The EMBO Journal vol.9 no.7 pp.2085 - 2093, 1990

OP-1 cDNA encodes an osteogenic protein in the TGF-β family

Engin Özkaynak, David C.Rueger, Eric A.Drier, Clare Corbett, Richard J.Ridge, T.Kuber Sampath and Hermann Oppermann

Creative BioMolecules, Inc., 35 South Street, Hopkinton, MA 01748, USA

Communicated by T.Graf

Amino acid sequences of two tryptic peptides derived from enriched bovine osteogenic protein preparations revealed considerable homology to two members of the TGF-β (transforming growth factor β) supergene family, DPP (decapentaplegic protein) of *Drosophila* and Vg-1 (vegetal protein) of *Xenopus*. Building upon this information we constructed a synthetic consensus gene to use as a probe to screen human genomic libraries. This resulted in the isolation of three interrelated genes. Among these were BMP-2b and BMP-3 which have recently been described by others. The third gene, termed OP-1 (osteogenic protein one), is new and was subsequently shown to encode the human homolog of a major component of bovine osteogenic protein. The genomic clones were used to isolate the corresponding complementary DNA (cDNA) clones. Sequence analysis indicates that OP-1 is a relative of the murine *Vgr*-1 (Vg-1 related gene). This report describes the cDNA structure and putative amino acid sequence of OP-1.

Key words: bone morphogenesis/consensus gene/differentiation/growth factor/supergene family

Introduction

The superfamily of transforming growth factor beta (TGF-β) related proteins has recently emerged. It includes the inhibins (Mason *et al.*, 1985; Forage *et al.*, 1986; Mayo *et al.*, 1986), the Müllerian inhibiting substance (MIS) (Cate *et al.*, 1986), *Drosophila melanogaster* decapentaplegic complex (DPP) (Padgett *et al.*, 1987) and the Vg-1 gene product of *Xenopus laevis* (Weeks and Melton, 1987). Four new additions have been reported: a set of bone morphogenetic proteins (BMP-2a, BMP-2b, BMP-3) (Wang *et al.*, 1988; Wozney *et al.*, 1988) and the Vg-1 related murine protein Vgr-1 (Lyons *et al.*, 1989).

The mature TGF-β like proteins share a distinctive pattern of seven cysteines, C...CXGXC...CC...CXCX, while TGF-β itself and the β chains (βA,βB) of inhibins and activins display two additional cysteines: C...CC...CXGXC...CC... CXCX. However, all members of the TGF-β superfamily are secreted as precursors approximately four times larger than the mature form and the sequences of the precursor portions of the family members show limited sequence homology. The mature protein is represented in the C-terminal domain and appears to be cleaved by a trypsin like protease (Brunner *et al.*, 1988; Pepinsky *et al.*, 1988; Tannahill and Melton, 1989). The pro region appears to be essential for proper *in vivo* folding of the mature region (Gray and Mason, 1990). All mature forms of the TGF-β like proteins occur as disulfide linked dimers and several of the precursors have been found associated as dimers (Miyazano *et al.*, 1988; Pepinsky *et al.*, 1988). Heterodimers of inhibin α and β chains [inhibins (αβ)] and homodimers of inhibin β chains [activins (ββ)] have been found to differ in biological activities (Ling *et al.*, 1986).

The mature domain of a TGF-β like protein is found to be highly conserved across animal species. TGF-β like proteins are involved in diverse activities in different cell types. For example, MIS causes the regression of the Müllerian duct in the male, which in the female develops into reproductive organs (Cate *et al.*, 1986). Inhibins and activins control the follicle stimulating hormone (FSH) secretion (Ling *et al.*, 1985, 1986; Vale *et al.*, 1986). The product of the DPP gene is involved in the dorsal ventral development of the embryo and possibly in other developmental events later in embryogenesis (Segal and Gelbart, 1985). Vg-1 is of maternal origin in *Xenopus* eggs and is thought to act as an intercellular signal for mesoderm induction (Rebagliati *et al.*, 1985). Another recent addition to the TGF-β superfamily is the *Vgr*-1 gene which is found to be expressed in both embryonic and adult tissues suggesting that it plays a role at different stages throughout development (Lyons *et al.*, 1989). The *Vgr*-1 gene is rather closely related to some of the recently described BMP genes. BMP-2a, BMP-2b and BMP-3 genes were identified on the basis of peptide sequences found in preparations of bovine osteogenic proteins (Wozney *et al.*, 1988; our unpublished work) and the recombinant expression product of BMP-2a has been observed to induce bone formation in rats (Wang *et al.*, 1990).

