Osteogenic Protein-2

A NEW MEMBER OF THE TRANSFORMING GROWTH FACTOR-β SUPERFAMILY EXPRESSED EARLY IN EMBRYOGENESIS*

Engin Özokaynak‡, Patrick N. J. Schnegelsberg§, Donald F. Jin, Gail Munro Clifford, Frederick D. Warren, Eric A. Drier, and Hermann Oppermann‡

From Creative BioMolecules, Inc., Hopkinton, Massachusetts 01748

Osteogenic protein-2, OP-2, a new member of the transforming growth factor-β (TGF-β) superfamily, closely related to the osteogenic/bone morphogenetic proteins, was discovered in mouse embryonic and human hippocampus cDNA libraries. The TGF-β domain of OP-2 shows 74% identity to OP-1, 75% to Vgr-1, and 76% to BMP-5, but in OP-2 may also have bone inductive activity. The genomic locus of OP-2 has seven exons, like OP-1, and spans more than 27 kilobases (kb). In the C-terminal TGF-β domain, OP-2 has a unique additional cysteine. Mouse embryos express relatively high levels of OP-2 mRNA at 8 days, two species of 3 and 5 kb. A careful study of mRNA expression of the osteogenic proteins in specific organs revealed discrete mRNA species for BMP-3, BMP-4, BMP-5, and BMP-6/Vgr-1 in lung or liver of young and adult mice. OP-1 is expressed in kidney; however, OP-2 and BMP-2 mRNAs were not detected in any organs studied, suggesting an early developmental role.

The extended TGF-β superfamily is widely represented in vertebrates and invertebrates. Examples among vertebrates are the Vg-1 gene product of Xenopus laevis (1), Mullerian inhibiting substance (MIS) (2), the inhibins and activins (3–5) and bone morphogenetic proteins (BMP-2, -3, -4, and -5) (6, 7), Vgr-1 (8), and OP-1 (9). More distantly related to these is GDF-1 (growth and differentiation factor-1) (10). Represented in invertebrates are the decapentaplegic protein (DPP) and 60A gene product of Drosophila melanogaster (11–15). Recently, receptors for activin and TGF-β were cloned and identified as serine threonine kinases (15–18). TGF-β-related proteins are secreted as a pro form, with the C-terminal domain yielding dimeric mature protein (16, 25). The C-terminal TGF-β domain is characterized by conserved residues including a pattern of 7 cysteines.

In their biological roles, these proteins act during embryogenesis and in adult animals. For example, the DPP gene product is a determinant of the dorsal-ventral pattern formation of the fly, and null mutations have been shown to be lethal (17). Inhibins and activins control secretion of follicle-stimulating hormone (3, 5). Moreover, activins A and B show mesoderm-inducing activity (18–20). BMP-4 also induced mesoderm in a similar assay (21). MIS causes regression of the Mullerian duct during male development (2, 22). Recombinant BMP-2 (23), BMP-4 (24), and OP-1 (25) induce new bone formation in a subcutaneous implant assay in rats (26). Recombinant OP-1 homodimers, produced in mammalian cell culture, have been effective in the repair of segmental bone defects in rabbit ulna (25, 27).

A major site of OP-1 expression was found to be the kidneys (28). Expression of a bone growth factor in the kidneys is consistent with their role in calcium regulation and bone homeostasis. The original source of osteogenic protein, the bone tissue, contains several TGF-β-related proteins as evidenced by protein sequence analysis of purified osteogenic activity (6, 29). The OP-1 gene and other BMPs were cloned using a consensus probe, based on partial protein sequence information and conserved elements of the TGF-β family (9). Screening of cDNA libraries with OP-1 probe led to discovery of OP-2, which is described here.

MATERIALS AND METHODS

Library Screening—All libraries were screened by an initial plating of 10^7 plaques (approximately 5 × 10^6 plaques/plate) and hybridizations were in 40% formamide, 5 × SSPE, 5 × Denhardt’s solution, and 0.1% SDS at 37 °C. Nonspecific counts were removed in 0.1 × SSPE, 0.1% SDS by shaking at 50 °C. Murine OP-2 cDNA was found in a 17-day mouse embryo 5′-stretch cDNA library (λ gt10, ML1029a, Clontech, Palo Alto, CA) and also in a teratocarcinoma (PCC4) cDNA library (ZAPII, 936301, Stratagene, La Jolla, CA). The complete human OP-2-coding sequence was derived from a cDNA clone in a human hippocampus cDNA library (λ ZAPII, 936303, Stratagene, from brain of a normal 2-year-old girl) and a genomic clone (EMBL-3 library, HL1067, Clontech). The latter library also yielded the other clones used to determine the human OP-2 genomic locus. The genomic structure of human OP-1 was determined using two different libraries (Clontech, HL1067, Stratagene, 940205). For identification of the splice junctions of OP-1 the following subclones were prepared. Exon 1 of OP-1, with both splice junctions, was subcloned as a 1.7-kb PstI fragment (pO142-12) and exon 2, with splice junctions, was cloned as a 0.6-kb HaeIII fragment (pO247-5). Part of the third exon and upstream junction was cloned as a 1.7-kb PstI fragment (pO243), whereas the downstream junction was cloned as a 0.3-kb Sau3AI fragment (pO248). Exon 4, with splice junctions, was subcloned as a 0.7-kb PstI fragment (pO229). Exons 4, 5, 6, and 7 were all present in the original OP-1 genomic clone (λ 13) (9).

