those in other osteocalcin genes (Fig. 1). Analysis of the amino acid sequence deduced from cDNA revealed a protein with 1) a

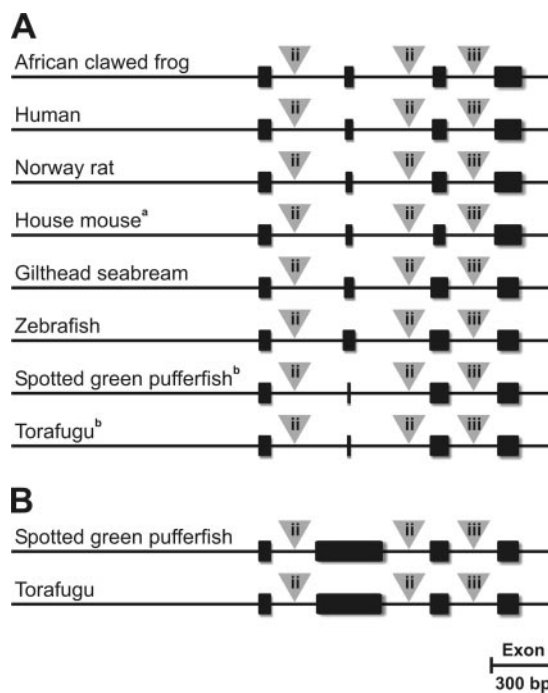


FIGURE 1. A, structural organization of osteocalcin coding sequence at the gene level. B, alternative organization of pufferfish osteocalcin genes. Black boxes indicate exons (or part of exons) representing the coding sequence, starting from the translation initiation codon and ending at the translation termination codon. Phase of intron insertion is indicated in gray triangles and is defined according to Patthy (16). ii, intron lying between the first and second nucleotides of a codon (phase 1 intron); iii, intron lying between the second and third nucleotides of a codon (phase 2 intron). a, mouse osteocalcin genes form a cluster where each gene has the same pattern. b, gene structure predicted from *in silico* analysis (9).

signal peptide, 2) a large propeptide (129 amino acids) containing 40 putative phosphoserine residues (phosphorylated by the Golgi apparatus casein kinase (17) at the SX(E/D/S(P)) recognition site), a γ -glutamyl carboxylase recognition site, a furin cleavage site, and 3) a mature peptide containing the osteocalcin signature (18) (Fig. 2).

Searching Tetraodon Assembly Release 7 using a partial *T. nigroviridis* osteocalcin gene sequence and BLAST facilities at Ensembl identified a genomic fragment (chr.Un_random) containing the complete gene sequence (supplemental Fig. 1). Only one nucleotide (positioned in intron 3) of 786 was found to be different in both sequences. The existence of a second osteocalcin transcript that would exhibit a structure similar to those already published (*i.e.* with a short exon 2) was investigated. Various attempts to clone it through PCR only amplified the same long transcript with no evidence for a shorter one.

Reconstruction of New Osteocalcin Sequences—The following nomenclature was adopted to designate osteocalcin isoforms, OC1 for osteocalcin short isoform (*i.e.* short exon 2) and OC2 for osteocalcin long isoform (*i.e.* long exon 2). Public sequence data bases were searched for osteocalcin sequences using the spotted green pufferfish osteocalcin sequence. Positive search results were clustered, and six cDNAs (red tail sheller OC2, three spined stickleback OC2, rainbow trout OC1 and OC2, zebrafish OC2, and channel catfish OC2) and two genes (torafugu OC2 and zebrafish OC2) could be reconstructed (Fig. 3 and supplemental Fig. 2). The phase of intron insertion in the newly identified genes was found to be identical to those in other osteocalcin genes. Interestingly, all OC2 sequences belong to a unique taxonomic group, the Osteichthyes or bony fish. No evidence for the presence of any OC2 sequences in mammalian, avian, or amphibian taxa was found.

Coexistence of Two Osteocalcins in Teleost Fish—Analysis of the isoform distribution in Osteichthyes (16 species represented) provided evi-