Bone morphogenetic activity was first described by Urist (1965). An *in vivo* assay to assess osteogenic activity in rats has permitted further purification of proteins involved in bone formation (Sampath and Reddi, 1981). After extensive purification the bovine osteogenic protein presents itself as disulfide linked dimer of 30 kd and can be reduced into distinct but related subunits of 16 kd and 18 kd (Sampath *et al.*, 1990).

In our strategy to isolate the gene(s) encoding osteogenic protein, tryptic peptides were generated from an enriched osteogenic protein preparation, sequenced and the information was used in a novel approach to construct a consensus gene probe, as described below. By screening a human genomic library with this probe we isolated the genes for BMP-2b and BMP-3 and a new gene termed OP-1. Here we report the isolation of the human OP-1 gene, which encodes a subunit of osteogenic protein, and show its relationship to the TGF-β superfamily.

Results and discussion

Purification of bovine osteogenic protein and isolation of peptides T2 and T4

The method for purification of osteogenic protein from bovine bone was described by Sampath *et al.* (1987, 1990). The biological activity was monitored using the *in vivo* bone induction assay in rats. In this assay, subcutaneous implantation of osteogenic protein with 4 M guanidine extracted rat matrix induces the formation of new bone. The osteogenic activity was evaluated after 12 days *in vivo* by histological examination and quantitated by measuring alkaline phosphatase activity and calcium content. Figure 1 shows an SDS protein gel of an enriched osteogenic protein preparation. The preparation was obtained from the C18 reverse phase chromatography step in the purification. Biologically active fractions were compared under non-reducing and reducing conditions using SDS−PAGE and Coomassie blue staining; a prominent diffuse band migrating with M_r 30 kd under non-reducing conditions disappears after reduction and several species in the range of M_r 15−18 kd appear instead.

It was previously observed that the biological activity in the rat implant assay is abolished by reducing agents in the presence of denaturants such as 4 M guanidine or SDS (Sampath *et al.*, 1987). In contrast, SDS treatment in the absence of reducing agent does not lead to irreversible inactivation. Hence, it was possible to use non-reducing SDS−PAGE as a further step in the identification of osteogenic protein (Sampath *et al.*, 1990). A non-reducing SDS gel, as shown in Figure 1, was horizontally sliced into eight sections, from which proteins were eluted by passive diffusion and separately assayed in the rat implant assay. Bone inducing activity, scored using the criteria mentioned above, was predominantly present in the slice containing the 30 kd species. SDS−gel slices containing 30 kd material were digested with trypsin and two peptides, T2 and T4, gave the following amino acid sequences: R/K-S-F-D-A-Y-Y-C-S-G-A-C-Q-F-P-(S)-P and R/K-?-C-C-V-P-T-E-L-S-A-I-S-M-L-Y-L-D-E-N-E, respectively. R or K at the beginning of the peptides is deducted because trypsin cleaves after arginine or lysine residues (N-terminal sequence analysis of the 30 kd protein had been unsuccessful).

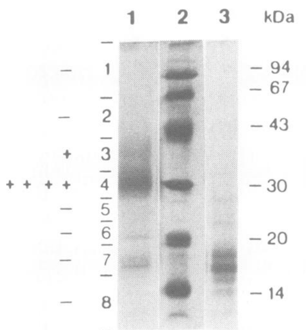

Fig. 1. Electrophoretic mobility of bone inducing activity on non-reducing SDS−PAGE. A pool of active fractions of C18 reverse phase chromatography purified osteogenic protein was analyzed by SDS−PAGE without reducing agent (**lane 1**) and after reduction with DTT (**lane 3**). Mol. wt standards (**lane 2**) are phosphorylase B (94 kd), bovine serum albumin (67 kd), ovalbumin (43 kd), carbonic anhydrase (30 kd), soybean trypsin inhibitor (20.1 kd), α-lactalbumin (14.4 kd). A separate lane, neither fixed nor stained, of the non-reduced sample in **lane 1** was sliced into eight sections as shown by the marks on the left. Slices were soaked in 0.1% SDS and the eluting components were assayed in the rat implantation assay. Extensive bone formation, indicated by + + + +, resulted from eluted protein of M_r 30 kd. No biological activity is found from reduced protein.

