JOURNAL OF BONE AND MINERAL RESEARCH
Volume 20, Number 8, 2005
Published online on March 21, 2005; doi: 10.1359/JBMR.050318

Identification of a New pebp2αA2 Isoform From Zebrafish *runx2* Capable of Inducing Osteocalcin Gene Expression In Vitro

Jorge P Pinto,[1,2] Natércia M Conceição,[1] Carla SB Viegas,[1] Ricardo B Leite,[1] Laurence D Hurst,[3] Robert N Kelsh,[4] and M Leonor Cancela[1]

ABSTRACT: The zebrafish runx2b transcription factor is an ortholog of RUNX2 and is highly conserved at the structural level. The runx2b *pebp2αA2* isoform induces osteocalcin gene expression by binding to a specific region of the promoter and seems to have been selectively conserved in the teleost lineage.

Introduction: RUNX2 (also known as CBFA1/Osf2/AML3/PEBP2αA) is a transcription factor essential for bone formation in mammals, as well as for osteoblast and chondrocyte differentiation, through regulation of expression of several bone- and cartilage-related genes. Since its discovery, Runx2 has been the subject of intense studies, mainly focused in unveiling regulatory targets of this transcription factor in high vertebrates. However, no single study has been published addressing the role of Runx2 in bone metabolism of low vertebrates. While analyzing the zebrafish (*Danio rerio*) *runx2* gene, we identified the presence of two orthologs of RUNX2, which we named *runx2a* and *runx2b* and cloned a *pebp2αA*-like transcript of the *runx2b* gene, which we named *pebp2αA2*.
Materials and Methods: Zebrafish *runx2b* gene and cDNA were isolated by RT-PCR and sequence data mining. The 3D structure of runx2b *runt* domain was modeled using mouse Runx1 *runt* as template. The regulatory effect of *pebp2αA2* on *osteocalcin* expression was analyzed by transient co-transfection experiments using a luciferase reporter gene. Phylogenetic analysis of available Runx sequences was performed with TREE_PUZZLE 5.2. and MrBayes.
Results and Conclusions: We showed that the *runx2b* gene structure is highly conserved between mammals and fish. Zebrafish *runx2b* has two promoter regions separated by a large intron. Sequence analysis suggested that the *runx2b* gene encodes three distinct isoforms, by a combination of alternative splicing and differential promoter activation, as described for the human gene. We have cloned a *pebp2αA*-like transcript of the *runx2b* gene, which we named *pebp2αA2*, and showed its high degree of sequence similarity with the mammalian *pebp2αA*. The cloned zebrafish *osteocalcin* promoter was found to contain three putative runx2-binding elements, and one of them, located at –221 from the ATG, was capable of mediating *pebp2αA2* transactivation. In addition, cross-species transactivation was also confirmed because the mouse Cbfa1 was able to induce the zebrafish *osteocalcin* promoter, whereas the zebrafish pebp2αA2 activated the murine *osteocalcin* promoter. These results are consistent with the high degree of evolutionary conservation of these proteins. The 3D structure of the runx2b *runt* domain was modeled based on the *runt* domain of mouse Runx1. Results show a high degree of similarity in the 3D configuration of the DNA binding regions from both domains, with significant differences only observed in non-DNA binding regions or in DNA-binding regions known to accommodate considerable structure flexibility. Phylogenetic analysis was used to clarify the relationship between the isoforms of each of the two zebrafish *Runx2* orthologs and other Runx proteins. Both zebrafish *runx2* genes clustered with other Runx2 sequences. The duplication event seemed, however, to be so old that, whereas Runx2b clearly clusters with the other fish sequences, it is unclear whether Runx2a clusters with Runx2 from higher vertebrates or from other fish.
J Bone Miner Res 2005;20:1440–1453. Published online on March 21, 2005; doi: 10.1359/JBMR.050318

Key words: *pebp2αA2*, runx2, osteocalcin, evolutionary conservation

INTRODUCTION

The transcription factor RUNX2 (also known as CBFA1, OSF2, PEPB2αA, or AML3) is a protein from the RUNT family, whose members regulate a broad spectrum of functions and share a 128 amino acid residue DNA-binding, or *runt*, domain, highly conserved across evolution.[1] The first member of this family to be identified was the *Drosophila* gene *runt*, named according to its function in the establishment of segmentation patterns during embryogenesis, and later discovered to also play a role in sex determination and neurogenesis. A second *Drosophila*

The authors have no conflict of interest.

[1]CCMAR, University of Algarve, Campus de Gambelas, Faro, Portugal; [2]Present address: Laboratory of Molecular Immunology, Institute for Molecular and Cell Biology, Rua do Campo Alegre, Porto, Portugal; [3]Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom; [4]Centre for Regenerative Medicine, Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom.

gene of the same family, *lozenge*, was found to be required for cell patterning in the eye and for hematopoiesis. Other invertebrate species in which runx proteins have been characterized include the nematode *Caenorhabditis elegans* and the echinoderm *Strongylocentrotus purpuratus*, where a single *runx* gene has thus far been identified.[2]

Three *Runx* genes have been identified in vertebrates to date. *Runx1* is required for hematopoiesis[3] and is frequently mutated in human leukemia. *Runx3* controls cell proliferation[4] and is frequently downregulated or silenced in human gastric and colorectal cancers.[5] A third gene, *Runx2,* has been shown to exert a critical role in regulating the switch from cartilage to bone and to be required for normal bone formation and osteoblast differentiation. Accordingly, deletion of the *Runx2* gene in mouse generated animals with a normally patterned but exclusively cartilaginous skeleton, because of an arrest of osteoblast differentiation. Additionally, a link was established between *Runx2* haploinsufficiency and cleidocranial dysplasia, an autosomal dominant skeletal dysplasia in mouse and humans.[6,7] Unexpectedly, mouse *Runx2* was also found to be required for bone remodeling[8,9] and chondrocyte differentiation and function.[10] Concomitant with this multiplicity of roles, Runx2 has been shown to regulate, in mammals, the expression of a variety of genes including osteocalcin,[11] type I and X collagens,[12,13] osteopontin,[14] bone sialoprotein,[15] sclerostenosis,[16] and vascular endothelial growth factor.[17] In addition to their role as transcriptional regulators, through direct binding to target promoters via their *runt* domains, the Runx2 proteins have been shown to establish complexes with proteins involving domains other than *runt*,[18] thereby influencing their regulatory capacities and function. Finally, studies in mammals have shown that *Runx* genes can function either as proto-oncogenes or tumor suppressors, through transcriptional regulation of genes implicated in tumor invasion and metastasis.[19]

Three major N-terminal isoforms of Runx2 are known, each regulated by distinct promoters.[11,20–24] They are designated Pebp2αA (type I/p56, starting with the sequence MRIPV), *til-1* (type II/p57, starting with the sequence MASNS), and Cbfa1/Osf2 (type III, beginning with MLHSPH). The Pebp2αA transcript is predominantly expressed in T cells,[20,25,26] but also in osteoblasts,[15,21] and its expression does not change with the differentiation status of the cells. In contrast, *til-1* expression is increased during differentiation of primary osteoblasts and is induced in osteoprogenitors and in premyoblast C2C12 cells in response to bone morphogenetic protein-2.[15] As for Cbfa1/Osf2, its expression seems to be restricted to bone and osteoblasts.[11]

Teleost fish are considered to have been the first group to develop a bony skeleton and, concomitantly, all the molecular machinery necessary for its formation and maintenance. Teleost *runx2* genes/cDNA have been isolated from medaka (*Oryzias latipes*),[27] fugu (*Takifugu rubripes*),[28] Tetraodon (*Tetraodon nigroviridis*; V Laizé and ML Cancela unpublished data, 2002), and zebrafish (*Danio rerio*),[29] but there is an almost complete lack of functional information for these proteins in lower vertebrates. Recently, the first case of a duplication of a *runx2* gene was described in zebrafish, the two copies being located at different sites in the genome (on LG 17 and 20), and presenting some differences in their temporal and spatial expressions.[29] The Runx2 transcription factor was first identified as a regulator of *osteocalcin* gene transcription and, since then, new target genes and functions have been unveiled. However, the mechanisms by which Runx2 regulates *osteocalcin* transcription seem to be similar to those involved in its interaction with other genes. Analysis of runx2/*osteocalcin* interactions in lower vertebrates will permit to address the question of whether the basic regulatory mechanisms of bone formation and osteoblast/chondrocyte differentiation have been conserved throughout evolution. We have therefore cloned the zebrafish *runx2* gene and studied the degree of conservation of its gene structure and function. Interestingly, we show strong conservation of *runx2* and remarkable conservation of transcriptional activation of the osteoblast marker *osteocalcin* in vitro across species. Additionally, we present the first structural model of the DNA-binding motif of a Runx2 protein and address the evolutionary hypothesis that arises from knowing the structure of the duplicated zebrafish *runx2* genes.

MATERIALS AND METHODS

Cloning of the zebrafish complete pebp2αA2 cDNA

Degenerate primers (zfRunx2F1 and zfRunx2R1; Table 1) were designed to obtain a zebrafish partial *runx2* cDNA by RT-PCR, which in turn, was used to construct specific primers to amplify the 5′-end (zfRunx2R2; Table 1) and to extend the available 3′ sequence (zfRunx2F2 and zfRunx2R3; Table 1) beyond the termination codon of the zebrafish *runx2* isoform I (*pebp2αA*). Amplification was performed by 5′- and 3′-RACE with Advantage cDNA polymerase mix (Clontech), using zebrafish Marathon cDNA libraries as templates, under conditions suggested by the supplier. All sequence alignments were performed with CLUSTAL W.[30]

Determination of the zebrafish runx2b gene structure

The exon/intron boundaries of the zebrafish *runx2b* gene (*runx2b*) were amplified by PCR, using specific exonic primers designed in regions adjacent to where conserved intronic insertions were known to be present in other species. Amplified sequences were confirmed and extended into introns after comparison with complete or partial intronic sequences resulting from in silico analysis of the available zebrafish genome (http://www.sanger.ac.uk/Projects/D_rerio/ and http://www.ncbi.nlm.nih.gov/genome/seq/DrBlast.html). All intronic/exonic boundaries were confirmed by direct sequence analysis.