DNA Sequencing—All sequencing was done according to Sanger et al. (30) using exonuclease III-mediated unidirectional deletion (31),
subcloning of restriction fragments, and synthetic primers. The coding regions of human and murine OP-2 DNA were sequenced on both strands. Compressions were resolved by performing reactions at 70 °C with Taq polymerase and using 7-deaza-GTP (U. S. Biochemical Corp.).

**Cloning of RNA and Northern Blot Analysis**—Mice, strain CD-1, and rats (Long-Evans) were from Charles River Laboratories, Wilmington, MA. Total RNA from mice was isolated by the acid guanidine thiocyanate-phenol-chloroform method (32). Poly(A)⁺ RNA was analyzed on 1.2% agarose-formaldehyde gels and blotted onto Nytran membranes (Schleicher & Schuell) with 10 × SSC (32). Hybridization conditions with 3P-labeled (33) probes were as described (28). Between hybridizations filters were depropped in 1 mm Tris-HCl, 1 mm EDTA, 0.1% SDS, pH 7.5, at 90–95 °C and exposed to film to assure complete removal of probe.

**Cloning of Vgr-1 and GDF-1 cDNA and BMP-5-specific cDNA Fragments**—A partial Vgr-1 clone, lacking some 300 nucleotides at the 5'-coding region was isolated from a mouse brain cDNA library (Clontech, ML1056a) using an OP-1 cDNA probe. A murine GDF-1 cDNA spanning the complete coding region was isolated by PCR. Oligo(dT) was used to prime first-strand synthesis from 200 ng of mouse brain poly(A)⁺ RNA. Reactions were in a 100-μl volume with 5 μl of 20 × PCR reaction buffer (1 M Tris-HCl, pH 9.0, 400 mM ammonium sulfate, 30 mM magnesium chloride), 2 μl of 10 mM dNTPs, 100 pmol of oligo(dT), 200 units of Moloney murine leukemia virus reverse transcriptase for 60 min at 37 °C. Reactions were subjected to 35 cycles of denaturation at 94 °C for 1 min, annealing at 45 °C for 2 min, and polymerization at 72 °C for 2 min. For mouse BMP-5, a 287-bp fragment was isolated by PCR from mouse embryonic mRNA under similar conditions as described for GDF-1. A 5'-BMP-5-specific PCR primer (CCATGGCATCCCATCTACCGAGGCGAAGCTTGGACACCTCCTGGGAGG) and 3'-GDF (GGAA-TGCACTTACCGGCAACGAAAAGG) were used for the first strand synthesis. A 5'-BMP-5-specific PCR primer (CCAGAC-CCAGACTTACCTACCGAGGCGAAGCTTGGACACCTG) was added for PCR amplification. PCR fragments were gel purified and cloned into a Bluescript KS(-) vector.

**RESULTS**

**Cloning of Murine and Human OP-2 cDNA**—In an effort to isolate additional OP-1 related genes, we screened several mouse cDNA libraries, using an OP-1 probe derived from the TGF-β domain (0.32-kb StuI-EcoRI fragment). A 17-day mouse embryo cDNA library yielded not only the murine homolog of the human OP-1 gene (28) but also the new gene, termed OP-2. Only one OP-2 clone was found compared to four of OP-1, indicating a low abundance at the 17-day stage. A murine OP-2 clone was also found in a teratocarcinoma (PC14) cDNA library.

The human OP-2 gene was isolated from a hippocampus cDNA library, previously the source of human OP-1 cDNA, by screening with murine OP-2 probe specific for the pro-region (0.3-kb EcoRI-BamHI fragment). A positive clone (λ O24 and subclone pO166) shared extensive sequence homology with murine OP-2. It contained 0.4 kb of 5'-noncoding sequences, the complete pro-region, and the first half of the TGF-β domain but lacked 0.14 kb from the C terminus. The last portion of the TGF-β domain was obtained from a human genomic library (EMBEC-3, Clontech; λ O28 contained the last four exons of human OP-2 and provided the missing part of the TGF-β domain on a BamHI-PstI fragment of 0.8 kb (pO173-2). Analysis of exons 4, 5, and 6 from this genomic clone revealed nucleotide sequences identical to the respective regions in the human cDNA. Subsequently, the entire human genomic OP-2 locus was isolated.