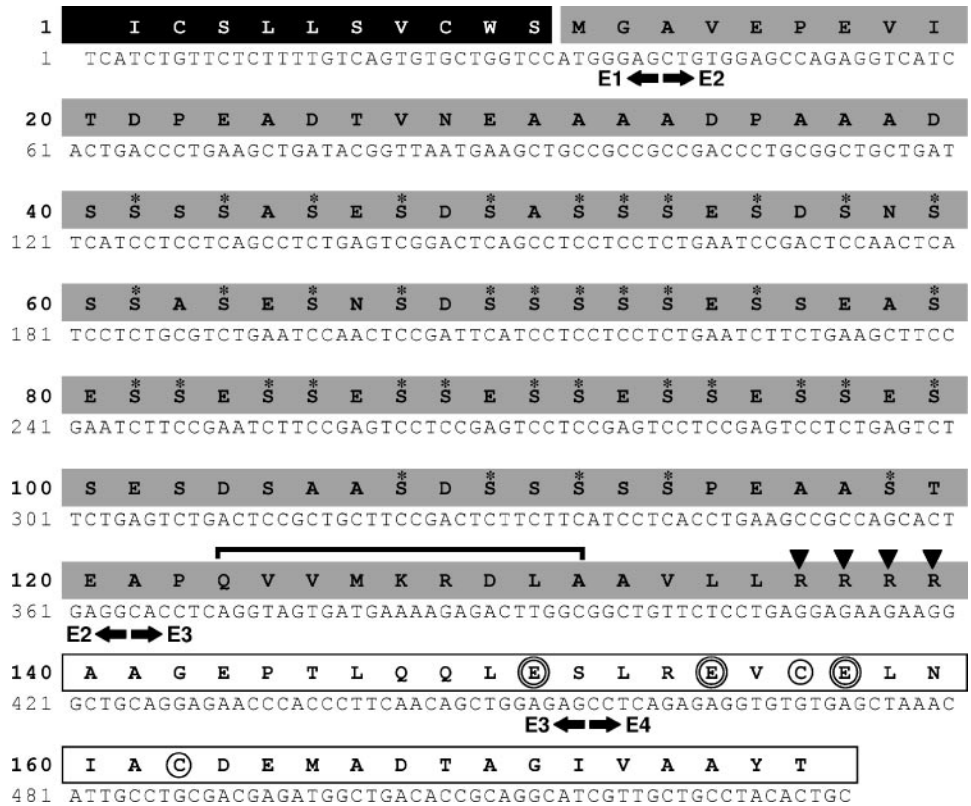


FIGURE 2. Spotted green pufferfish (*T. nigroviridis*) osteocalcin cDNA and deduced amino acid sequence. Black arrows below the nucleotide sequence indicate exon (E) limits. Black box indicates signal peptide (partial). Gray boxes indicate propeptide. White boxes indicate mature peptide (partial). Osteocalcin canonical cysteine residues are circled and predicted Gla residues are double-circled. Solid black triangles indicate the furin proteolytic cleavage site. Black line indicates putative γ -glutamyl carboxylase binding site. Asterisks indicate putative phosphorylated serine residues at SX(E/D/S)(P) sites.

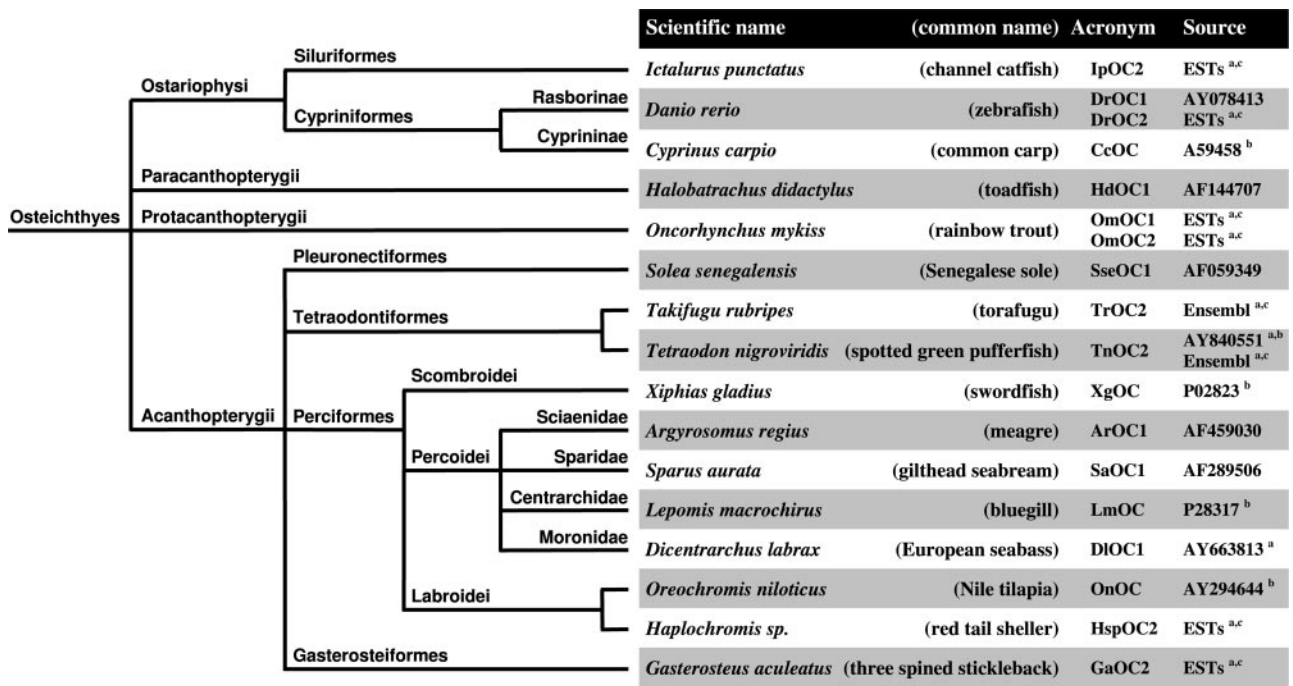


FIGURE 3. Osteocalcin sequences used in this study and taxonomy of represented species. Taxonomic data were retrieved (11–23-2004) from the Integrated Taxonomic Information System at www.itis.usda.gov. a, this study; b, partial sequence; c, reconstructed sequence presented in supplementary material.