Cloning and analysis of the zebrafish runx2b proximal and osteocalcin promoters

The *osteocalcin and runx2b* proximal promoter sequences were amplified with the Universal Genome Walker kit (Clontech), using reverse specific primers within the zebrafish *osteocalcin* cDNA (PY Gavaia and ML Can-

Table 1. Oligonucleotides Used for PCR Amplification

Name	*Sequence**
Sense primers	
zfRunx2F1	ATCATCGC(G/C)GACCACCC(A/G)GCCG
zfRunx2F2	CCCAACTTTCTCTGCTCGGTTTTGCCTTC
zfBGP WtF2	*CCGGAGCTCGAG*ACTCTGACCACACAGAGTC
zfBGP WtF3	*CCGGAGCTCGAG*CAACAAACCCCACACGACATC
zfBGP MutF2	*CCGGAGCTCGAG*ACTCTGAGAACACAGAGTC
zfBGP MutF3	*CCGGAGCTCGAG*CAACAAACCGAACACGACATC
Antisense primers	
zfBGPR1	*CACGCAAGCTTGCTGATC*TGTGTGCGTGTG
zfRunx2R1	GCCACCTGGTTCTTCATCAC
zfRunx2R2	CCTGGTTCTTCATCACCCCTGAGGCATTC
zfRunx2R3	CTGAAAGCCAGGCGAATACATC
zfRunx2R4	CCAATGAGAAGGCAAAACCGAGCAGAGAAA
zfRunx2R5	AAGCACCGTATGTCCTCCGTTTGCGTGAGA

* Underlined and italic sequences in sense primers are *Xho*I sites, underlined and italic sequences in antisense primers are *Hin*dIII sites.

cela, unpublished data, 2001) and the *runx2b* proximal promoter (zfRunx2R4 and zfRunx2R5; Table 1), under conditions specified by the manufacturer. Identification of putative binding motifs in the sequences obtained was performed with MatInspector v.2.2 and Genomatix software.[31] Comparative sequence analysis of zebrafish *Pebp2αA2* and mouse *Pebp2αA* promoters was made with Pipmaker.[32]

Determination of the 3D structure of the zebrafish Pebp2αA2 runt domain

The comparative model of the DNA binding domain of *Pebp2αA2* (*runt* domain) was obtained using the MODELLER 6v2 software.[33] This software requires a pre-existent structure, with a predicted high similarity with the query structure. The most appropriate structure found was the *runt* domain of the mouse Runx1 protein, with the Protein Data Bank code 1EAN.[34] This structure has a 1.25 Å resolution and reveals 90% of sequence identity with the putative runx2b DNA binding domain, spanning a 114 amino acid region.

The spatial constraints were derived by transferring the spatial features from the known protein structure to the unknown structure. The model showing the smallest pseudo-energy was taken as final modeled structure. The stereochemistry of the final structure was checked using PROCHECK,[35] which showed no residues in disallowed regions of the Ramachandran plot. Figures 6A and 6B were created using DeepView 3.7 SP5.

Reporter plasmids

Constructs ocProm-Luc and pebp2αA2ORF-pCMX-PL1 were obtained by cloning, respectively, the promoter of zebrafish *osteocalcin* (nucleotides –25 to –840 from the initiation codon; accession number AY178836) into the pGL2-Luc reporter plasmid (Promega) and the zebrafish *pebp2αA2* open reading frame (ORF; nucleotides –25 to +1426 from the ATG) into the pCMX-PL1 expression vector (kindly provided by Dr Roland Schuele, University of Freiburg, Freiburg, Germany).

Osteocalcin promoter deletion constructs, ocwt2 and ocwt3, were obtained by PCR amplification of the full *osteocalcin* promoter, using forward primers designed on the motifs R2 and R3, respectively, and bearing a specific restriction site at their 5′ ends (zfBGP WtF2, zfBGP WtF3; Table 1) and a reverse primer zfBGPR1 (Table 1), bearing a specific restriction site at its 3′ end. Point mutations were generated in the putative runx2-binding motifs R2 (ocMut2) and R3 (ocMut3) by PCR amplification of the wildtype sequence with forward primers (zfBGP MutF2, zfBGP MutF3; Table 1) containing a two-base pair mutation in motif R2 and R3, respectively, and the same specific reverse primer (zfBGPR1; Table 1). PCR products were cloned into the pGL2-Luc reporter plasmid and used in transfection experiments. All constructs obtained were confirmed by DNA sequence analysis. Plasmids used for transfection studies were prepared using Qiagen plasmid Maxi Kit. The mouse *Osf2/Cbfa1* ORF and o*steocalcin* promoter (mOG2) constructs were a kind gift from Dr Gerard Karsenty (Baylor College of Medicine, Houston, TX, USA).

Cell culture, transfection, and luciferase assays

The *Xenopus laevis* A6 cell line (ATCC#CCL102) was cultured at 24°C in 0.6× L15 medium supplemented with 5% FCS and 1% penicillin/streptomycin (Gibco BRL), in a humidified air atmosphere. Cells were seeded at 60% confluency in 6-well plates, and transient transfection assays were carried out using Fugene6 as carrier (Roche Molecular Biochemicals). Luciferase (LuC) activity was assayed as recommended by the manufacturer (Promega) in a TD-20/20 luminometer (Turner Designs). Relative light units were normalized to protein concentration using the Coomassie dye binding assay (Pierce). All experiments were repeated at least five times. As a positive control for each transfection, the mouse *Osf2/Cbfa1* ORF and mouse *Osteocalcin* gene (mOG2)[11,36] were co-transfected in A6 cells.

Phylogenetic analysis of the duplicated runx2 isoforms

All complete protein sequences of Runt-related transcription factors (family HBG004221) were extracted from HOVERGEN[37] (http://pbil.univ-lyon1.fr/search/query_fam.php), aligned using MULTALIN,[38] and

cleaned through Gblocks.[39] The resulting alignment was examined using TREE_PUZZLE 5.2, under the JTT model, with one invariable class and eight γ rates with all parameters estimated from the data. Numbers associated with each branch represent support figures. The same alignment was examined using MrBayes (v3.0b4),[40] with a mixed amino acid model. MrBayes was executed for 1,000,000 generations resulting in 10,000 trees. A burn-in of 7000 trees was used. Support values indicate the proportion of the 3000 remaining trees supporting a given node under the 50% majority consensus rule.

RESULTS

Cloning and analysis of the zebrafish pebp2αA2 cDNA

Using RT-PCR with a combination of degenerate and specific primers and 5′RACE amplification, we cloned a zebrafish *runx2* isoform (*pebp2αA2*) cDNA (accession AY176052) spanning 2490 bp corresponding to an 87-bp 5′UTR, a 1350-bp ORF, and a 1053-bp partial 3′ UTR.

Alignment of available or deduced complete Pebp2αA protein sequences from *D. rerio, T. rubripes,* and *H. sapiens,* of til-1 sequences of *D. rerio* and *H. sapiens* and of the partial pebp2αA sequence from the cartilaginous fish *Raja eglanteria* (Fig. 1), shows an overall 81% identity between pebp2αA2 and pebp2αA1, with strongest homologies found in the *runt* domain (97%), the PST (proline/serine/threonine-rich) region (80%), the putative nuclear localization signal (NLS; 78%), and the putative C-terminal repression motif (100%). Comparative analysis of all Pebp2αA protein sequences from bony fish and *H. sapiens* showed an overall identity of 64%, again focused in the *runt* domain (96%), the PST region (60%), the presumptive NLS repression domain (78%), and the C-terminal VWRPY repression motif (100%), found almost invariably in all known Runt proteins. Interestingly, the glutamine/alanine (Q/A) repeat region was absent in all fish pebp2αA sequences, whereas it seems to be a characteristic of higher vertebrate sequences (from chicken to humans).

Determination of the zebrafish runx2b gene structure

The structure of the *runx2b* gene region that generates the *pebp2αA2* isoform mRNA (Fig. 2) has eight exons (Table 2), and all intronic/exonic borders were well conserved in comparison with mammals (data not shown). Introns 2 and 4–7 were not sequenced in full, nor were they available in the database of the zebrafish genome sequencing project. The high degree of conservation between zebrafish and human *Runx2* genomic sequences strongly suggests the existence, in zebrafish, of the same three runx2 isoforms, corresponding to types I to III, previously described in mammals. The isoforms starting with MASN and MSHSP, being the most N-terminal transcripts, could result from alternative splicing coupled with partial exon skipping, leading to the use of two different ATGs, located either in exon 1 (for the *Cbfa1*/*Osf* 2 transcript) or exon 2 (for the *Cbfa1* transcript; Fig. 2). Control of transcription is probably exerted by the distal promoter (P1). On the other hand, transcription of *pebp2αA2* is driven by a different promoter (P2), located within intron 2 (Fig. 2), its transcription start site, 5′UTR, and initiation codon (ATG3) being also located in this intron (within the ψ domain, see Fig. 2).

Cloning and analysis of the runx2b proximal promoter

Analysis of the 1795-bp fragment (GenBank AY509034) of the *runx2*b proximal promoter with TRANSFAC and MatInspector software revealed the existence of several putative binding motifs for nuclear proteins (Fig. 3A). We have not found a canonical TATA box, although it is possible that the motif TAATTT, located between –12 and –17 bp, could serve as a RNA polymerase II recognition site. Additionally, this promoter has a putative CAAT box located at –151 bp and among others, putative binding sites for CREB/ATF (–71, –178, and –1722), Runx2 (–598 and –913), and CBF1 (–104, –785, –1358, and –1386).