Based on cDNA and genomic DNA sequence, the predicted size for human pre-pro-OP-2 is 402 amino acids (Fig. 1). The putative secretion signal peptide contains 19 amino acids and the pro-protein contains 383 amino acids. The sequence RTPRJA is proposed as the maturation cleavage site of OP-2, alanine being the first amino acid of the mature protein. Therefore, the N terminus of the mature protein, up to the first cysteine in the TGF-β domain, comprises 37 aa with mature OP-2 being 138 amino acids in length, like OP-1 and Vgr-1.

**Alignments of the TGF-β Domains of OP-2 and Related Proteins**—A distinctive feature of OP-2 is the presence of an additional cysteine in the TGF-β domain, which corresponds to a tyrosine residue in OP-1. OP-2 is the only TGF-β-related protein that breaks the pattern of 7 conserved cysteines in the TGF-β domain, as seen in the alignment with OP-1, BMP-5, Vgr-1, BMP-2, BMP-3, BMP-4, DPP, 60A, Vg-1, GDF-1, activins A and B, and TGF-β1 (Fig. 2). This alignment places OP-2 near OP-1, Vgr-1, BMP-5, and also 60A. Similarities, as percentage of identical amino acids, are compiled in Table I. OP-2 is equally related (in its TGF-β domain) to OP-1 (74%), Vgr-1 (BMP-6) (75%), and BMP-5 (76%) and also very close to the Drosophila 60A protein (65%). It is notably more distant from BMP-2 and BMP-4 (55%), which have arginine rather than histidine at the C terminus (Fig. 2). A glycosylation site in the center of the TGF-β domain is shared by OP-2, OP-1, BMP-5, BMP-6, 60A, DPP, BMP-2, and BMP-4, but absent in BMP-3, and more distantly related proteins.

**Comparison of Human and Murine OP-2**—Members of the TGF-β superfamily typically show extensive conservation across species. The alignment of human and mouse OP-2 in (Fig. 1) shows only 4 amino acid changes in the TGF-β domain while the mature proteins differ by 13 amino acids. In the pro-region, murine OP-2 is 3 amino acids shorter than human OP-2. Some diversions is found near residue 90 of OP-2, where OP-2 and OP-1 diverged the most (Fig. 3). Overall, a total of 59 aa changes between human and murine pre-pro-OP-2
OP-2, a New Member of the TGF-β Superfamily

FIG. 2. Sequence alignments of 7-cysteine domains. The amino acid sequence of OP-1 is aligned with that of BMP-5, BMP-6/Vgr-1, Drosophila 60A, BMP-2, BMP-4, DPP, Vg-1, BMP-3, GDF-1, activin A, activin B, TGF-β1. Black bars indicate complete conservation among all sequences and the shaded areas indicate partial conservation. Dashes (-) indicate gaps introduced for alignment.

TABLE I

| Identities between the 7-cysteine-TGF-β domains |
|------------------------------------------------|
| The percent identities in the 7-cysteine domains, which are aligned in Fig. 2, are arranged in this matrix as groups of similarity, separated by double lines. |

| OP-2 | OP-1 | Vg-1 | BMP-5 | BMP-6 | 60A | BMP-2 | BMP-4 | DPP | Vg-1 | BMP-3 | GDF-1 | Act. A | Act. B | TGF-β1 |
|------|------|------|-------|-------|-----|-------|-------|-----|------|-------|-------|-------|-------|-------|
| OP-2 | 96   | 95   | 94    | 93    | 92  | 90    | 90    | 89  | 88   | 87    | 87    | 86    | 85    | 85    |
| OP-1 | 98   | 97   | 96    | 95    | 94  | 92    | 92    | 91  | 90   | 90    | 90    | 89    | 89    | 89    |
| Vg-1 | 99   | 99   | 98    | 97    | 96  | 94    | 94    | 93  | 92   | 92    | 92    | 92    | 92    | 92    |
| BMP-5| 98   | 98   | 97    | 96    | 95  | 93    | 93    | 92  | 91   | 91    | 91    | 90    | 90    | 90    |
| BMP-6| 97   | 97   | 96    | 95    | 94  | 92    | 92    | 91  | 90   | 90    | 90    | 89    | 89    | 89    |
| 60A  | 96   | 96   | 95    | 94    | 93  | 91    | 91    | 90  | 90   | 90    | 90    | 89    | 89    | 89    |
| BMP-2| 95   | 95   | 94    | 93    | 92  | 91    | 91    | 90  | 90   | 90    | 90    | 89    | 89    | 89    |
| BMP-4| 94   | 94   | 93    | 92    | 91  | 90    | 90    | 89  | 89   | 89    | 89    | 88    | 88    | 88    |
| DPP  | 93   | 93   | 92    | 91    | 90  | 89    | 89    | 88  | 88   | 88    | 88    | 87    | 87    | 87    |
| Vg-1 | 92   | 92   | 91    | 90    | 89  | 88    | 88    | 87  | 87   | 87    | 87    | 86    | 86    | 86    |
| BMP-3| 91   | 91   | 90    | 89    | 88  | 87    | 87    | 86  | 86   | 86    | 86    | 85    | 85    | 85    |
| GDF-1| 90   | 90   | 89    | 88    | 87  | 86    | 86    | 85  | 85   | 85    | 85    | 84    | 84    | 84    |
| Act. A| 89   | 89   | 88    | 87    | 86  | 85    | 85    | 84  | 84   | 84    | 84    | 83    | 83    | 83    |
| Act. B| 88   | 88   | 87    | 86    | 85  | 84    | 84    | 83  | 83   | 83    | 83    | 82    | 82    | 82    |
| TGF-β1| 87  | 87   | 86    | 85    | 84  | 83    | 83    | 82  | 82   | 82    | 82    | 81    | 81    | 81    |

significantly exceed the 11 changes between human and murine OP-1 (28). Diminished conservation across species is also found with BMP-3 and GDF-1 (10).