dence for the presence of at least one isoform in 14 different species (five with OC1, five with OC2, and four with an undetermined OC) and of both isoforms in two fish (zebrafish and rainbow trout) belonging to distinct taxonomic groups (Fig. 3). Pairwise comparison of sequences outside the prodomain revealed that OC1 and OC2 are 65 and 54% identical in zebrafish and rainbow trout, respectively, suggesting that both isoforms are

most probably coded by different genes. Reconstruction of both zebrafish genes from genomic fragments obtained in the laboratory of M. L. Cancela (DrOC1, results not shown) and available from GenBank™/Ensembl data bases (DrOC2, supplemental Fig. 2) confirmed the existence of two genes in zebrafish. The DrOC2 gene was found to exhibit the same phase of intron insertion as other known osteocalcin genes.

Second Osteocalcin Isoform in Fish

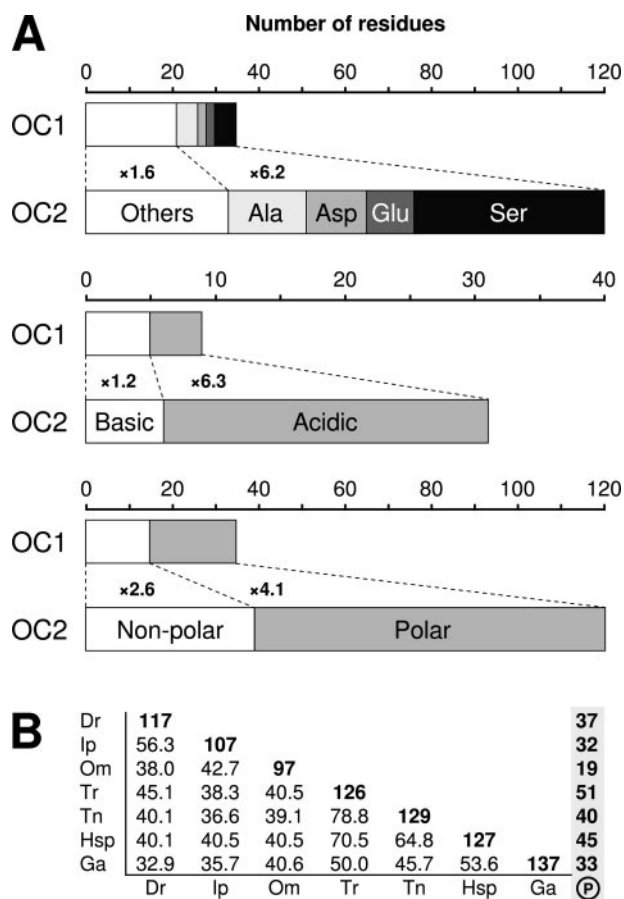


FIGURE 4. A, amino acid composition of OC1 and OC2 prodomains. Values are the mean of seven sequences for each isoform (see Fig. 3). B, pairwise percent identities among OC2 prodomain sequences. Diagonal bold values are sequence lengths and vertical bold values, putative phosphorylated serine residues. Dr, *Danio rerio*; Ip, *Ictalurus punctatus*; Om, *Oncorhynchus mykiss*; Tr, *T. rubripes*; Tn, *T. nigroviridis*; Hsp, *Haplochromis* sp.; Ga, *Gasterosteus aculeatus*.

Prodomain of OC2 Is Large, Acidic, and Contains Numerous Phosphorylation Sites—In osteocalcin genes, the entire exon 2 codes for the N-terminal part of the prodomain, a region quite variable and with no associated function. The analysis of osteocalcin prodomains from signal peptide cleavage, as predicted by SignalP (19), to propeptide cleavage at RXXR (18) revealed that 1) OC2 prodomain is more than three times longer than that of OC1 and 2) OC2 prodomain exhibits higher acidic/basic and polar/non-polar residue ratios, mainly as a result of an enrichment in aspartate and glutamate (acidic) and serine (polar) residues (Fig. 4A and Table 1). Consequently, the OC2 prodomain is highly negatively charged and has a low isoelectric point, whereas hydrophathy remains unchanged (Table 1). Another consequence of the selective enrichment in Asp, Glu, Ser, and also Ala is the low complexity of OC2 prodomain.