It has been shown that sequence comparison between teleost and mammalian genes can point to conserved regulatory regions.[41] We have performed a comparative sequence analysis of runx2b proximal promoters (P2) from zebrafish and mouse using the PipMaker software but failed to identify significantly conserved regions (results not shown).

Cloning and analysis of the zebrafish osteocalcin promoter

To address the question of whether the highly conserved Runx2 protein might be accompanied by conserved regulatory targets, we cloned 870 bp of the zebrafish *osteocalcin* 5′ flanking region (Fig. 3B). Sequence analysis shows the existence of a putative TATA box, located at –60 bp and a CAAT box, at –92 bp (numbered from the first nucleotide of the longest cDNA obtained; PY Gavaia and ML Cancela, unpublished data, 2001). Flanking the CAAT box, between –85 and –104 bp, a motif sharing a 70% conservation with human and rat *osteocalcin* boxes[42] was identified. In addition, several putative transcription factor binding motifs were also predicted with this analysis, namely for Runx2 (–187 to –201, –221 to –215, and –350 to –364), AP1 (–165 to –185 and –690 to –710), RXR (–654 to –638), CEBPβ (–389 to –376; –432 to –435; –613 to –626), and E4BP4 (–283 to –274). A repetitive motif ($[CA]_{n\ =\ 67}$) was also identified, located at –596 to –447.

Pebp2αA2 induces expression of the bone-related osteocalcin gene in A6 cells

Co-transfection of A6 cells with the zebrafish *pebp2αA2* ORF and o*steocalcin* promoter (Fig. 4A) showed a 2-fold induction of the o*steocalcin* promoter construct. The same result was obtained when the same o*steocalcin* promoter construct was co-transfected with the mouse pCMV-Osf2/Cbfa1 construct (Fig. 4A). In addition, the zebrafish *pebp2αA2* ORF was able to activate the mouse o*steocalcin* promoter to the same extent as the zebrafish o*steocalcin* promoter (Fig. 4A).

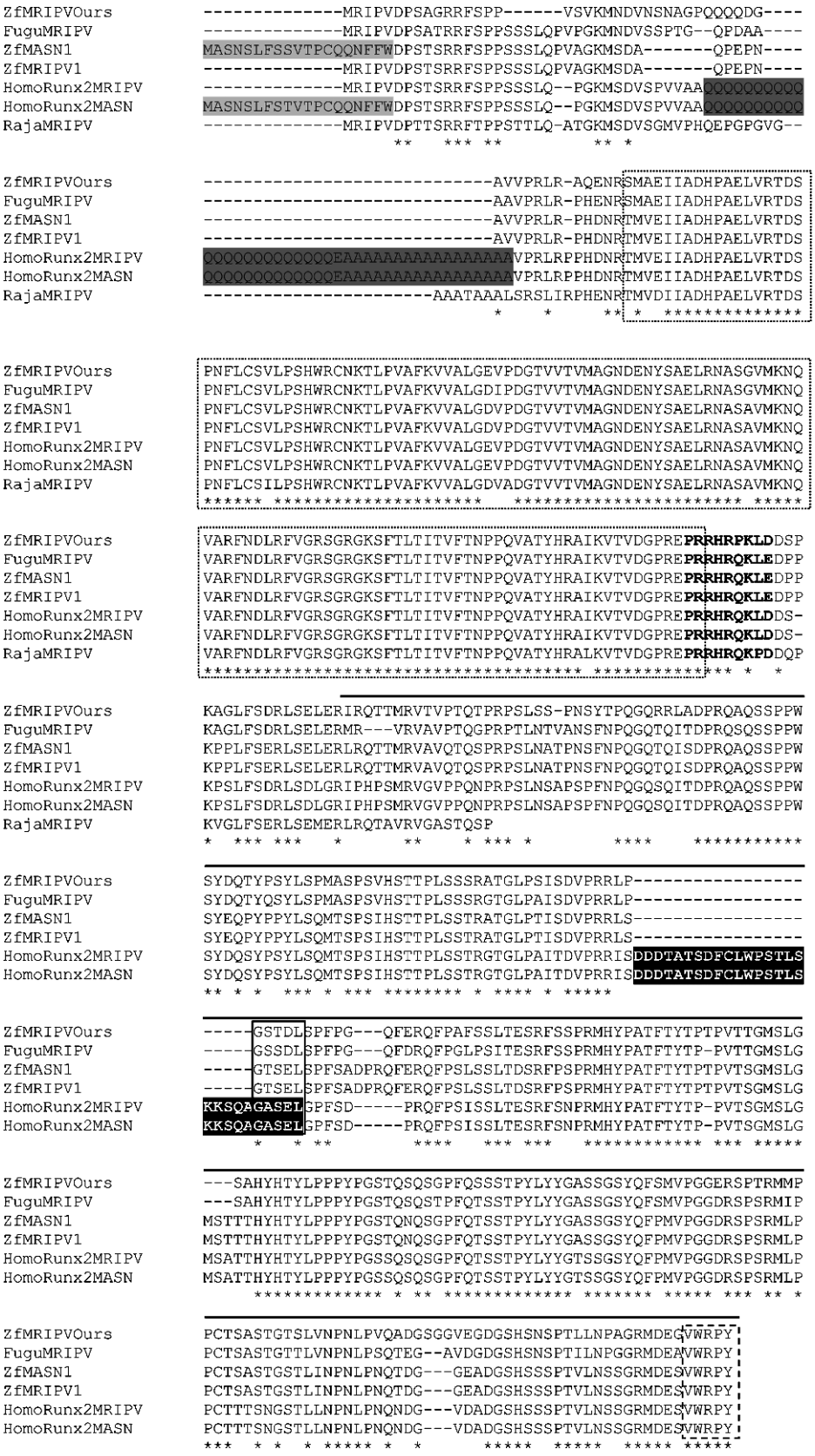

FIG. 1. Alignment of Pebp2αA protein sequences from *D. rerio* (pebp2αA2: AY176052 and pebp2αA1: AY380819), *T. rubripes* (deduced from the gene sequence AF477383), and *H. sapiens* (NM_004348), of the til-1 protein sequence of *D. rerio* (AY380820) and *H. sapiens* (AF001450) and of the partial pebp2αA sequence of *Raja eglanteria* (AY444494). Identical residues between all species are signaled with an asterisk. The activation domain 1 (AD1) of Cbfa1 is in light gray and the Q/A domain (AD2) of human RUNX2 isoforms is shaded in dark gray. The *runt* domain is dot-boxed and the proline/serine/threonine-rich (PST) region is indicated by a black line on top. The putative nuclear localization signal (NLS) is in bold, and the C-terminal 27 amino acids of AD3 are shaded in black with white letters. The "GASEL" motif, involved in the transactivation function of AD3, is boxed, and the "VWRPY" C-terminal repressor motif is dash-boxed.

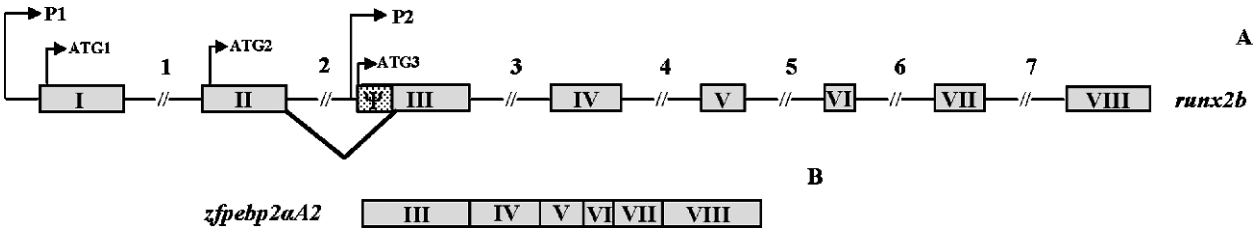

FIG. 2. Zebrafish *runx2b* gene organization and correspondence with *pebp2αA2* transcript. (A) The *runx2* gene consists of eight exons and two different alternative promoters (P1 and P2; not at scale). Introns are denoted by black lines (not at scale) and exons by gray boxes. (B) The *pebp2αA2* cDNA results from the transcription of exons III to VIII including a part of intron 2, denoted as Ψ. All exons, except exon VIII and Ψ, are at scale.

TABLE 2. zfrunx2b GENE ORGANIZATION AND EXON/INTRON SIZES

Exon	*Exon size (nt)*	*Intron*	*Intron size (nt)*
I	142	1	127
II	117	2	>5000
III	354	3	1386
IV	155	4	>1340
V	108	5	>478
VI	170	6	>657
VII	165	7	>9475
VIII	>1538		

nt, nucleotides.

In mammals, the Cbfa1 isoform has been shown to interact with the *osteocalcin* promoter, more specifically with two well-conserved elements, named OSE1 and OSE2.[36] Consequently, we analyzed the interaction between pebp2αA2 and the putative Runx2-binding elements (R2 and R3) identified in the zebrafish *osteocalcin* promoter. Site-directed mutagenesis of R2 (Fig. 4B), completely abolished the transactivation of the *osteocalcin* promoter by pebp2αA2 (Fig. 4A). Accordingly, no activation was seen when Wt3 or mutant R3 were used, indicating that presence of R2 is required for activation of transcription (Fig. 4A), and therefore R2 appears to be a target of pebp2αA2 in this assay system. Supporting this hypothesis is the fact that similar induction values were obtained when using the complete promoter (867 bp) or the nonmutated WT2 promoter fragment (226 bp), suggesting the absence of additional binding motifs for pebp2αA2 in the zebrafish *osteocalcin* promoter sequence analyzed.