Comparison with OP-1—Alignments of human OP-2 with OP-1 (Fig. 3) may be used to decipher the structure and possible role of the pro-regions and to confirm the deduced sequences. OP-2 has only marginal homology with OP-1 in the signal peptide and adjacent pro-region (the first 33 residues). A nearby RXXR protease site of OP-1 is replaced in OP-2 by RXXG, at residue 37. However, a substitute RXXR pattern appears, shifted by 5 amino acids.

The pro-region of OP-2 and OP-1 share a potential glycosylation site (present also in BMP-5 and -6). While OP-1 lacks cysteines outside of the TGF-β domain, OP-2 contains a cysteine in the signal peptide and a second one at residue 31. A third cysteine is at residue 205 in the pro-region of human but not murine OP-2.

Proteolytic Maturation Sites in the Pro-domain—The proteolytic cleavages resulting in removal of pro-regions from the mature proteins occur immediately past the sequence FXXR in members of this family (7, 16, 28). Mutants of OP-1 with a minor alteration of the RXXR sequence are not properly cleaved. Several genes of this family encode multiple RXXR patterns in the pro-region. For example, human OP-2 has three additional RXXR sites in the pro-region. N-terminal sequencing of mature OP-1 has shown the maturation site to be the most downstream RXXR pattern (25). However, additional upstream RXXR maturation sites can also be processed. Fig. 4 shows alignments of mature N termini for different members of the TGF-β superfamily. Comparison of 16 examples shows that these RXXR sites are typically followed by serine or alanine. On occasion, RXXR is followed by aspartic acid, glycine, or glutamine.

N Termini of the Mature Proteins—An alignment of the

1 Kaplan, P., Dorai, H., Özkanay, E., and Oppermann, H., unpublished results.

2 R. Tucker, personal communication.
mature N termini, up to the first cysteine of the TGF-β domain, for different superfamily members leads to interesting observations, regarding the length and composition (Fig. 4). Long N termini are found for the closely related OP-1, OP-2, and BMP-6 (37 aa) and BMP-5 (36 aa). N-terminal residues of BMP-5, BMP-6, and OP-1 differ mainly by conservative changes. The alignment of charged residues is nearly identical in all members of the superfamily that have osteogenic potential. It is absent in others, such as the BMP-2 group, the activins, and MIS.

**OP-2 mRNA Expression**—In order to detect OP-2 expression we screened mRNA preparations from several organ tissues of 2-day-old rats, such as brain, calvaria, heart, kidney, and lung by Northern blot hybridization (Fig. 5). For this analysis brain was chosen since human OP-2 was isolated upstream of the TGF-β domain. This feature is found in all members of the superfamily that share a high content of basic amino acid residues. This feature is found in all members of the superfamily that have osteogenic potential. It is absent in others, such as the TGF-β group, the activins, and MIS.

**Other TGF-β Superfamily Members in Specific Organs**—In order to detect OP-2 expression in grown animals we screened several mouse organs by Northern hybridization. We extended this analysis to eight TGF-β-related genes (OP-1, OP-2, Vgr-1, BMP-2, BMP-3, BMP-4, BMP-5, and GDF-1) as no stringent or comprehensive study was done with specific probes; murine OP-1 (0.7-kb BstXI-BglI fragment), murine OP-2, 3’-untranslated region (0.42-kb XmnI-Hind111 fragment), murine Vgr-1 (0.26-kb PouI-SsaI fragment), and human BMP-4 (0.7-kb BstEII fragment). The location of probes can also be seen in the line diagrams on the right panel of Fig. 7. Markings on the right of each hybridization correspond to 8.48-, 7.46-, 4.40-, 2.37-, and 1.35-kb RNA (0.24-9.49-kb RNA ladder; Gibco-BRL).

were detected in 6-day postnatal animals and kidneys from 2-week-old animals.

**mRNA Expression of Other TGF-β Superfamily Members in Specific Organs**—In order to detect OP-2 expression in grown animals we screened several mouse organs by Northern hybridization. We extended this analysis to eight TGF-β-related genes (OP-1, OP-2, Vgr-1, BMP-2, BMP-3, BMP-4, BMP-5, and GDF-1) as no stringent or comprehensive study of mRNA expression is found in the literature. Cross-hybridization was minimized by using specific probes corresponding to highly diverged sequences. Poly(A)⁺ RNA from brain, spleen, lung, heart, liver, and kidney of 2-week-old and 6-9-month-old male mice was analyzed on 1.2% agarose-formaldehyde gels, transferred to a membrane and hybridized sequentially to different probes (Fig. 7).