A T-Coffee alignment of OC2 prodomain sequences was used to evaluate interspecies sequence homology and identify conserved residues. Calculated pairwise sequence identity revealed that OC2 prodomain sequences are moderately conserved (Fig. 4B) with an average identity of 44.6% (values from closely related species, *i.e.* spotted green pufferfish and torafugu, were excluded). As a comparison, OC2 mature peptide has an average sequence identity of 80.9%. Similarly, only 30 residue positions of 172 were totally conserved among OC2 prodomains (supplemental Fig. 3). A detailed search for conserved domain/motif in OC2 prodomain only identified the RXXR signature for the furin cleav-

TABLE 1
Physicochemical analysis of OC1 and OC2 prodomain

Properties	OC1 ^a	OC2 ^a	<i>t</i> Test
Length (amino acids) ^b	35 ± 1	120 ± 14	<i>p</i> < 0.001
Molecular mass (kDa) ^b	3.7 ± 0.2	11.9 ± 1.5	<i>p</i> < 0.001
Charge ^b	+0.2 ± 2.4	-19.2 ± 7.1	<i>p</i> < 0.001
Isoelectric point ^b	6.7 ± 2.7	3.6 ± 0.2	<i>p</i> = 0.01
Hydrophathy ^c	-0.70 ± 0.22	-0.72 ± 0.15	NS ^d

^a Mean ± S.D., *n* = 7.

^b Calculated using Peptide Statistic Tool.

^c Calculated using ProtParam.

^d NS, difference is not significant.

age site. Interestingly, this signature was found to be isoform-specific; RQ(K/T)R and R(R/H/S)(R/K)R for OC1 and OC2, respectively.

Analysis of osteocalcin prodomains for phosphorylation revealed, depending on the species, 19–51 residues, mostly serines, possibly phosphorylated at the (S/T)X(E/D/S)(P) motif in OC2 prodomains (Fig. 4B). No evidence for the presence of similar phosphorylation sites was found in OC1 prodomains.

Prodomain of OC2 Is Similar to Phosphoproteins—Public sequence data bases were searched for proteins with similarity to OC2 propeptide using the BLAST facilities at NCBI. OC2 propeptide was found to have significant homology to 1) dentin phosphoprotein (a cleavage product of dentin sialophosphoprotein) and 2) phosvitin (a cleavage product of the egg yolk protein vitellogenin). Both proteins have been reported to be highly phosphorylated, and their similarity with OC2 prodomain is restricted to their phosphorylated domain.

Mature Peptide Is Isoform-specific—All mature peptides were aligned using T-Coffee and their relationship analyzed in a phylogenetic tree using the neighbor-joining method. OC1 and OC2 mature peptides were found to cluster in separate groups (Fig. 5) demonstrating their specificity. It is therefore possible to distinguish osteocalcin isoforms not only by their prodomain but also by their mature peptide. Following this idea, osteocalcin from the common carp (CcOC), the swordfish (XgOC), the bluegill (LmOC), and the Nile tilapia (OnOC), where sequence information for the propeptide is missing, were included in this analysis. CcOC, LmOC, and OnOC were found to cluster with OC1 isoforms, whereas XgOC clustered with OC2 sequences (Fig. 5). To further validate this observation, the furin cleavage site of OnOC, available with mature peptide, was found to be R-Q-K-R, fitting the signature of the OC1 proteolytic site.

Mature peptides of each osteocalcin isoform (10 sequences for OC1 and 8 sequences for OC2; see Fig. 3) were separately aligned using T-Coffee and alignments displayed as Logo sequences (Fig. 6). Analysis of each consensus sequence revealed 21 and 22 fully conserved residues (Fig. 6, *black lettering*) in mature peptide of OC1 and OC2, respectively, and 15 of those were found to be conserved in both proteins, indicating the presence of isoform-specific conserved residues (six in OC1 and seven in OC2). These residues are likely to be responsible for the clustering of OC1 and OC2 in Fig. 5. As expected, the recently published signature of osteocalcin Gla domain (18) is in total agreement with the reported OC1 consensus sequence, whereas it differs from those of OC2 at a single position (Asp in OC1 and Asp or Glu in OC2).

DISCUSSION

We present here for the first time evidence for two different osteocalcin isoforms in vertebrates. A second osteocalcin (OC2) has been identified in teleost fish and coexists in two species with the previously identified osteocalcin (OC1). The main difference between OC1 and OC2 is in the prodomain, which is large, acidic and most probably highly phosphorylated in OC2.

FIGURE 5. Phylogenetic analysis of fish osteocalcin mature peptide using MEGA version 3.0 (15).