Rationale for the synthetic oligonucleotide probe

Given the above peptide sequence information we could have constructed mixed oligonucleotide probes (in consideration of the degeneracy of the genetic code) for screening of gene libraries. However, important additional information became available via a homology search of the PIR/NSF protein

```
                                     highly conserved                                            peptide T2
           L  Y  V  D  F  Q  R :D::V::G::W::D::D::W::I::I::A::P: V  D =F==D==A==Y==Y==C==S==G==A==C==Q==F==P= S  A  D  H  F  N  S
probe:     CTGTACGTGGACTTCCAGCGCGACGTGGGCTGGGACGACTGGATCATCGCCCCCGTCGACTTCGACGCCTACTACTGCTCCGGAGCCTGCCAGTTCCCCTCTGCGGATCACTTCAACAG
OP-1:                                  **A***C*********C*G***************G**
BMP-2:                                  ****T**********A*T*********TG*G*****
BMP-3:                                                 ******AG***A*****T**TT*****     ******T*****T***************G************

                         DPP derived                                                   peptide T4
        -T--N--H--A--V--V--Q--T--L--V--N--N--M--N--P--G--K--V--P--K--P =C==C==V==P==T==E==L==S==A==I==S==M==L==Y==L==D==E==N= S  T  V
probe:  CACCAACCACGCCGTGGTGCAGACCCTGGTGAACAACATGAACCCCGGCAAGGTACCCAAGCCCTGCTGCGTGCCCACCGAGCTGTCCGCCATCAGCATGCTGTACCTGGACGAGAATTCCACCGTG
OP-1:   *************A*C*********G*****CC**TT***C*****G*AA*C***G**************T*C******GC****
BMP-2:    ********T***A*T******************                                           **T**T*********T**A***AGT******TC****************T***T**

          V  L  K  N  Y  Q  E  M  T  V  V  G  C  G  C  R  *
probe:    GTGCTGAAGAACTACCAGGAGATGACCGTGGTGGGCTGCGGCTGCCGCTAA
```

Fig. 2. Synthetic gene probe based on a consensus protein. Shown is the nucleotide sequence of the probe. Major nucleotide homologies of this gene probe with three different genes (OP-1, BMP-2b and BMP-3) are marked by asterisks. Four elements shaped the design of this probe: (i) the tryptic peptide T2(R/K-S-F-D-A-Y-Y-C-S-G-A-C-Q-F-P-[S]-P) originating from BMP-3, and the tryptic peptide T4(R/K-?-C-C-V-P-T-E-L-S-A-I-S-M-L-Y-L-D-E-N-E) originating from BMP-2 (marked by = = =); (ii) the most conserved regions in the TGF-β family (marked by :::), occurring in the first third of the TGF-β domains; (iii) partial consensus of the various TGF-β related proteins; (iv) the amino acid sequence of DPP (marked by ---) which was most closely related to peptides T2 and T4, and which was used where the sequence conservation in the TGF-β family is weak. All three genes, OP-1, BMP-2b and BMP-3, were detected via the highly conserved region near the sequence G-W-D-W-I. Moreover, BMP-3 was detected via the reverse translated peptide T2 sequence, BMP-2 via peptide T4 and homology with DPP, finally, OP-1 via its similarity to DPP.

database, namely, T2 and T4 showing close homology with *Xenopus* Vg-1 and *Drosophila* DPP (BMP-2 and BMP-3 were still unknown then). The relative positions of T2 and T4 peptides in alignments with Vg-1 and DPP, can be seen in Figure 5. A possible strategy would then be to use the Vg-1 or DPP genes as alternative probes. These probes, however, are compromised by the differences in genetic codon usage between *Xenopus, Drosophila* and man.

Therefore, we decided to combine all available information, the peptide sequences T2 and T4 as well as the relationship to the TGF-β family, in particular Vg-1 and DPP, in a synthetic consensus gene (Figure 2), using human codon bias, as found in human TGF-β. This also allowed us to take advantage of the highly conserved amino acid patterns (see also Figure 5) of the TGF-β like proteins, thus enhancing the chances of successful screens. This consensus approach, novel in the TGF-β field, proved itself as it resulted in the identification of the different components of the osteogenic protein, a complexity we had not anticipated. For example, we did not expect that the peptides T2 and T4 were derived from two distinct polypeptides (BMP-3 and BMP-2a). A special benefit of using the consensus probe was the isolation of the OP-1 gene via conserved amino acids and its homology to DPP, even though OP-1 peptides had initially not been found in tryptic digests generated from osteogenic protein preparations. At a later stage, the presence of OP-1 in purified bovine osteogenic protein was established by the identification of various peptides (Sampath *et al.*, 1990).