Clear results were obtained for most genes but no mRNA was detected in any organs studied for OP-2 (data not shown) and BMP-2. OP-1 message is mainly present in the kidneys and at a lower level in brain tissue as a total of four species (1.8, 2.2, 2.4, and 4.0 kb) (28). BMP-6/Vgr-1 mRNA, a single species of about 3.7 kb, is found in lungs and at low level in kidneys. BMP-5 mRNA is located in lungs and liver as a total of four species of about 3.7 kb, is found in lungs and at low level in kidneys. BMP-5 mRNA is located in lungs and liver as a single species of 4.2 kb. BMP-4 mRNA is mainly in lungs and at much lower levels in kidneys (two species of 1.8 and 2.1 kb). BMP-4 specific signal was also detected in spleen and liver, however, migrating slightly different. Possibly this sig-
Fig. 6. Expression of OP-2 mRNA in murine embryos and 6-day postnatal animal. Equal amounts of poly(A)+ RNA (10 µg) were loaded into each lane. Hybridizations were done with probes specific for the pro-region of murine OP-2 (0.31-kb EcoRI-BamHI fragment), and the 3′-untranslated region of murine OP-2 (0.42-kb XmnI-HindIII fragment). The locations of the probes are shown with bold lines underneath the line diagram corresponding to the murine OP-2 cDNA. Lane 1, poly(A)+ kidney RNA (a negative control); lanes 2–7, poly(A)+ RNA obtained from 8-day embryos (lane 2), 10-day embryos (lane 3), 17-day embryos (lane 4), 6-day postnatal animals (lane 5), 2-week-old animal kidneys (lane 6), 8-day embryos (lane 7). Lanes 1–6, hybridization to a pro-region specific OP-2 probe, and lane 7, hybridization to a 3′-untranslated region specific OP-2 probe. The gray box in the OP-2 line diagram indicates the signal peptide and the hatched box corresponds to the TGF-β domain. An arrow marks the location of the maturation site. The sizes of the major OP-2 mRNA species are shown on the left.

3 kb
5 kb
EF-Tu

Fig. 7. mRNA expression of TGF-β superfamily members in organs of mice. Expression of OP-1, Vgr-1, BMP-5, BMP-4, BMP-2, BMP-3, and GDF-1 mRNA was studied in selected organ tissues of 2-week- and 6–9-month-old mice. Equal amounts (5 µg) of poly(A)+ RNA were loaded onto each lane of a 1.2% agarose-formaldehyde gel. Lanes 2, 4, 6, 8, 10, and 12 correspond to poly(A)+ RNA obtained from 2-week-old mice, lanes 3, 5, 7, 9, 11, and 13 correspond to poly(A)+ RNA obtained from 6–9-month-old mice. Lanes 1 and 14 contain poly(A)+ RNA (from brain and liver of 2-week-old mice). The same membrane was used for sequential hybridizations with probes specific for: murine OP-1 (0.7-kb BstXI-BglII fragment); murine Vgr-1 (0.26-kb PvuII-Sacl fragment); murine BMP-5 (0.28-kb PstI fragment, PstI sites were introduced during the PCR); human BMP-4 (0.7-kb BstEII fragment); murine BMP-3 (0.27-kb EcoRI-EcoRV fragment); human BMP-2 (0.7-kb NcoI fragment); murine GDF-1 (0.47-kb ClaI-PstI fragment, ClaI site from the vector); murine EF-Tu (0.5-kb HindIII-BamHI fragment, BamHI from the vector). Locations of the probes are shown in bold lines underneath the cDNA diagrams. The gray boxes in the diagrams indicate the signal peptides and hatched boxes correspond to the 7-cysteine/TGF-β domains. Arrows mark the location of maturation sites. The sizes of the major RNA species are shown on the left. Markings on the right of each hybridization correspond to 9.49-, 7.46-, 4.40-, 2.37-, and 1.35-kb RNA (0.24–0.49-kb RNA ladder).

Presence of Multiple or Large Transcripts—The presence of oversized and multiple mRNA transcripts of OP-2, two species of 3 and 5 kb in 8-day embryos, has ample precedence among the other related genes. Most members of the superfamily have transcripts much larger than the expected 1.4–1.8-kb size and several have different size transcripts (Fig. 7). Northern blots showed four OP-1 mRNA species: 1.8, 2.2, 2.4, and 4 kb (28). BMP-3 also has multiple size transcripts (2.5 and 7.0 kb). BMP-5 and Vgr-1 (8) have single size transcripts of about 4 kb, much longer than expected.