Determination of the 3D structure of the zebrafish pebp2αA2 runt domain

The *runt* motif is the DNA-binding domain of all Runx proteins and thus is essential for protein function. Because one crystal structure has been published for a Runx protein, that of the mouse Runx1,[34] and given the high degree of sequence conservation (90% at the amino acid level) between the *runt* domain of the mouse Runx1 and that of the zebra fish *pebp2αA2,* we asked whether their 3D structure was also conserved between fish and mammals. The mouse Runx1 *runt* domain was thus used as template for the 3D modeling of the zebrafish protein runt domain. The model obtained for the pebp2αA2 *runt* motif displays a high degree of similarity with the mouse Runx1 equivalent, being made of 12 β strands separated by flexible loops, adopting the fold of an immunoglobulin (Ig)-like β sandwich,[43] similar to that shown by the crystallographic structures of *runt* domains bound to DNA and to Cbfβ (Fig. 5A).[44–46] The amino acid substitutions between mouse and zebrafish runt domains occur in β2 (S to C), L2 (T to S), L3 (S to C and I to V), L4 (D to E), β5 (L to V), β7 (T to S, A to G, and A to V), and β12 (I to V). However, none of these differences in amino acid residues are located in regions described as essential for DNA binding (Figs. 5B and 5C).

Phylogenetic analysis of the duplicated runx2 isoforms

The amino acid sequences of the zebrafish duplicated pebp2αA proteins were aligned with all known Runx proteins, and their phylogenetic relationships were analyzed (Figs. 6A and B). The results clearly showed a clustering of both Runx2b and Runx2a isoforms with other Runx2 proteins, separate from Runx1 and Runx3. However, numerous features of the topology of the phylogeny are sensitive to analysis methods. Notably, the precise relationship between the three family groups is unclear, and whereas Runx2b segregates with the Runx2 orthologs of other fish species, Runx2a location is unclear. The Bayesian analysis (Fig. 6B) fails to adequately resolve the node, whereas the quartet puzzling method provides weak support for it being closer to the mammalian and bird orthologs. It is safest to conclude that the duplication postdates the expansion of the three families, but is otherwise hard to resolve. Furthermore, it does not seem to be a recent *D. rerio*–specific duplication.

DISCUSSION

The zebrafish *D. rerio* is widely accepted as a valuable model for studies of vertebrate development[47,48] and an emerging model system for human disease,[49] presenting unique advantages over other vertebrate models. The Runx2 transcription factor is essential for osteoblastic differentiation and transcription of osteoblastic marker genes such as *osteocalcin*. However, these studies have been markedly biased toward higher vertebrates and the osteoblast-specific Cbfa1 isoform. Thus, other models that could contribute to understand the physiological role and evolutionary path of this gene and its multiple transcripts remain

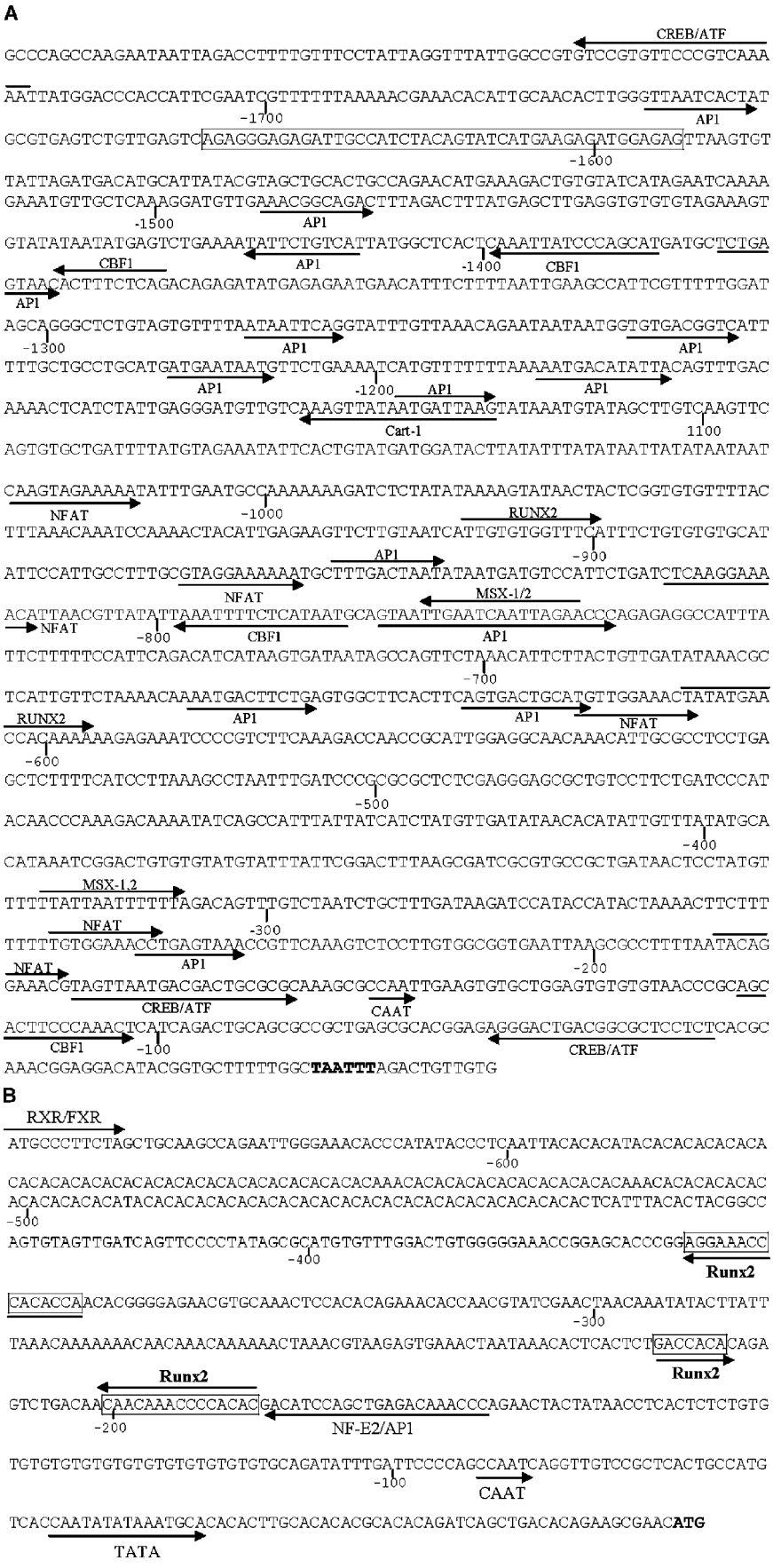

FIG. 3. Sequence analysis of (A) *runx2b* proximal promoter and (B) *osteocalcin* promoter. Several putative binding motifs were identified by MatInspector v.2.2 and Genomatix analysis and signaled in the sequence with solid arrows. A purine-rich region (boxed) was also identified in the *runx2b* promoter. Numbering from the ATG of the longest cDNA obtained is shown throughout the sequences.

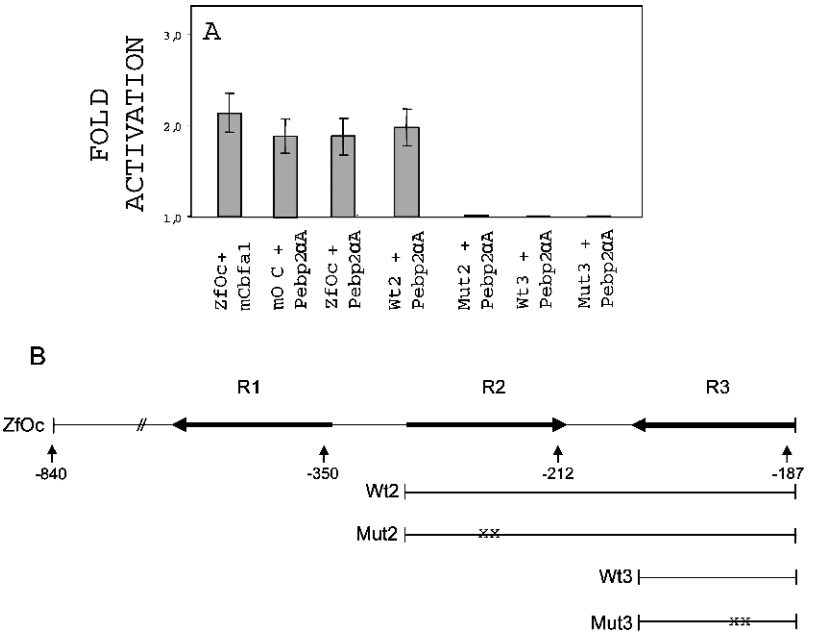

FIG. 4. Characterization of zebrafish *osteocalcin* promoter regulation. (A) Functional characterization of the *osteocalcin* promoter. Transfections were performed in A6 cells plated in 6-well plates and transiently cotransfected with either zfpebp2αA2ORF-pCMX-PL1 or mouse *Osf2/Cbfa1* ORF expression vectors together with either zebrafish *osteocalcin* (ZfOc) promoter or Wt2 (containing motifs R2 and R3) fused to the luciferase reporter gene. Mutation of motif R2 (Mut2) inhibits ZfOc promoter activation by zfpebp2αA2. Presence of either intact or mutated motif R3 (Wt3 vs. Mut3) show no effect on ZfOc promoter activation by zfpebp2αA2. (B) Schematic representation of the zebrafish *osteocalcin* promoter with the three putative runx2-binding elements. Localization of each motif within the *osteocalcin* promoter is shown by horizontal arrows, numbered from the initiation codon (see Fig. 3B). Nucleotides mutated by site-directed mutagenesis are identified by an "x".