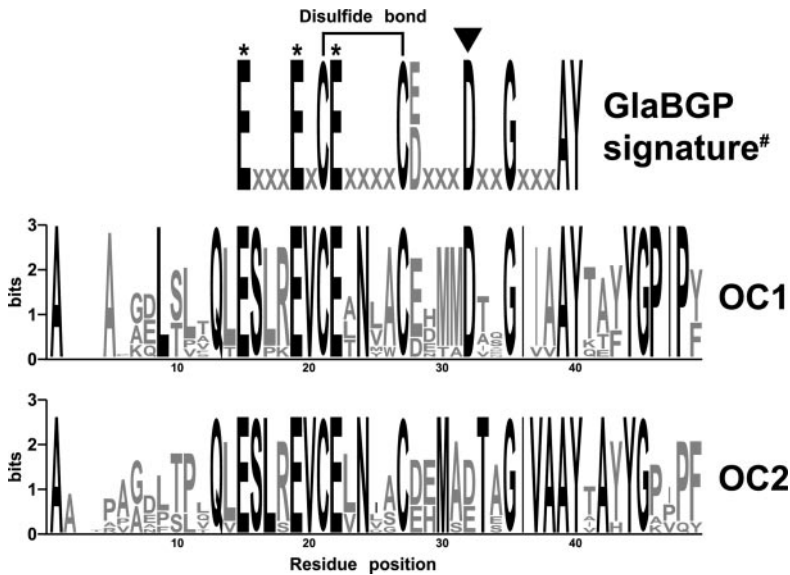
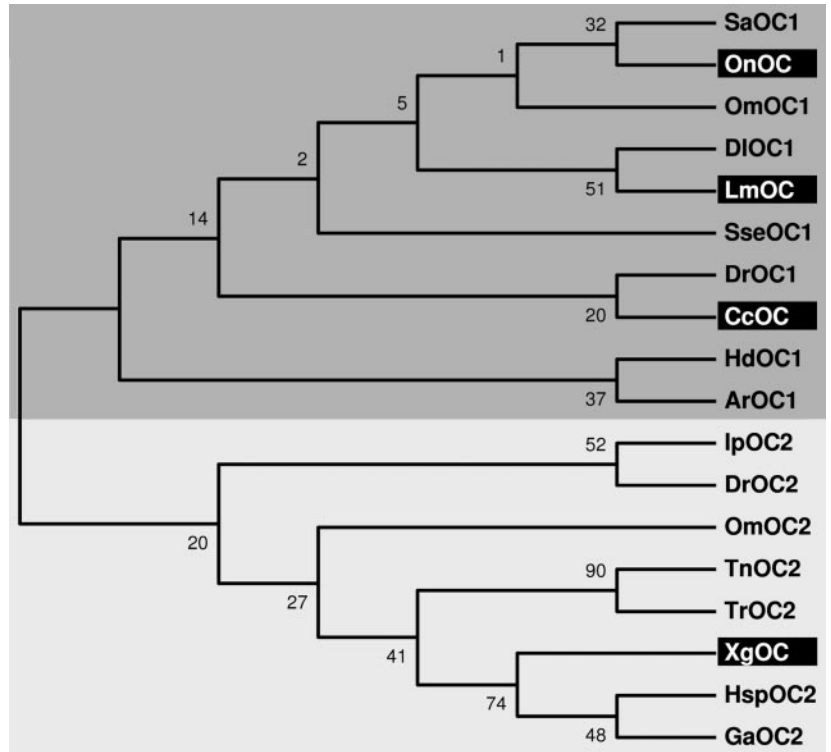


FIGURE 6. Sequence logos of fish OC1 and OC2 mature peptides. Sequence logos reveal the conservation of residues at particular positions in the protein sequence. Highly conserved residues are in black. *, Gla residues; #, signature of osteocalcin Gla domain (including non-fish sequences) as determined by Laizé *et al.* (18). Black triangle indicates variable position in OC2 when compared with signature of osteocalcin Gla domain.

Computer Predictions Must Be Experimentally Confirmed—Although invaluable and quite reliable, computer predictions have to be confirmed whenever possible by biological experiments. Nishimoto and colleagues (9) recently published a gene-based prediction of *T. nigroviridis* osteocalcin that was shown here to be erroneous with the cloning of the complementary DNA. Our work emphasizes the importance of biological support for any computer prediction.

Two Osteocalcins in Teleost Fish—Although its presence in higher vertebrates (especially in mammals) cannot be ruled out, our data provide strong indications that OC2 may be restricted to teleost fish, which represent ancient vertebrates (bony fishes diverged from tetrapod lineage approximately 450 million years ago). The analysis of isoform distribution using available genome sequence information has shown that most fishes exhibit one isoform or the other, whereas both isoforms

were found to coexist only in two species, zebrafish and rainbow trout. Comparison of osteocalcin isoforms from these species clearly showed that they share a common origin and are likely to have emerged from a gene duplication event (at a gene, chromosome, or genome level). Because both fish belong to distinct and distant taxonomic groups, this duplication event must be ancient and in fact is likely to have affected most bony fish (Fig. 7). Our hypothesis is strongly supported by recent studies analyzing zebrafish, medaka, and green pufferfish genomes, which have suggested that a whole-genome duplication event occurred in the teleost fish lineage after divergence from tetrapods (20, 21). After this event, osteocalcin duplicates have diverged and were either lost or retained. It seems from the data presented here that one of the duplicates, either OC1 or OC2, was lost in most fish species analyzed, whereas both were retained in zebrafish and rainbow trout. This is in

Second Osteocalcin Isoform in Fish

total agreement with recent evidence suggesting a massive elimination of DNA after the duplication resulting in the retention of only a subset of the duplicates in most modern teleost genomes (22, 23). Interestingly, it was found that more duplicates have been retained in the zebrafish genome than in those of the spotted green pufferfish (21). The apparent presence of only one isoform (and always OC1) in amphibians, birds, and mammals is also in total agreement with the scenario developed in Fig. 7.