Design of the synthetic consensus gene

The synthetic gene mimicked the C-terminal TGF-β like domain that is the hallmark of this protein family, but it included only the C-terminal six cysteines. The amino acid sequence was derived by incorporating the following elements: the peptide sequences of T2 and T4; the amino acids conserved in DPP, Vg-1, activin, MIS and TGF-β or in at least two of these proteins; and where consensus was absent, the sequence of DPP. To assign nucleotide sequence, the amino acid sequence was reverse translated into a DNA sequence with human codon bias, as found in TGF-β. The synthetic gene was assembled from five pairs of long oligonucleotides as described in Materials and methods. Briefly, pairs of complementary oligonucleotides up to 80 nucleotides long, spanning major restriction sites, were cloned into pUC vectors and thereafter combined into the complete gene in three ligation steps.

Screening of a human genomic library

We initially screened a genomic library rather than a cDNA library, which may or may not contain a particular cDNA. Also, we used a human DNA library although the peptide sequences were obtained from bovine osteogenic protein. The rationale used was based on the observation that specific TGF-β related genes, especially in their TGF-β domains, show an extremely high degree of protein sequence conservation in different animal species. In addition, it had been established that bovine osteogenic protein is active in several different species tested (Sampath and Reddi, 1983).

Approximately 5×10^5 recombinant clones from a human genomic phage library were screened using the synthetic gene as a probe. Twenty four lambda clones, giving positive hybridization signals, were isolated and further analyzed. Restriction fragments, obtained by *Eco*RI digestion of recombinant λ phage clones, were analyzed by Southern hybridization with the consensus probe. Positive fragments were subcloned into the Bluescript phagemid and their nucleotide sequences were determined. The nucleotide sequences obtained were analyzed by nucleic acid homology searches. Moreover, the sequences were translated into amino acids in all six possible reading frames and searched for the canonic cysteine patterns that characterize the TGF-β family namely, CXGXC, then CC, finally CXCX. This search yielded three genes, BMP-2b, BMP-3 and OP-1. The first two genes have been reported (Wozney *et al.*, 1988) while OP-1 is a new gene. Figure 2 shows the consensus gene probe and some of its major nucleotide homologies with the sequences of OP-1, BMP-2b and BMP-3, all of which were directly isolated using this probe. These results show that the strategy of a consensus probe may prove useful in studying other extended gene families.

The genomic clones of BMP-2b displayed the entire TGF-β domain, containing no intron in this region. BMP-3 only annealed via the upstream region of the 7-cysteine domain, which is interrupted by an intron. The OP-1 gene initially revealed only two small exons, which, when translated into amino acid sequence showed homology to the last two thirds of the TGF-β domains of DPP and Vg-1, particularly with respect to cysteine patterns (see above). The nearest upstream exon, which is still part of the conserved C-terminal TGF-β domain, was found by generating further upstream sequence information from the same genomic clone and searching for cysteines and conserved amino acid patterns after translation in several reading frames. These last three exons of OP-1 span 78 bp, 111 bp and 147 bp, and are separated by two introns of ~2 kb each. The exact locations of these splice sites are shown in Figure 4. The pro region was not identified due to its lack of homology to a known protein (Vgr-1 sequence was not available at the

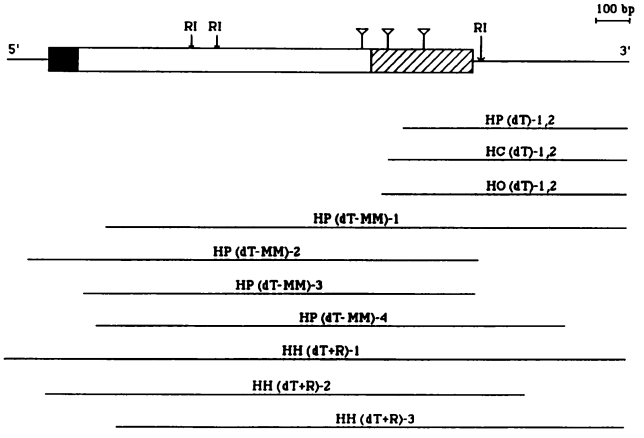

Fig. 3. Line diagram of the 1.9 kb OP-1 cDNA. The solid box indicates the putative signal peptide and the hatched box corresponds to the TGF-β like domain that contains the seven cysteine residues. Triangles indicate the three introns present in this area (genomic sequence information was available only for this region). The lines below the diagram indicate size of the clones obtained from various cDNA libraries. Human placenta, HP(dT); human calvaria, HC(dT); human osteosarcoma, HO(dT); human placenta (5'-stretch), HP(dT-MM); human hippocampus, HH(dT+R). 'dT' denotes the library was constructed by priming with oligo(dT), 'dT-MM' denotes the library was constructed in the presence of methyl mercuric hydroxide after priming with oligo(dT) and 'dT+R' denotes that the library was constructed by priming with a mixture of random primers and oligo(dT).