EF-Tu

Genomic Structure of the Human OP-2 Gene—The human OP-2 locus was isolated from a genomic library on three overlapping phage clones (λ O34, O35, O28), the OP-2 coding information spread over 7 exons and 27 kb. The entire nucleotide sequence with exception of the largest two introns, 1 and 3, was determined from 20 subclones (Fig. 8). The exon-intron boundaries are shown in Fig. 9. A comparison of exon-intron boundaries in the TGF-β domain showed matching locations with those of OP-1 (9). The first OP-2 exon contains 334 bp of coding sequence (111 amino acids), including the signal peptide, and is followed by the largest intron (14.6 kb). The second exon (190 bp, 64 amino acids) is separated by a short intron (0.4 kb) from exon 3 (149 bp, 49 amino acids). It follows a large third intron of 9.5 kb. The fourth exon (195
Fig. 8. Genomic structure of OP-1, OP-2, and BMP-4. OP-1 and OP-2 genes contain seven exons, whereas the BMP-4 gene contains two exons. Some subclones used to derive the genomic structures are shown in bold lines, with the corresponding phage clones in parentheses. The OP-2 gene spans 27 kb of genomic distance (contained in three phage clones). K and X indicate KpnI and XhoI sites, respectively. R and P indicate EcoRI and PstI sites in the BMP-4 locus. Dotted lines in the OP-1 diagram indicate undetermined intron lengths. Genomic subclones used to derive the OP-1 structure: p0142 (1.7-kb PstI fragment), p0247 (0.6-kb XhoI fragment), p0243 (1.7-kb PstI fragment), p0248 (0.3-kb Sau3AI fragment), p0229 (0.7-kb PstI fragment), p064 (5-kb EcoRI fragment), p057 (3-kb BamHI fragment). Genomic subclones used to determine the OP-2 structure: p0210 (1.7-kb PstI fragment), p0208 (9-kb XhoI fragment), p0217 (9-kb PstI fragment), p0174 (0.8-kb PstI fragment), p0179 (2.8-kb StuI fragment), pol74 (0.8-kb PstI fragment), pol79 (2.8-kb StuI fragment), p0247 (0.6-kb HaeIII fragment), p0248 (0.3-kb Sau3AI fragment), p0249 (0.6-kb PstI fragment), p0247 (0.6-kb XhoI fragment). Genomic subclones used to determine the BMP-4 structure: p050 (2.2-kb EcoRI fragment), p049 (1.9-kb PstI fragment), p044 (1.9-kb EcoRI fragment).

bp, 65 amino acids) encodes the RTPRA maturation site and is followed by a 0.8-kb intron. The TGF-β domain resides on exons 5–7: exon 5 (80 bp, 27 amino acids) encodes the first cysteine of mature OP-2 and is followed by intron 5 (0.5 kb in length), exon 6 (111 bp, 37 amino acids) is separated by a 2.5-kb intron from the seventh, last exon with 147 bp (49 amino acids) of coding sequence.

Genomic Structure of Human OP-1—To complete the genomic comparison of OP-2 and OP-1 we isolated the remaining exons of OP-1 from a human genomic library on three EcoRI fragments (p050, p044). The complete BMP-4 coding region is present in only two exons, 0.37 and 0.86 kb, interrupted by an intron of 0.96 kb (Fig. 8).

DISCUSSION

The present study describes a new gene which is closely related to OP-1, BMP-5, Vgr-1, and Drosophila 60A. Based
A close inspection of the proteolytic maturation sites of several precursors shows that the pattern RXRR can be further specified as RXRXR[A] or RXRXRI[S], which are found in the majority of cases. In OP-2 the TGF-β domain follows two potential maturation sites, 21 residues apart. Multiple RXRR sites near the TGF-β domain are also seen in BMP-2 and BMP-4, about 30 residues apart. Additional proteolytic maturation sites are found further upstream in the pro-region. In human OP-2 the multiple maturation sites could release five potential polypeptides, spanning from amino acid 20 (signal peptide cleavage site) to as 42, from 43 to 101, from 102 to 242, from 243 to 263, and the mature C-terminal domain, with the possibility of additional biological activities.

The mature N termini of different members of the TGF-β superfamily are quite diverse in their amino acid sequence. The extensive variations found in the mature N termini may provide distinction between the otherwise conserved mature proteins. Possibly this region has diverged because it is not crucial for receptor binding or protein folding. However, for a given protein this region is still relatively well conserved among animal species, in support of a functional role. N-terminal sequences may supply the individual proteins with more specific recognition for their respective receptors. They may allow targeting of specific tissues via binding to extra-cellular matrix. The N terminus of OP-2 has some similarity in composition to nuclear localization sequences that contain several prolines interspersed with lysines or arginines (36), a feature not seen for the other osteogenic proteins.