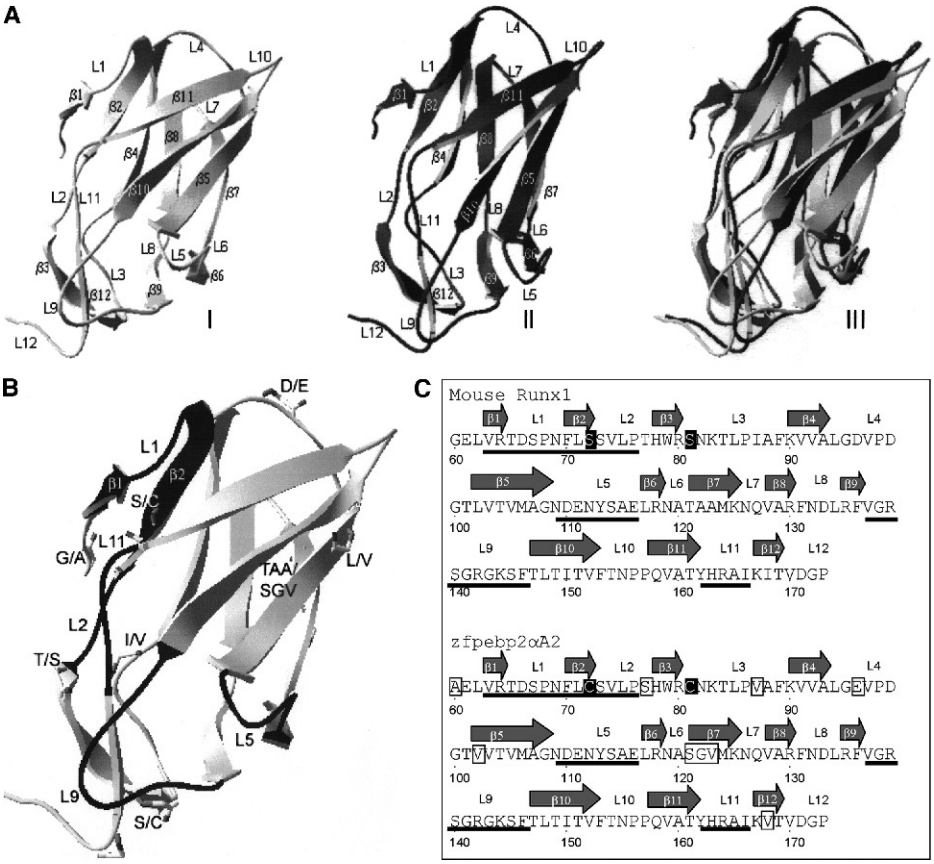

FIG. 5. 3D modeling of the pebp2αA2 *runt* domain. (A) Comparison with murine Runx1 *runt* domain. (I) Ribbon diagram of the mouse Runx1 *runt* domain structure (1EAN; amino acids 60–173).[34] (II) Ribbon diagram of the structure predicted for pebp2αA2 *runt* domain (residues Ala60 to Pro173). (III) Overlap of the pebp2αA2 and mRunx1 *runt* structures (pebp2αA2 *runt* in dark gray). The β strands and loops in I and II are numbered according to the mRunx1 model derived from its crystal structure and are labeled sequentially, from β1 to β12 and L1 to L12, respectively. (B) Ribbon diagram of the mRunx1 *runt* domain (1EAN; amino acids 60–173),[34] showing amino acid substitutions relatively to the pebp2αA2 DNA-binding motif. Protein regions intervening in DNA-binding are identified and shown in black. (C) Amino acid sequence and secondary structure of mouse Runx1 (1EAN; amino acids 60–173)[34] and zebrafish pebp2αA2 *runt* domains. Arrows denote the β-strands formed by each amino acid group, and each loop is identified by an L and the respective number. Amino acids are numbered below the sequence and amino acid substitutions between mouse and zebrafish are boxed. Protein domains shown to be critical for DNA/Cbfβ binding are signaled with a black bar. Amino acids shaded in black correspond to residues changed by engineered mutation to facilitate structure determination.[34]

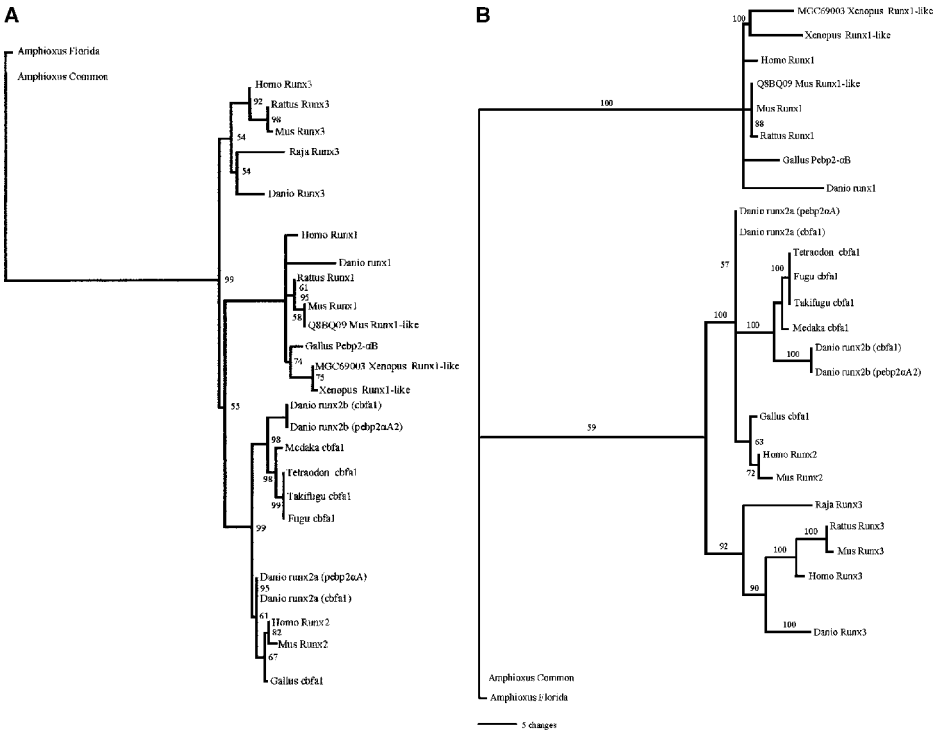

FIG. 6. Phylogenetic tree of all complete Runx type proteins in vertebrates. (A) Analysis using Tree-Puzzle. The numbers associated with nodes indicate support values derived from Tree-puzzle. (B) Analysis using MrBayes. The numbers associated with nodes indicate support values derived from 50% majority consensus rule applied to 3000 equally likely trees.

to be explored. Within this context, and in line of our previous work focused on teleost fish as models for analyzing expression of bone- and cartilage-related genes and skeletal development,[50–54] we hereby report the cloning, promoter analysis, and functional study of the *runx2b* gene and its pebp2αA2 isoform from the zebrafish *D. rerio*.

pebp2αA2 isoform shows all the characteristics of lower vertebrate orthologs

The high degree of homology between *pebp2αA2* cDNA and *Pebp2αA* from other species (higher and lower vertebrates) allowed us to predict the existence, in the hereby identified protein, of particular regions (Fig. 1), previously described as important for transcriptional activation of target genes in other species: (1) a PST region, containing the putative nuclear localization signal, along with an activator/repressor domain, shown to be partially responsible for the transactivation function of Runx2/Cbfa1 protein[23,55]; (2) a *runt* domain, encompassing the DNA binding and heterodimerization domains[1,20,56,57]; (3) the two cysteine residues, important for the redox regulation of Pebp2αA DNA binding[58]; and (4) a VWRPY motif, located at the C-terminus of the protein, which can act as inhibitor of the transactivation function of Runx2/Cbfa1, possibly by interacting with mammalian homologs of Groucho, such as TLE2.[23] This motif was previously described in *Drosophila* runt, where it also acts as a repressor of transcription[59] and more recently has been shown to be involved in the regulation of normal osteoblast proliferation.[60] The striking conservation of all four domains throughout evolution is strongly suggestive of their importance for protein function.

In contrast, a relatively low evolutionary conservation was observed for motifs important for Runx2 function in higher vertebrates: (1) the GASEL motif, which participates in the transactivation function of AD3[23] (only two in five amino acids conserved between human and zebrafish); and (2) the C-terminal 27 amino acids of AD3, required for optimum transactivation in mammals[61] and absent in *pebp2αA* from lower vertebrates. Finally, it is apparent that the Q/A domain is characteristic of higher vertebrates (Fig. 2), the highest known degree of expansion being observed in mouse, with 29Q and 18A. The Q/A domain was shown[23] to hinder heterodimerization of Runx2 proteins with Cbfβ, an ubiquitously expressed nuclear factor,[25,62] which increases affinity of RUNX1 to DNA.[63] In addition, this domain acts equally as an activation domain[23] in the context of the native protein. The presence of only a single Q/A sequence in zebrafish in the position of the expanded Q/A repeat of mammals suggests either the absence of interactions between Runx2 proteins and Cbfβ in this group or the possibility that Q/A transactivation function may be accomplished here by different regions of the protein, such as the PST domain, as was suggested for RUNX1.[64] Additionally, the fact that this domain is equally absent from all known Runx1 and Runx3 proteins suggests that it was absent from the ancestral gene, being an acquisition of the Runx2 proteins in higher vertebrates.

Gene structure of runx2b is highly conserved relative to mammalian orthologs

During the preparation of this manuscript, a paper describing the cloning and tissue expression of duplicated zebrafish *pebp2αA* cDNA was published.[29] The reported

isoform, encoded by the gene located on chromosome 20, seems to correspond to the one we describe here, which we have also mapped to LG 20 (results not shown). However a difference of 19 amino acids (4%) between the two ORF sequences was observed, and further comparisons are needed to identify which are, for example, strain–specific sequence differences. The *runx2* isoform described in this paper was called pebp2αA2 and, accordingly, the *RUNX2* ortholog from which *pebp2αA2* is transcribed was named *runx2b*.