Driving Force behind Isoform Selection—After the duplication of their genome, some fish species have retained both isoforms, whereas others retained either OC1 or OC2. The careful analysis of fish species presented in this study has not yet permitted us to connect the observed distribution of fish osteocalcins with their adaptation to a particular habitat (*e.g.* fresh water *versus* marine fish, deep-sea *versus* surface fish, tropical *versus* temperate fish), improved performances, faster embryonic development, or a specific evolutionary trait (cellular bone *versus* acellular bone). Similarly, analysis of data provided with each expressed sequence tag did not permit us to identify any tissue or developmental stage specific for OC2 gene expression. OC2 expressed sequence tags were found in embryo, fry, juvenile, and adult stages and in jaw, eye, and

heart tissue. Additional studies aiming at the characterization of osteocalcin isoform distribution in more fish species, as well as the determination of sites of gene expression and protein accumulation for both isoforms, will be required to better understand the driving force behind osteocalcin isoform selection. Another important question to be answered in future studies is whether both proteins retain their original function. The analysis of both mature peptides has revealed differences in amino acid composition, but key positions, as well as the peptide architecture/structure, have been conserved, suggesting conservation of the function.

Function of proOC2—The propeptide of OC1 is small (average size is 35 amino acids) and would only serve as a high affinity recognition site for vitamin K-dependent γ -glutamyl carboxylase (24). Despite exhibiting a similar carboxylase recognition site, OC2 propeptide is quite different from that of OC1, which is much longer (average size is 120 amino acids), more acidic, and serine-rich. Whether these serine residues, if any, are phosphorylated remains to be demonstrated, but the presence of phosphate groups would clearly confer a new dimension to proOC2. Numerous phosphorylated and acid residues would provide a protein with the ability to bind high amounts of calcium or calcium-containing crystals, suggesting an active role in biomineralization (25, 26). Various secreted proteins shown to contain clusters of multiple phosphoryl residues have been associated with tissue mineralization (teeth and bone) or involved in soluble calcium phosphate stabilization to prevent spontaneous precipitation (see Ref. 27 for review). Several of these proteins share some similarity with OC2, essentially within the phosphorylation domain. They are 1) the dentin phosphoprotein (a possible nucleator/promoter of crystal growth in dentin), 2) the extracellular phosphoglycoprotein osteopontin (an inhibitor of biomineralization), 3) the phosphovitin (a nucleator of crystal formation), and 4) the matrix Gla protein (a potent inhibitor of ectopic mineralization). However, the theory of proOC2 being involved in calcium binding or nucleation and growth of the inorganic calcium/phosphorus crystals becomes valid only if OC2 propeptide is found in the extracellular compartment at the proximity of the mineral. ProOC2 must be therefore secreted. Osteocalcin propeptide has been shown to be cleaved from the mature peptide by the furin, an endoprotease localized in the *trans*-Golgi network (28). Thereafter, the fate of OC1 propeptide is controversial. Some studies have suggested that proOC1 remains in the cell and is not co-secreted with osteocalcin (29), others have shown that

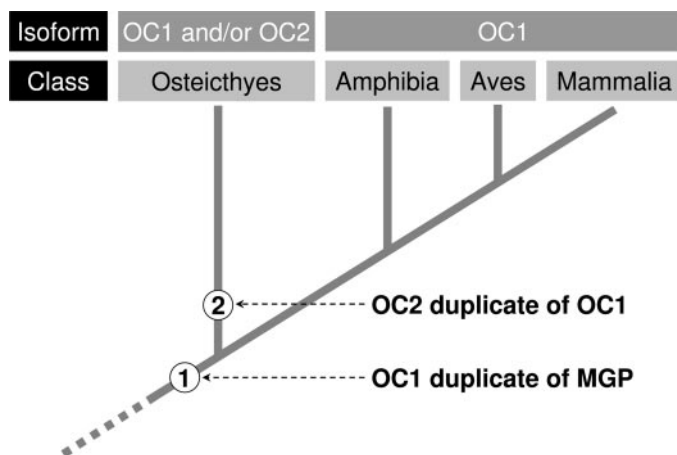


FIGURE 7. Emergence of osteocalcin isoforms 1 and 2 (OC1 and OC2) during vertebrate evolution as a result of duplication events (represented as ① and ②). Taxonomic groups (at the class level) and osteocalcin isoforms are indicated in the top panels. OC is for osteocalcin; MGP is for matrix Gla protein.