The high content of basic amino acids in the N terminus is characteristic for the osteogenic proteins. In contrast, a mature N terminus of only 10-14 amino acids with few basic residues is characteristic for the more distantly related proteins, TGF-β 1-5, the activins, inhibin α, and MIS. The positively charged N termini of the osteogenic proteins may permit binding of the proteins to hydroxylapatite, a resin traditionally used during purification of the osteogenic protein (29). In vivo, the basic N termini may mediate the deposition of osteogenic proteins in bony tissue. However, in the subcutaneous bone induction assay the N termini are not essential for biological activity (25) perhaps due to the local administration of the protein, immobilized on collagen matrix. The insect proteins, DPP and 60A, also display positively charged N termini, even though their role is not bone induction.

The analysis of mRNA species for osteogenic proteins has been difficult due to the extremely low level of natural expression of these proteins, and published data are sparse. Analysis by dot blot or by in situ hybridization is confounded by the possibility of cross hybridization of probes to different members. In our study we have minimized this possibility by careful choice of probes. Selection of poly(A)+ RNA on oligo(dT)-cellulose is necessary due to low levels of expression of these mRNAs. The low level of expression may be a reflection of regulatory function and high biological activity. Nanogram amounts suffice for subcutaneous bone induction in the rat model (29).

Northern analysis of several organs with probes for various family members indicates that mRNA for most of them is expressed in specific organs. OP-1 mRNA is expressed mainly in kidneys and bladder (28), which may explain the epithelial osteogenesis, discovered by Huggins over 60 years ago. Huggins (38) noted that urinary tract epithelia implanted into the abdominal wall of dogs evoked large amounts of bone within 12 days. OP-1 mRNA is also expressed in brain which is the sole site of GDF-1 expression. In contrast, we find BMP-3, BMP-4, BMP-5, and BMP-6/Vgr-1 to be primarily expressed in the lungs. The lungs may participate in the growth regulation of bone and connective tissues in an endocrine manner, as proposed for the kidney (28).

OP-2- and BMP-2-specific mRNAs were not detected in any of the adult organs by our Northern hybridization analysis. OP-2 cDNA was found at low abundance in a hippocampus cDNA library and may be expressed at low levels in brain. However, OP-2 mRNA was found at relatively high levels in 8-day mouse embryos, indicating a developmental role; in the adult animal OP-2 and BMP-2 may be expressed in a more discrete location, or primarily during tissue regeneration. While the timing of expression for BMP-5 and OP-1 seems to be directly related to growth, an inverse relationship was found for Vgr-1 and BMP-4. The level of BMP-3 expression (in lungs) did not change with the age of the animal.

The early embryo displays two oversize OP-2 mRNA species. Hence, the OP-2 locus may be considerably larger that the 27 kb cloned so far. The multiple size transcripts observed for OP-1, OP-2, BMP-3, and BMP-4 may result from splicing events that affect coding or untranslated regions, or they may represent bicistronic mRNA species (10). Diverse transcripts have also been seen with DPP. These multiple DPP-specific mRNAs have been shown to be due to alternatively spliced 5′ untranslated exons (22).

There is an apparent redundancy of osteogenic proteins in the TGF-β superfamily. Since bone has different architectures and fine structures depending on the anatomical localization (for example, long bones, facial bones, skull plates, and dentin), the osteogenic proteins may have evolved along with the different types of bone. Roles other than bone formation are likely since analogs of the osteogenic proteins have been found in invertebrates (DPP and 60A of Drosophila). Multiple functions are indicated by the fact that the bone morphogenetic proteins are expressed in early embryonic development as well as later in life.

Acknowledgments—We thank Scott Bagley for excellent sequencing work. We thank John E. Smith for stimulating discussions and Donald Engelman and John Lee for their overall input in the bone project.