The *runx2b* gene structure is highly similar to that described for its mammalian[61,65] and Fugu[28] orthologs (Fig. 2). From the three putative transcription initiation sites, only ATG1 remains to be tested, the activity of ATG2 and ATG3 having been established.[29] Therefore, the available evidence indicates that the plasticity of the *Runx2* gene has been conserved during evolution. Similarly, the regulation of expression of *Runx* genes by two distinct promoters seems to be the rule in all vertebrates.[61,66–69] These striking similarities strongly suggest that a comparable structure/function was already present in the *runx2* gene(s) of the ancestor vertebrate and are consistent with the essential role of the extant *Runx2* gene products in a diversity of biological functions.[70–72]

runx2b and Runx2 promoters show conservation of motifs

Comparison of the *runx2b* proximal promoter sequence to that of mouse *Runx2* showed some common consensus binding motifs, namely for NFAT, AP1, Runx2, GATA1, MSX-1,2, DLX1, and Cart-1. Of these, it has been shown that Runx2[24] and Msx-2[73] are able to regulate the expression of Runx2/Cbfa1 in mammals. They are thus strong candidates for regulating expression of *pebp2αA2* isoform in the same way. Additionally, the purine-rich region identified between –1594 and –1637 in the proximal *runx2b* promoter was also detected in the rat promoter,[24] suggesting a possible function of this region.

Zebrafish osteocalcin promoter shows motifs conserved with mammalian orthologs

The zebrafish *osteocalcin* promoter presents several motifs similar to mammalian *osteocalcin* promoters (Fig. 3B). Of these, the three putative runx2-binding sites are highly conserved compared with the mouse OSE motifs, strongly suggesting a role for runx2 in the transcriptional regulation of this bone-specific gene. Putative VDR and RXR binding sites in the zebrafish *osteocalcin* promoter, similarly to what we have described for other lower vertebrates,[51] suggests a transcriptional regulation of zebrafish *osteocalcin* by these molecules. The same holds true for the estrogen- and PTH-related elements (CEBPβ and E4BP4, respectively) identified, consistent with the described regulatory action of these molecules on osteocalcin expression.[74] In 1989, Lian et al.[42] identified a 24 nucleotide motif in the promoter of the rat *osteocalcin* involved in the efficient transcription of this gene. They named it *osteocalcin* box (*Oc* box) and it was later found in promoters of other mammalian *osteocalcin* genes. The –85 to –104 motif identified in the promoter of zf *osteocalcin* is the first reported motif in a lower vertebrate sharing high homology with the *Oc* box, suggesting that this sequence may be important in the regulation of zebrafish *osteocalcin*. Although functional studies need to be performed, the presence of these elements led us to hypothesize that the zebrafish *osteocalcin* gene may be under similar regulatory constraints as its mammalian orthologs, suggesting a conservation of these mechanisms over >200 million years of evolution.

Ability of pebp2αA2 to induce expression of the osteoblast-specific osteocalcin gene

To date, no information is available concerning the ability of this transcription factor to regulate *osteocalcin* gene transcription in lower vertebrates, despite the putative presence of OSE elements.[51,75] In this work, we show that pebp2αA2, one of the isoforms of *runx2b*, is able to direct a 2-fold activation of a zebrafish-*osteocalcin* promoter-luciferase chimeric gene (Fig. 4A), a level of induction similar to that reported in other models.[15,21,76] Runx2-induced *osteocalcin* gene transcription in mammals has been shown to be mediated by protein–DNA interactions located at two OSE elements, designated OSE1 and OSE2, respectively.[36] We show that induction of the zebrafish *osteocalcin* promoter by pebp2αA2 is completely abolished after site-directed mutagenesis of motif R2 (Figs. 4A and B), implicating this 15nt motif in the pebp2αA2-mediated transactivation of *osteocalcin* gene transcription. Ducy and Karsenty[36] described a similar element in the promoter of the mouse *osteocalcin* gene that binds an ubiquitously expressed factor, which they propose to be a basic helix-loop-helix protein. We suggest that, at least in zebrafish, this factor is pebp2αA2. Moreover, we found no evidence of interaction between pebp2αA2 and motifs R1 and R3, which may correspond, in view of their location, respectively, to the OSE1 and OSE2 motifs described in the mouse gene promoter (data not shown). This is in agreement with the restricted use of OSE1 and OSE2 by the osteoblast-specific isoforms of *Runx2*. Moreover, the presence of similar induction values observed with either the 867nt (full promoter) or the 226nt (WT2) construct exclude the existence of any extra pebp2αA2 binding sites 5′ to R2 within the DNA fragment analyzed. We show, additionally, that pebp2αA2 is able to activate the mouse *osteocalcin* promoter, although it remains to be shown whether it interacts specifically with the OSE motifs previously identified or with other regions of the mouse promoter.

Determination of the 3D structure of the zebrafish pebp2αA2 runt domain

The high degree of sequence conservation and the ability of zebrafish pebp2αA2 to bind the mouse *osteocalcin* promoter prompted us to compare the 3D structure of this zebrafish runx2-encoded isoform and mouse Runx1 (the only runt motif with a published structure) and study if the amino acid differences observed between the two *runt* domains could induce significant protein structural changes. Modeling results show that amino acid differences observed in the zebrafish DNA-binding motif do not induce major

structural changes relatively to the murine model (Fig. 5A). Amino acids known to play a key role in Runx protein function because of their participation in DNA binding are Arg80, Lys83, Arg135, Arg139, Arg142, Gly143, Lys167, Val170, Asp171, and Arg174,[45,46] whereas residues involved in Runx/Cbfβ interactions are Asp66, Pro68, Asn69, Met106, Ala107, Tyr113, Ser114, Thr149, Thr151, Phe153, Pro156, Pro157, Gln158, Val159, and His163.[45] Additionally, other amino acids involved in Runx2 dysfunction or clinical disease include Arg80, Lys83, Arg135, Arg139, Ser140, and Arg142.[77,78] None of these amino acids are changed in pebp2αA2, which probably reflects their importance for the correct structure and function of this protein. The only significant amino acid change is located on the β7 strand (Figs. 5B and 5C). However, with the exception of the Ala123Val change, all amino acid substitutions are conservative and do not produce a significant conformation shift in this region of the pebp2αA2 protein (Fig. 5A). This Ala123Val conformational change is not expected to influence the DNA/protein-binding capabilities of pebp2αA2, because this strand is reported not to be directly involved in DNA binding.[45] Similarly, the conformational change observed in the DNA binding loop L9, known to suffer a large structural change after complex formation, is not expected to affect protein function, because it is possible to observe a variety of conformations on this loop in the absence of DNA,[33] a fact that has been attributed to its glycine residues.[79]

Overall, results obtained with the modeling of the zebrafish pebp2αA2 *runt* domain lead to the conclusion that substitutions in this domain are strictly controlled, being only allowed in non-DNA/protein-binding regions. The single DNA-binding motif where substitutions seem to be under a more loose control is loop 9, a region where structural variation is allowed. The high structural similarity between the zebrafish and mouse *runt* domains explains the ability of the zebrafish pebp2αA2 to induce expression of the mouse *osteocalcin* gene and highlights the striking structural and functional evolutionary conservation of these proteins.

Phylogenetic analysis of runx2b and pebp2αA2

The identification in zebrafish of the first duplication of a *Runx2* gene prompted us to perform an in-depth evolutionary analysis of these genes in this species. Phylogenetic analysis of all known Runx sequences from fish, *xenopus,* chicken, and mammals (Figs. 6A and 6B) indicated that both zebrafish *runx2* genes are true orthologs of mammalian Runx2, showing a robust separation from Runx1 and Runx3. The relationship between the two runx2 orthologs, however, is not straightforward, because their relative position is very sensitive to the methodology. As expected, the support values for the relevant nodes are very low. However, both approaches show clearly that the duplication of the runx2 ancestor gene occurred at or around the time of the divergence between teleost and the tetrapod ancestor, because the amount of divergence accumulated between the two genes places one of them (*runx2b*) consistently within the fish cluster and the other (*runx2a*), either in an independent cluster or in a cluster with a relatively recent common ancestor with mammalian and avian clusters. In both cases, the separation seems to have occurred either shortly before, during, or shortly after the divergence between the lines that originated fish and mammals. Both methodologies suggest that this duplication occurred before the divergence of the zebrafish ancestor from other teleosts.

In summary, we have isolated one of the copies of the duplicated zebrafish *runx2* gene and found that it has a structure similar to its previously described mammalian and Fugu orthologs, suggesting the capability of zebrafish *runx2* to code for three protein isoforms with different N-terminal regions.

The cDNA encoding *pebp2αA2* isoform was isolated and shown to be highly conserved between lower and higher vertebrates. We show for the first time that, in a lower vertebrate, *pebp2αA2* not only is competent for transcriptional regulation of the *osteocalcin gene*, through interaction with a well-defined binding motif in this promoter (Motif R2), but is equally able to induce transcription of a reporter gene controlled by the mammalian *Oc* promoter.

Modeling of the *runt* motif of zebrafish pebp2αA2 shows a remarkable similarity between the 3D structure of its DNA binding motif and that of mouse Runx1, indicative of an intense selective pressure against mutations in this motif, the hallmark of all Runx proteins.

Finally, we analyzed the evolutionary relationships of the two known *runx2 genes*. Both are unambiguously members of the Runx2 family. The duplication seems to have occurred at or around the time of the divergence between teleost and the tetrapod ancestor, although it proved impossible to further resolve the timing of this occurrence. Runx2b seems to be the ortholog of the *runx2* genes found in other fish. The high degree of conservation of the *runt* motif in the evolutionary scale allows the interspecific transcriptional regulation of the Runx2 proteins between teleosts and mammals, reflecting the importance of this particular DNA binding motif in the course of evolution.