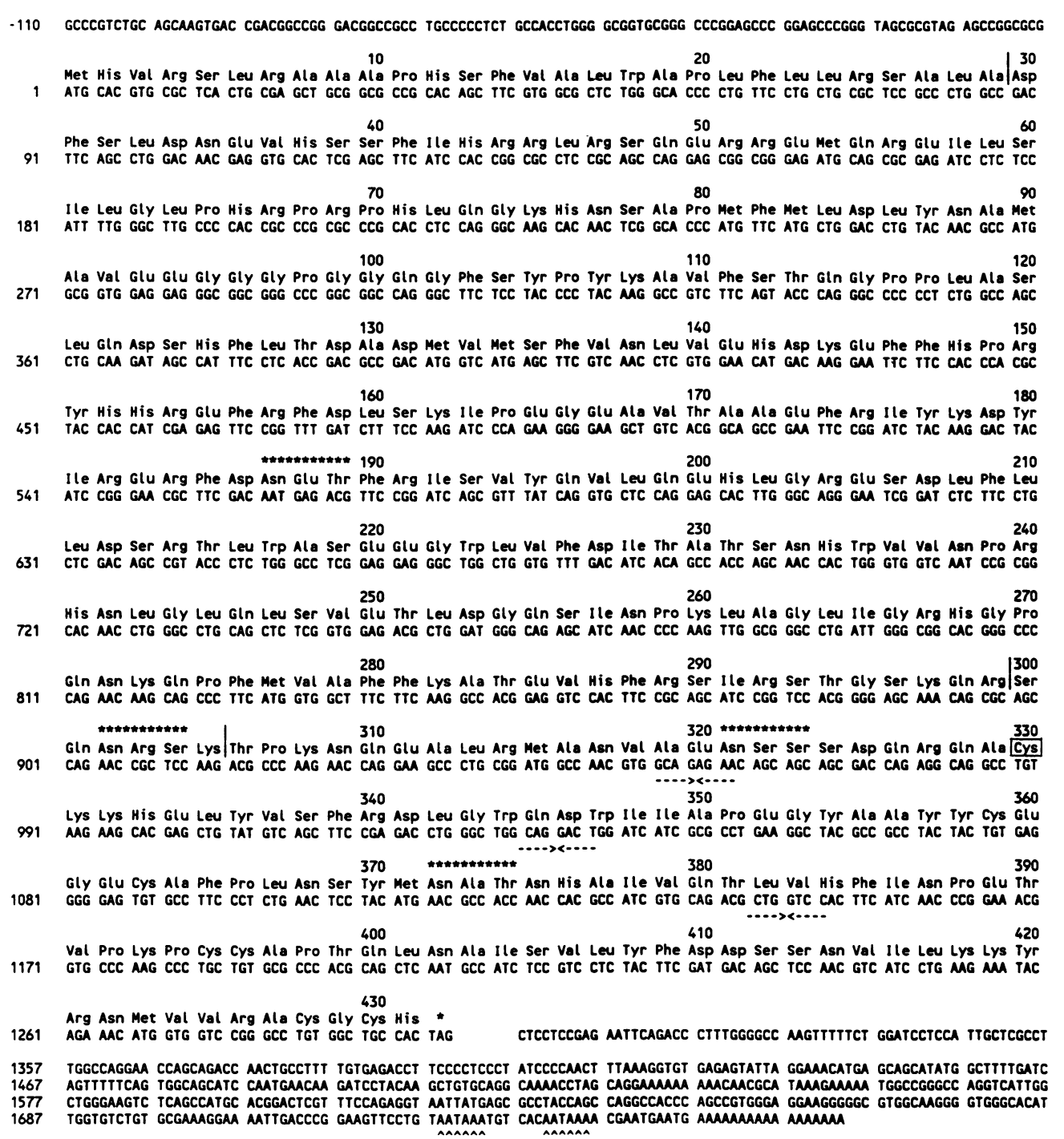

Fig. 4 Nucleotide sequence and the deducted amino acid sequence of OP-1 cDNA. The numbering of amino acids (above the sequence) and of nucleotides (on the left margin) starts with the first ATG of the OP-1 coding sequence. N-glycosylation sites are shown by *****. The putative signal peptidase cleavage site and the proposed maturation sites are marked by vertical lines, |. The three splice junctions in the C-terminal region as determined from partial genomic OP-1 sequence are indicated by ---><---. The first of seven cysteines is amino acid 330 (boxed). Two poly(A) addition signals (AATAAA) are seen at 1728 and 1740 bp, marked by arrowheads, ^ .

time). Consequently, we turned to cDNA libraries using the OP-1 exons as probes.