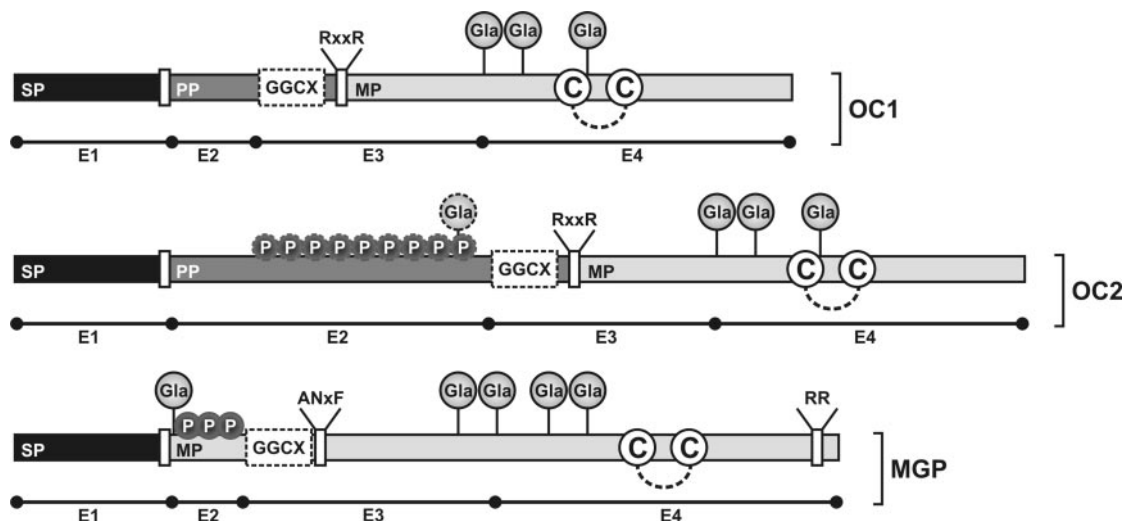


FIGURE 8. Structural organization of osteocalcin isoforms OC1 and OC2 and matrix Gla protein. Exonic structure of OC1, OC2, and MGP coding sequences is indicated below each protein structure. Dashed line represents intramolecular disulfide bond. SP, signal peptide; PP, propeptide; MM, mature peptide; Gla, γ -carboxyglutamic acid residue; GGCX, γ -glutamyl carboxylase recognition site; P, phosphorylation; C, conserved cysteine residues; E, exon; ANXF, RXXR, and RR indicate proteolytic cleavage sites.

proOC1 is present in blood serum and have even proposed that it could be used as a marker of osteoblastic function (30). The signature of the furin cleavage site was found to be isoform-specific, raising the possibility of differential cleavage efficiency. Whether OC2 propeptide is indeed cleaved and/or secreted remains uncertain and will need to be investigated in another study.

Evolutionary Relationship between OC and MGP—It has been recently proposed that osteocalcin and matrix Gla protein could be closely related, OC resulting from MGP duplication during a genome duplication, which would have occurred around 400 million years ago (18). Comparison of OC1, OC2, and MGP protein structure shows that MGP and OC2 are closer to each other than to OC1 due to the presence in MGP and OC2 of a serine phosphorylation domain located within the prodomain and entirely encoded by exon 2 (Fig. 8). A glutamic acid residue, located at the end of the putative phosphorylation domain, was found to be conserved in all OC2 prodomains (Fig. 8 and supplemental Fig. 3); could this conserved Glu residue be γ -carboxylated, as reported in prodomain-like MGP (31), reinforcing the similarity between MGP and OC2? Following this idea, could OC2 (thanks to its specific furin cleavage) be secreted as a propeptide and be, in suitable conditions, cleaved extracellularly, as it occurs for MGP at the ANXF cleavage site? In this case, both proteins would have a N-terminal phosphorylated domain capable of being cleaved, which could represent an additional step for their functional regulation. Finally, could an unprocessed OC2 (no furin cleavage) have a role similar to that of MGP (*i.e.* inhibitor of tissue calcification)? Additional studies aiming at the characterization of OC2 processing (γ -carboxylation, phosphorylation, and proteolytic cleavage of prodomain), as well as the determination of protein function, will be required to better understand the relationship between OC and MGP.

CONCLUSIONS

Our results support the hypothesis that 1) there exists in teleost fish a second osteocalcin gene (OC2), which originated from a duplication event (genome, chromosome, or gene duplication) that probably occurred in the teleost fish lineage soon after divergence from tetrapods; 2) this gene codes for a protein with a large acidic prodomain rich in phosphorylated residues that may be involved in the binding of calcium or calcium-containing crystals; and 3) OC2 could be involved as a secreted matrix-associated phosphoprotein in the regulation of biomineralization. Finally, our work emphasizes the importance for biological support of any computer prediction.

Acknowledgments—We are grateful to Dr. P. J. Gavaia and J. Sendão from the CCMAR, University of Algarve, Faro, Portugal, for the maintenance of the fish and for help in preparing tissue samples.