ACKNOWLEDGMENTS

This work was supported in part by research funds from CCMAR and a research grant POCTI/BCI/48748/2002 from the Portuguese Science and Technology Foundation (FCT). Work was also supported by a Luso-British Collaborative Action (B-72/04) from CRUP, between MLC and RK. JPP and NC were the recipients of postdoctoral fellowships (PRAXIS XXI/BPD/20229/99 and SFRH/BPD/9451/2002, respectively) from FCT. RL was the recipient of a research fellowship from CCMAR.

REFERENCES

1. Kagoshima H, Shigesada K, Satake M, Ito Y, Miyoshi H, Ohki M, Pepling M, Gergen JP 1993 The Runt domain identifies a new family of heteromeric transcriptional regulators. Trends Genet **9:**338–341.
2. Coffman JA 2003 Runx transcription factors and the developmental balance between cell proliferation and differentiation. Cell Biol Int **27:**315–324.

3. Okuda T, van Deursen J, Hiebert SW, Grosveld G, Downing JR 1996 AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. Cell **84:**321–330.
4. Fukamachi H, Ito K 2004 Growth regulation of gastric epithelial cells by Runx3. Oncogene **23:**4330–4335.
5. Ku JL, Kang SB, Shin YK, Kang HC, Hong SH, Kim IJ, Shin JH, Han IO, Park JG 2004 Promoter hypermethylation downregulates RUNX3 gene expression in colorectal cancer cell lines. Oncogene **23:**6736–6742.
6. Komori T, Yagi H, Nomura S, Yamaguchi A, Sasaki K, Deguchi K, Shimizu Y, Bronson RT, Gao Y-H, Inada M, Sato M, Okamoto R, Kitamura Y, Yoshiki S, Kishimoto T 1997 Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. Cell **89:**755–764.
7. Otto F, Thornell AP, Crompton T, Denzel A, Gilmour KC, Rosewell IR, Stamp GWH, Beddington RSP, Mundlos S, Olsen BR, Selby PB, Owen MJ 1997 Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. Cell **89:**765–771.
8. Gao YH, Shinki T, Yuasa T, Kataoka-Enomoto H, Komori T, Suda T, Yamaguchi A 1998 Potential role of Cbfa1, an essential transcriptional factor for osteoblast differentiation, in osteoclastogenesis: Regulation of mRNA expression of osteoclast differentiation factor (ODF). Biochem Biophys Res Commun **252:**697–702.
9. Liu W, Toyosawa S, Furuichi T, Kanatani N, Yoshida C, Liu Y, Himeno M, Narai S, Yamaguchi A, Komori T 2001 Overexpression of Cbfa1 in osteoblasts inhibits osteoblast maturation and causes osteopenia with multiple fractures. J Cell Biol **155:**157–166.
10. Takeda S, Bonnamy JP, Owen MJ, Ducy P, Karsenty G 2001 Continuous expression of Cbfa1 in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues Cbfa1-deficient mice. Genes Dev **15:**467–481.
11. Ducy P, Zhang R, Geoffroy V, Ridall AL, Karsenty G 1997 Osf2/Cbfa1: A transcriptional activator of osteoblast differentiation. Cell **89:**747–754.
12. Lee KS, Kim HJ, Li QL, Chi XZ, Ueta C, Komori T, Wozney JM, Kim EG, Choi JY, Ryoo HM, Bae SC 2000 Runx2 is a common target of transforming growth factor beta1 and bone morphogenetic protein 2, and cooperation between Runx2 and Smad5 induces osteoblast-specific gene expression in the pluripotent mesenchymal precursor cell line C2C12. Mol Cell Biol **20:**8783–8792.
13. Zheng Q, Zhou G, Morello R, Chen Y, Garcia-Rojas X, Lee B 2003 Type X collagen gene regulation by Runx2 contributes directly to its hypertrophic chondrocyte-specific expression in vivo. J Cell Biol **162:**833–842.
14. Inman CK, Shore P 2003 The osteoblast transcription factor Runx2 is expressed in mammary epithelial cells and mediates osteopontin expression. J Biol Chem **278:**48684–48689.
15. Banerjee C, Javed A, Choi J-Y, Green J, Rosen V, van Wijnen A, Stein J, Lian J, Stein G 2001 Differential regulation of the two principal Runx2/Cbfa1 N-terminal isoforms in response to bone morphogenetic protein-2 during development of the osteoblast phenotype. Endocrinology **142:**4026–4039.
16. Sevetson B, Taylor S, Pan Y 2004 Cbfa1/RUNX2 directs specific expression of the sclerostenosis gene (SOST). J Biol Chem **279:**13849–13858.
17. Zelzer E, Glotzer DJ, Hartmann C, Thomas D, Fukai N, Soker S, Olsen BR 2001 Tissue specifc regulation of VEGF expression during bone development requires Cbfa1/Runx2. Mech Dev **106:**97–106.
18. McCarthy TL, Chang WZ, Liu Y, Centrella M 2003 Runx2 integrates estrogen activity in osteoblasts. J Biol Chem **278:**43121–43129.
19. Jiménez M, Balbín M, López J, Alvarez J, Komori T, López-Otín C 1999 Collagenase 3 is a target of Cbfa1, a transcription factor of the *runt* gene family involved in bone formation. Mol Cell Biol **19:**4431–4442.
20. Ogawa E, Maruyama M, Kagoshima H, Inuzuka M, Lu J, Satake M, Shigesada K, Ito Y 1993 PEBP2/PEA2 represents a family of transcription factors homologous to the products of the Drosophila runt gene and the human AML1 gene. Proc Natl Acad Sci USA **90:**6859–6863.
21. Harada H, Tagashira S, Fujiwara M, Ogawa S, Katsumata T, Yamaguchi A, Komori T, Nakatsuka M 1999 Cbfa1 isoforms exert functional differences in osteoblast differentiation. J Biol Chem **274:**6972–6978.
22. Stewart M, Terry A, Hu M, O'Hara M, Blyth K, Baxter E, Cameron T, Onions DE, Neil JC 1997 Proviral insertions induce the expression of bone-specific isoforms of PEBP2alphaA (CBFA1): Evidence for a new myc collaborating oncogene. Proc Natl Acad Sci USA **94:**8646–8651.
23. Thirunavukkarasu K, Mahajan M, McLarren KW, Stifani S, Karsenty G 1998 Two domains unique to osteoblast-specific transcription factor Osf2/Cbfa1 contribute to its transactivation function and its inability to heterodimerize with Cbfβ. Mol Cell Biol **18:**4197–4208.
24. Drissi H, Luc Q, Shakoori R, Chuva De Sousa Lopes S, Choi JY, Terry A, Hu M, Jones S, Neil JC, Lian JB, Stein JL, Van Wijnen AJ, Stein GS 2000 Transcriptional autoregulation of the bone related CBFA1/RUNX2 gene. J Cell Physiol **184:**341–350.
25. Ogawa E, Inuzuka M, Maruyama M, Satake M, Naito-Fujimoto M, Ito Y, Shigesada K 1993 Molecular cloning and characterization of PEBP2 beta, the heterodimeric partner of a novel Drosophila runt-related DNA binding protein PEBP2 alpha. Virology **194:**314–331.
26. Sakakura C, Yamaguchi-Iwai Y, Satake M, Bae SC, Takahashi A, Ogawa E, Hagiwara A, Takahashi T, Murakami A, Makino K, Nakagawa T, Kamada N, Ito Y 1994 Growth inhibition and induction of differentiation of t(8;21) acute myeloid leukemia cells by the DNA-binding domain of PEBP2 and the AML1/MTG8(ETO)-specific antisense oligonucleotide. Proc Natl Acad Sci USA **91:**11723–11727.
27. Inohaya K, Kudo A 2000 Temporal and spatial patterns of *cbfal* expression during embryonic development in the teleost *Oryzias latipes*. Dev Genes Evol **210:**570–574.
28. Eggers JH, Stock M, Fliegauf M, Vonderstrass B, Otto F 2002 Genomic characterization of the RUNX2 gene of Fugu rubripes. Gene **291:**159–167.
29. Flores MV, Tsang VW, Hu W, Kalev-Zylinska M, Postlethwait J, Crosier P, Crosier K, Fisher S 2004 Duplicate zebrafish runx2 orthologues are expressed in developing skeletal elements. Gene Expr Patterns **4:**573–581.
30. Thompson JD, Higgins DG, Gibson TJ 1994 CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res **22:**4673–4680.
31. Quandt K, Frech K, Karas H, Wingender E, Werner T 1995 MatInd and MatInspector - New fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Res **23:**4878–4884.
32. Schwartz S, Zhang Z, Frazer KA, Smit A, Riemer C, Bouck J, Gibbs R, Hardison R, Miller W 2000 PipMaker—a web server for aligning two genomic DNA sequences. Genome Res **10:**577–586.
33. Sali A, Blundell TL 1993 Comparative protein modelling by satisfaction of special restrains. J Mol Biol **234:**779–815.
34. Backstrom S, Wolf-Watz M, Grundstrom C, Hard T, Grundstrom T, Sauer UH 2002 The RUNX1 Runt domain at 1.25A resolution: A structural switch and specifically bound chloride ions modulate DNA binding. J Mol Biol **322:**259–272.
35. Laskowski RA, MacArthur MW, Moss DS, Thorton JM 1993 PROCHECK: A program to check the stereochemical quality of protein structures. J Appl Crystallogr **26:**283–291.
36. Ducy P, Karsenty G 1995 Two distinct osteoblast-specific *cis*-acting elements control expression of a mouse osteocalcin gene. Mol Cell Biol **15:**1858–1869.