REFERENCES

1. Rebagnoli, M. R., Weeks, D. L., Harvey, R. P., and Melton, D. A. (1985) Cell 42, 769-777.  
2. Cate, R. L., Mattaliano, R. J., Hession, C., Tizard, F., Varner, N. M., Cingari, A., Nifza, G., Chow, E. F., Fehilly, R. A., Bertoni, J. M., Torres, G., Walmer, B. P., Ramachandran, K. L., Ragan, R. C., Manganaro, T. F., MacLauchlin, D. T., and Donahue, P. F. (1986) Cell 45, 607-618.  
3. Mason, A. J., Hayflick, J. S., Ling, E., Esch, F., Ueno, N., Ying, S.-Y., Guillemin, R., Niall, H., and Seeburg, P. H. (1985) Nature 318, 659-663.  
4. Ling, N., Ying, S.-Y., Ueno, N., Shimasaki, S., Esch, F., Hotta, M., and Guillemin, R. (1986) Nature 321, 777-780.  
5. Vale, W., Roger, J., Vaughan, J., McClintock, R., Corrigan, A., Woo, W., Karr, D., and Spiess, J. (1986) Nature 321, 777-770.  
6. Wozney, J. M., Rosen, V., Celeste, A. J., Mitochon, L. M., Whitters, M., Derynck, R., Kritz, R. W., Hewick, R., and Wang, E. (1988) Science 242, 1528-1533.  
7. Celeste, A. J., Innarezi, J. A., Taylor, R. C., Hewick, R. M., Roson, V., Wang, E. A., and Wozney, J. M. (1990) Proc. Natl. Acad. Sci. U. S. A. 87, 8643-8647.  
8. Lyons, K., Griswod, J. L., Lee, A., Hashimi, S., Lindenquist, P., Chen, E., Hogen, B. L., and Derynck, R. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 4545-4548.  
9. Oksiyanuk, F., Raeger, D., Drier, E. A., Corbett, C., Hidge, R. J., Sampata, T. K., and Oppermann, H. (1990) EMBO J. 9, 2085-2093.  
10. Lee, S.-J. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 4250-4254.  
11. Segal, D., and Gelb, W. M. (1985) Genetics 109, 119-143.  
12. Wartort, K. A., Thomasen, G. H., and Gelbort, W. M. (1991) Proc. Natl. Acad. Sci. U. S. A. 88, 9214-9218.  
13. Mathews, L. S., and Vale, W. W. (1991) Cell 65, 973-982.  
14. Lin, H. Y., Wang, X.-F., Neaton, E., Weinberg, R., and Lodish, H. F. (1992) Cell 70, 775-785.  
15. Massague, J. (1992) Cell 69, 1067-1070.  
16. Parhaman, G. E., Coffee, C., Ittleson, M. D., and Hoffmann, F. M. (1990) Mol. Cell. Biol. 10, 2690-2677.  
17. Irish, V. F., and Gelbort, W. M. (1987) Genes Dev. 1, 868-879.  
18. Vigorita, V. J., Fronza-Van Rani, A., Zuch, J. M., van Zoelent, E., J. J., van Nimmen, K., Koster, C. H., Snek, G. T., Durston, A. J., and Haylock, D. (1990) Nature 346, 752-754.  
19. Smith, J. C., Price, B. M., van Nimmen, K., and Haylock, D. (1990) Nature 345, 729-731.
OP-2, a New Member of the TGF-β Superfamily

20. Thomsen, G., Woolf, T., Whitman, M., Sokol, S., Vaughan, J., Vale, W., and Melton, D. A. (1990) Cell 63, 455-453
21. Koster, M., Plessow, S., Clement, J. H., Lorenz, A., Tiedemann, H., and Knöchel, W. (1991) Mech. Dev. 33, 181-200
22. Behringer, R. R., Cate, R. L., Froelick, G. J., Palmieri, R. D., and Brinster, R. L. (1990) Nature 345, 167-170
23. Wang, E. A., Rosenberg, V., D’Alessandro, J. S., Bauduy, M., Cordes, P., Harada, T., Hesp, D., Hewick, R., Kers, K. M., Lapan, P., Luxenberg, D. F., Mróz, D., Muses, I. K., Nov, J., and Wornoy, J. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 2220-2224
24. Hammonds, R. G. Jr, Schwall, R., Dudley, A., Berkemeier, L., Lai, C., Lee, J., Cunningham, N., Redd, A. H., Wood, W. L., and Mason, A. J. (1991) Mol. Endocrinol. 5, 149-155
25. Sampath, T. K., Matiskel, J. C., Hauschka, P. V., Jones, W. K., Sasa, H., Tucker, R. F., White, K., Coughlin, J. E., Tucker, M. M., Peng, R. H. L., Corbett, C., Ozkaynak, E., Oppermann, H., and Rueger, D. C. (1992) J. Biol. Chem. 267, 20352-20362
26. Sampath, T. K., and Reddi, A. H. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 7599-7603
27. Cook, S. D., Baffes, G. C., Sampath, T. K., and Rueger, D. C. (1991) J. Mineral Res. 4, Suppl. 1 (Abstr. 283)
28. Ozkaynak, E., Schnegelsberg, P. N. J., and Oppermann, H. (1991) Biochem. Biophys. Res. Commun. 179, 116-123
29. Sampath, T. K., Coughlin, J. E., Whetstone, R. M., Banach, D., Corbett, C., Ridge, R. J., Ozkaynak, E., Oppermann, H., and Rueger, D. C. (1990) J. Biol. Chem. 265, 13198-13205
30. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467
31. Ozkaynak, E., and Putney, S. D. (1987) BioTechniques 5, 770-773
32. Chomczynski, P., and Sacchi, N. (1987) Anal. Biochem. 162, 156-159
33. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (eds) (1990) Current Protocols in Molecular Biology, Wiley-Interscience, New York
34. Daopin, S., Piez, K. A., Ogawa, Y., and Davies, D. R. (1992) Science 257, 369-373
35. Schlesinger, M. P., and Gruetter, M. G. (1992) Nature 358, 430-434
36. Wagner, P., Kunz, J., Koller, A., and Hall, M. N. (1990) FEBS Lett. 275, 1-5
37. St. Johnston, R. D., Hoffmann, F. M., Blackman, R. K., Segal, D., Grimmel, R., Padgett, R. W., Irick, H. A., and Gelbart, W. M. (1990) Genes & Dev. 4, 1114-1127
38. Huggins, C. B. (1931) Arch. Surg. 22, 377-408