REFERENCES

- Poser, J. W., and Price, P. A. (1979) *J. Biol. Chem.* **254**, 431–436
- Hauschka, P. V., and Wians, F. H. (1989) *Anat. Rec.* **224**, 180–188
- Sommer, B., Bickel, M., Hofstetter, W., and Wetterwald, A. (1996) *Bone* **19**, 371–380
- Hauschka, P. V., and Carr, S. A. (1982) *Biochemistry* **21**, 2538–2547
- Frazão, C., Simes, D. C., Coelho, R., Alves, D., Williamson, M. K., Price, P. A., Cancela, M. L., and Carrondo, M. A. (2005) *Biochemistry* **44**, 1234–1242
- Ducy, P., Desbois, C., Boyce, B., Pinero, G. J., Story, B., Dunstan, C., Smith, E., Bonadio, J., Goldstein, S., Gundberg, C., Bradley, A., and Karsenty, G. (1996) *Nature* **382**, 448–452
- Hunter, G. K., Hauschka, P. V., Poole, A. R., Rosenberg, L. C., and Goldberg, H. A. (1996) *Biochem. J.* **317**, 59–64
- Boskey, A. L., Gadaleta, S., Gundberg, C., Doty, S. B., Ducy, P., and Karsenty, G. (1998) *Bone* **23**, 187–196
- Nishimoto, S. K., Waite, J. H., Nishimoto, M., and Kriwacki, R. W. (2003) *J. Biol. Chem.* **278**, 11843–11848
- Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25**, 4876–4882
- Notredame, C., Higgins, D. G., and Heringa, J. (2000) *J. Mol. Biol.* **302**, 205–217
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004) *Genome Res.* **14**, 1188–1190
- Stothard, P. (2000) *BioTechniques* **28**, 1102–1104
- Kumar, S., Tamura, K., and Nei, M. (2004) *Briefings in Bioinformatics* **5**, 150–163
- Patthy, L. (1987) *FEBS Lett.* **214**, 1–7
- Christensen, B., Nielsen, M. S., Haselmann, K. F., Petersen, T. E., and Sorensen, E. S. (2005) *Biochem. J.* **390**, 285–292
- Laizé, V., Martel, P., Viegas, C. S. B., Price, P. A., and Cancela, M. L. (2005) *J. Biol. Chem.* **280**, 26659–26668
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) *Protein Eng.* **10**, 1–6
- Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., Boudet, N., Castellano, S., Anthouard, V., Jubin, C., Castelli, V., Katinka, M., Vacherie, B., Biemont, C., Skalli, Z., Cattolico, L., Poulain, J., De Berardinis, V., Cruaud, C., Duprat, S., Brottier, P., Coutanceau, J. P., Gouzy, J., Parra, G., Lardier, G., Chapple, C., McKernan, K. J., McEwan, P., Bosak, S., Kellis, M., Volff, J. N., Guigo, R., Zody, M. C., Mesirov, J., Lindblad-Toh, K., Birren, B., Nusbaum, C., Kahn, D., Robinson-Rechavi, M., Laudet, V., Schachter, V., Quetier, F., Saurin, W., Scarpelli, C., Wincker, P., Lander, E. S., Weissenbach, J., and Roest Crollius, H. (2004) *Nature* **431**, 946–957
- Taylor, J. S., Braasch, I., Frickey, T., Meyer, A., and Van de Peer, Y. (2003) *Genome Res.* **13**, 382–390
- Woods, I. G., Wilson, C., Friedlander, B., Chang, P., Reyes, D. K., Nix, R., Kelly, P. D., Chu, F., Postlethwait, J. H., and Talbot, W. S. (2005) *Genome Res.* **15**, 1307–1314
- Amores, A., Suzuki, T., Yan, Y.-L., Pomeroy, J., Singer, A., Amemiya, C., and Postlethwait, J. H. (2004) *Genome Res.* **14**, 1–10
- Engelke, J. A., Hale, J. E., Suttie, J. W., and Price, P. A. (1991) *Biochim. Biophys. Acta* **1078**, 31–34
- He, G., Ramachandran, A., Dahl, T., George, S., Schultz, D., Cookson, D., Veis, A., and George, A. (2005) *J. Biol. Chem.* **280**, 33109–33114
- Gorski, J. P. (1992) *Calcif. Tissue Int.* **50**, 391–396
- Huq, N. L., Cross, K. J., Ung, M., and Reynolds, E. C. (2005) *Arch. Oral Biol.* **50**, 599–609
- Nakayama, K. (1997) *Biochem. J.* **327**, 625–635
- Gundberg, C. M., and Clough, M. E. (1992) *J. Bone Miner. Res.* **7**, 73–80
- Hosoda, K., Kanzaki, S., Eguchi, H., Kiyoki, M., Yamaji, T., Koshihara, Y., Shiraki, M., and Seino, Y. (1993) *J. Bone Miner. Res.* **8**, 553–565
- Price, P. A., and Williamson, M. K. (1985) *J. Biol. Chem.* **260**, 4971–4975