37. Duret L, Mouchiroud D, Gouy M 1994 HOVERGEN: A database of homologous vertebrate genes. Nucleic Acids Res **22:**2360–2365.
38. Corpet F 1988 Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res **16:**10881–10890.
39. Castresana J 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol **17:**540–552.
40. Huelsenbeck JP, Ronquist F 2001 MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
41. Aparicio S, Morrison A, Gould A, Gilthorpe J, Chaudhuri C, Rigby P, Krumlauf R, Brenner S 1995 Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes.* Proc Natl Acad Sci USA **92:**1684–1688.
42. Lian J, Stewart C, Puchacz E, Mackowiak S, Shalhoub V, Collart D, Zambetti G, Stein G 1989 Structure of the rat osteocalcin gene and regulation of vitamin D-dependent expression. Proc Natl Acad Sci USA **86:**1143–1147.
43. Bork P, Holm L, Sander C 1994 The immunoglobulin fold. Structural classification, sequence patterns and common core. J Mol Biol **242:**309–320.
44. Warren AJ, Bravo J, Williams RL, Rabbitts TH 2000 Structural basis for the heterodimeric interaction between the acute leukaemia-associated transcription factors AML1 and CBF-beta. EMBO J **19:**3004–3015.
45. Tahirov TH, Inoue-Bungo T, Morii H, Fujikawa A, Sasaki M, Kimura K, Shiina M, Sato K, Kumasaka T, Yamamoto M, Ishii S, Ogata K 2001 Structural analyses of DNA recognition by the AML1/Runx-1 Runt domain and its allosteric control by CBF-beta. Cell **104:**755–767.
46. Bravo J, Li Z, Speck NA, Warren AJ 2001 The leukemia-associated AML1 (Runx1)-CBF beta complex functions as a DNA-induced molecular clamp. Nat Struct Biol **8:**371–378.
47. Kimmel CB 1989 Genetics and early development of zebrafish. Trends Genet **5:**283–288.
48. Driever W, Solnica-Krezel L, Schier AF, Neuhauss SC, Malicki J, Stemple DL, Stainier DY, Zwartkruis F, Abdelilah S, Rangini Z 1996 A genetic screen for mutations affecting embryogenesis in zebrafish. Development **123:**37–46.
49. Zon LI 1999 Zebrafish: A new model for human disease. Genome Res **9:**99–100.
50. Cancela ML, Williamson MK, Price PA 1995 Amino-acid sequence of bone Gla protein from the African clawed toad *Xenopus laevis* and the fish *Sparus aurata*. Int J Pept Protein Res **46:**419–423.
51. Pinto JP, Ohresser M, Cancela ML 2001 Cloning of the bone Gla protein gene from the teleost fish *Sparus aurata*. Evidence for overall conservation in molecular structure and pattern of expression from fish to man. Gene **270:**77–91.
52. Pinto JP, Conceição N, Gavaia PJ, Cancela ML 2003 Matrix Gla protein gene expression and protein accumulation colocalize with cartilage distribution during development of the teleost fish *Sparus aurata*. Bone **32:**201–210.
53. Simes DC, Williamson MK, Ortiz-Delgado JB, Viegas CS, Price PA, Cancela ML 2003 Purification of matrix Gla protein from a marine teleost fish, *Argyrosomus regius*: Calcified cartilage and not bone as the primary site of MGP accumulation in fish. J Bone Miner Res **18:**244–259.
54. Pombinho AR, Laize V, Molha DM, Marques SM, Cancela ML 2004 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:**393–406.
55. Cui CB, Cooper LF, Yang X, Karsenty G, Aukhil I 2003 Transcriptional coactivation of bone-specific transcription factor Cbfa1 by TAZ. Mol Cell Biol **23:**1004–1013.
56. Imai Y, Kurokawa M, Izutsu K, Hangaishi A, Takeuchi K, Maki K, Ogawa S, Chiba S, Mitani K, Hirai H 2000 Mutations of the AML1 gene in myelodysplastic syndrome and their functional implications in leukemogenesis. Blood **96:**3154–3160.
57. Michaud J, Wu F, Osato M, Cottles GM, Yanagida M, Asou N, Shigesada K, Ito Y, Benson KF, Raskind WH, Rossier C, Antonarakis SE, Israels S, McNicol A, Weiss H, Horwitz M, Scott HS 2002 In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: Implications for mechanisms of pathogenesis. Blood **99:**1364–1372.
58. Akamatsu Y, Ohno T, Hirota K, Kagoshima H, Yodoi J, Shigesada K 1997 Redox regulation of the DNA binding activity in transcription factor PEBP2. The roles of two conserved cysteine residues. J Biol Chem **272:**14497–14500.
59. Aronson BD, Fisher AL, Blechman K, Caudy M, Gergen JP 1997 Groucho-dependent and -independent repression activities of Runt domain proteins. Mol Cell Biol **17:**5581–5587.
60. Pratap J, Galindo M, Zaidi SK, Vradii D, Bhat BM, Robinson JA, Choi JY, Komori T, Stein JL, Lian JB, Stein GS, van Wijnen AJ 2003 Cell growth regulatory role of Runx2 during proliferative expansion of preosteoblasts. Cancer Res **63:**5357–5362.
61. Geoffroy V, Corral DA, Zhou L, Lee B, Karsenty G 1998 Genomic organization, expression of the human *CBFA1* gene, and evidence for an alternative splicing event affecting protein function. Mamm Genome **9:**54–57.
62. Golling G, Li L-H, Pepling M, Stebbins M, Gergen JP 1996 Drosophila homologs of the proto-oncogene product PEBP2/CBF beta regulate the DNA-binding properties of Runt. Mol Cell Biol **16:**932–942.
63. Ito Y, Bae S-C 1997 The runt Domain Transcription Factor PEBP2/CBF, and Its Involvement in Human Leukemia. Birkhauser Verlag, Basel, Switzerland.
64. Bae SC, Ogawa E, Maruyama M, Oka H, Satake M, Shigesada K, Jenkins NA, Gilbert DJ, Copeland NG, Ito Y 1994 PEBP2 alpha B/mouse AML1 consists of multiple isoforms that possess differential transactivation potentials. Mol Cell Biol **14:**3242–3252.
65. Xiao ZS, Thomas R, Hinson TK, Quarles LD 1998 Genomic structure and isoform expression of the mouse, rat and human Cbfa1/Osf2 transcription factor. Gene **214:**187–197.
66. Levanon D, Bernstein Y, Negreanu V, Ghozi MC, Baram I, Aloya R, Goldenberg D, Lotem J, Groner Y 1996 A large variety of alternatively spliced and differentially expressed messenger RNAs are encoded by the human acute myeloid leukemia gene AML1. DNA Cell Biol **15:**175–185.
67. Rini D, Calabi F 2001 Identification and comparative analysis of a second runx3 promoter. Gene **273:**13–22.
68. Miyoshi H, Ohira M, Shimizu K, Mitani K, Hirai H, Imai T, Yokoyama K, Soeda E, Ohki M 1995 Alternative splicing and genomic structure of the AML1 gene involved in acute myeloid leukemia. Nucleic Acids Res **23:**2762–2769.
69. Ahn MY, Bae SC, Maruyama M, Ito Y 1996 Comparison of the human genomic structure of the runt domain-encoding PEBP2/CBFα gene family. Gene **168:**279–280.
70. Karsenty G 2000 Role of Cbfa1 in osteoblast differentiation and function. Cell Dev Biol **11:**343–346.
71. Blyth K, Terry A, Mackay N, Vaillant F, Bell M, Cameron ER, Neil JC, Stewart M 2001 Runx2: A novel oncogenic effector revealed by in vivo complementation and retroviral tagging. Oncogene **20:**295–302.
72. Cameron ER, Blyth K, Hanlon L, Kilbey A, Mackay N, Stewart M, Terry A, Vaillant F, Wotton S, Neil JC 2003 The Runx genes as dominant oncogenes. Blood Cells Mol Dis **30:**194–200.
73. Satokata I, Ma L, Ohshima H, Bei M, Woo I, Nishizawa K, Maeda T, Takano Y, Uchiyama M, Heaney S, Peters H, Tang Z, Maxson R, Maas R 2000 Msx2 deficiency in mice causes pleiotropic defects in bone growth and ectodermal organ formation. Nat Genet **24:**391–395.
74. Yu XP, Chandrasekhar S 1997 Parathyroid hormone (PTH 1-34) regulation of rat osteocalcin gene transcription. Endocrinology **138:**3085–3092.
75. Viegas CS, Pinto JP, Conceicao N, Simes DC, Cancela ML 2002 Cloning and characterization of the cDNA and gene encoding *Xenopus laevis* osteocalcin. Gene **289:**97–107.
76. Xiao ZS, Hinson TK, Quarles LD 1999 Cbfa1 isoform overex-

pression upregulates osteocalcin gene expression in non-osteoblastic and pre-osteoblastic cells. J Cell Biochem **74:**596–605.

77. Zhou G, Chen Y, Zhou L, Thirunavukkarasu K, Hecht J, Chitayat D, Gelb BD, Pirinen S, Berry SA, Greenberg CR, Karsenty G, Lee B 1999 CBFA1 mutation analysis and functional correlation with phenotypic variability in cleidocranial dysplasia. Hum Mol Genet **8:**2311–2316.
78. Quack I, Vonderstrass B, Stock M, Aylsworth AS, Becker A, Brueton L, Lee PJ, Majewski F, Mulliken JB, Suri M, Zenker M, Mundlos S, Otto F 1999 Mutation analysis of core binding factor A1 in patients with cleidocranial dysplasia. Am J Hum Genet **65:**1268–1278.
79. Bartfeld D, Shimon L, Couture GC, Rabinovich D, Frolow F, Levanon D, Groner Y, Shakked Z 2002 DNA recognition by the RUNX1 transcription factor is mediated by an allosteric transition in the RUNT domain and by DNA bending. Structure **10:**1395–1407.

Address reprint requests to:
ML Cancela, PhD
Center of Marine Sciences
University of Algarve
Campus de Gambelas
Faro 8005-139, Portugal
E-mail: lcancela@ualg.pt

Received in original form October 1, 2004; revised form January 19, 2005; accepted March 16, 2005.