Screening of cDNA libraries

OP-1 cDNA was found in several cDNA libraries, including human placenta, human calvaria and human osteosarcoma. However, the OP-1 cDNA clones in these oligo(dT) primed cDNA libraries were all ~700 nucleotides long and comprised only the 3′ untranslated region and part of the TGF-β like domain. They did not include more than the last two exons and ended at various points in the 5′ portion of the second to the last exon (Figure 3). This observation was inconsistent with the average insert size of ~1.5 kb in these cDNA libraries. Hence, we concluded that cDNA synthesis

Table I. Identities between the 7-cysteine-TGF-β domains.

	OP-1	Vgr-1	BMP-2a	BMP-2b	DPP	Vg-1	Inhibin βA	BMP-3	Inhibin βB	TGF-β1
OP-1		88	60	58	58	57	44	42	39	34
Vgr-1	88		62	59	58	59	45	44	39	35
BMP-2a	60	62		92	74	58	45	49	44	32
BMP-2b	58	59	92		76	56	43	48	44	31
DPP	58	58	74	76		48	44	43	40	36
Vg-1	57	59	58	56	48		47	49	39	36
Inhibin βA	44	45	45	43	44	47		37	64	39
BMP-3	42	44	49	48	43	49	37		38	32
Inhibin βB	39	39	44	44	40	39	64	38		31
TGF-β1	34	35	32	31	36	36	39	32	31	

The percentage identities of the amino acid sequences of the TGF-β like proteins which are aligned in Figure 5 were calculated by the Align program (DNAstar, Inc.) based on the method of Needleman and Wunsch (1970). In this comparison substitutions are not scored. This table is arranged in order of similarities with OP-1 (the blank boxes in the table represent 100% identity.

from OP-1 mRNA may have aborted due to mRNA secondary structure at or near the boundary of the second to the last exon. Small variations in the exact location may have been due to different reaction conditions during construction of the libraries. Therefore, we screened cDNA libraries that were simultaneously primed by random primers and oligo(dT) or that were reverse transcribed under reaction conditions specifically designed to eliminate RNA secondary structure. These libraries (human hippocampus from Stratagene and human placenta from Clontech) yielded complete or nearly complete cDNA clones, predicting an mRNA size of ~2 kb. Figure 3 illustrates the length of different OP-1 cDNA clones.

Nucleotide sequence of OP-1 and putative protein product

The cDNA clones were sequenced using a combination of our unidirectional deletion technique (Özkaynak and Putney, 1987), synthetic internal primers, and subcloning of fragments. Moreover, various restriction fragments of OP-1 were expressed as fusion proteins in *Escherichia coli*, in order to verify the proper reading frame. The N-terminal portion in these fusions is a modified synthetic *trp* LE gene (Huston *et al.*, 1988), combined with a polylinker. Natural restriction sites in human OP-1 cDNA, namely *Sma*I, 26 nucleotides upstream of the first ATG (methionine), *Xho*I at amino acid 38, *Nco*I at amino acid 89, *Eco*RI at amino acid 173, and *Stu*I at amino acid 328, were spliced with the *E.coli trp* LE derived fusion partner at several restriction sites that suited the predicted reading frame. This resulted in expression of proteins that formed inclusion bodies in *E.coli*. The proteins, solubilized in SDS with DTT, migrated on SDS−PAGE with the expected mobilities, hence confirming the proposed reading frame at several locations (data not shown). In such experiments, gene fusions that lead to frameshift of *E.coli* leader protein and OP-1 produce polypeptides that contain no OP-1 sequence and appear truncated due to the occurrence of termination codons in the nonsense reading frames. Based on these data the cDNA sequence (Figure 4) predicts a primary translation product of ~49 kd which includes a 27−29 amino acid secretion signal peptide, a pro sequence of ~270 amino acids and the TGF-β domain. The deduced amino acid sequence thus obtained was compared with that of other members of the TGF-β family.

Comparison of OP-1 with mature domains of TGF-β related genes

The conserved C-terminal domain of OP-1 was compared with several closely related representatives of the TGF-β superfamily, namely Vgr-1, DPP, VG-1, BMP-2a. BMP-2b, BMP-3, inhibin βA and βB chain, and TGF-β1, which were also compared against each other by the method of Needleman and Wunsch (1970). The similarities, expressed as percentage of identical amino acids, are compiled in Table I, in order of resemblance to OP-1. The close similarity between OP-1 and Vgr-1 is most striking. This study also reveals that OP-1 and Vgr-1 are related to DPP (58%) as closely as they are related to Vg-1 (57% and 59%); in contrast, BMP-2a and BMP-2b show more homology to DPP (74% and 76%) than to Vg-1 (58% and 56%).

Sequence alignments of the C-terminal amino acid sequences are shown in Figure 5, in order of similarity to OP-1. From this alignment, it is apparent that the conservation of sequence is reduced in the central region of the TGF-β domain, where inhibin and TGF-β differ the most. Also, TGF-β and inhibin βB and βA chains distinguish themselves from the subfamily of Vg-1, DPP and BMP proteins in that they contain a double cysteine at the beginning of the 'TGF-β domain' (Derynck *et al.*, 1988). TGF-β, moreover, lacks the two basic residues that follow the first cysteine, a feature found in OP-1, Vgr-1, Vg-1, DPP, BMP and inhibin βA chain.

The pro region of OP-1

In addition to the 7-cysteine region we compared the pro regions for several pairs of proteins. A striking identity of 62% was observed when the pro regions downstream of amino acid 117 of OP-1 (*PLAS*) and amino acid 124 of Vgr-1 (*PLTS*), respectively, were aligned; for BMP-2a and BMP-2b 52% of the upstream region was matching. Much lower values were obtained for other upstream regions (OP-1 and BMP-2a: 22%, OP-1 and BMP-2b: 23%, OP-1 and BMP-3: 21%, Vgr-1 and BMP-2a: 23%, Vgr-1 and BMP-2b: 28%, Vgr-1 and BMP-3: 19%).

In contrast to the consensus C-terminal domains within the extended TGF-β family, the pro regions of the TGF-β family are not conserved, and can be used to distinguish otherwise closely related proteins. Whereas the pro regions of TGF-β1, 2, 3, 4 and 5 can be aligned with each other (Kondaiah *et al.*, 1990), they cannot be aligned with other

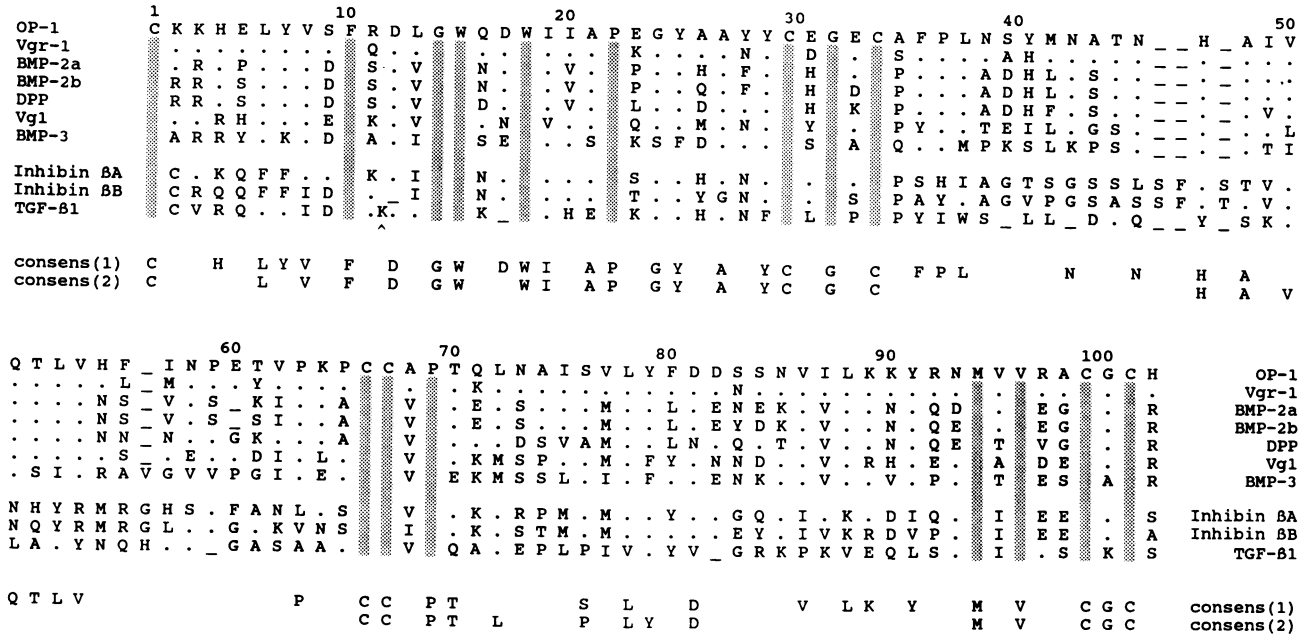

Fig. 5. Sequence alignments of 7-cysteine domains. The amino acid sequence of OP-1 is aligned with that of Vgr-1, BMP-2a, BMP-2b, DPP, inhibin βA chain, inhibin βB chain, and TGF-β1. Shaded bars indicate complete consensus among all sequences shown. Dots indicate amino acids which are identical to those of OP-1. Dashes (−) indicate gaps introduced for alignment. The line 'consensus 1' depicts sequence conservation for the close relatives of OP-1 (permitting one variation) and the line 'consensus 2' shows the overall agreement (permitting two exceptions).

TGF-β like proteins. Similarly, the pro regions of BMP-2a and BMP-2b can be aligned with each other (Wozney *et al.*, 1988), but not with BMP-3. The homology between the pro regions of BMP-2, BMP-3, DPP and Vgr-1 is very poor, with the exception of certain motifs: the penta amino acid sequence *NTVRS* in the pro regions of BMP-2a, BMP-2b and BMP-3; and the relatively conserved sequence MLDLY, also found in OP-1, located ~20−25 amino acids upstream of *NTVRS*. However, OP-1 and Vgr-1 show excellent alignments in certain parts of their pro regions (Figure 6 and see below).

While the alignments of pro regions among distantly related TGF-β like proteins are very poor they still conform in size. For example, the overall size of the pre-pro form of TGF-β1 is 391 amino acids (aa) (Derynck *et al.*, 1985), TGF-β2 is 414 aa (Webb *et al.*, 1988), TGF-β3 is 412 aa (Jakowlew *et al.*, 1988), and that of Vgr-1 is 438 aa (Lyons *et al.*, 1989a). Similarly, our cDNA clones of OP-1 yielded only one likely initiator ATG, predicting a primary translation product of 431 aa. The first methionine of OP-1 is followed by a signal peptide which fulfills the criteria of von Heijne (1986), including a basic N-terminal region, a central hydrophobic region and a more polar C-terminal region. The predicted signal peptidase cleavage site is located following amino acid 29, where an alanine at −1, leucine at −2, and alanine at −3 conform with the statistical observations. By comparison, the signal peptide of TGF-β1 spans 29 aa (Gentry *et al.*, 1988; Okada *et al.*, 1989) and that of TGF-β3 spans 18 or 23 aa (Jakowlew *et al.*, 1988). We have confirmed the entire open reading frame by expressing our cDNA from the *Sma*I site located 26 nucleotides upstream of the first ATG, through the complete coding region as a fusion protein. This resulted in the synthesis of a protein consistent with the predicted size for the putative precursor (data not shown).

Comparison of OP-1 and Vgr-1

As mentioned above, OP-1 and Vgr-1 show striking homology in their pro regions, similar to that found between BMP-2a and BMP-2b. However, alignment of OP-1 and Vgr-1 produces some notable disparities (see Figure 6). The alignment is continuous downstream from the sequence *PLAS* at amino acid 117, but there are two interruptions of homology in the upstream sequences which were initially not detected during alignment of the entire pro regions under standard alignment conditions. When shorter segments of the amino acid sequence were compared, the interrupted upstream homologies were found. The homology of OP-1 and Vgr-1 is first broken near position 70 in the OP-1 pre-pro sequence by a 29 amino acid insert in the Vgr-1 sequence which contains a string of 10 glutamine residues in the center. The homology then resumes partially for ~20 amino acids and is interrupted a second time for a stretch of ~25 amino acids in OP-1; these are countered by ~50 non-homologous amino acids of the Vgr-1 sequence. This is followed by very good alignment until about amino acid 300 of OP-1, whereupon the level of sequence conservation is relatively low for ~28 amino acids, up to the highly conserved 7-cysteine domain.

This comparison also revealed that the putative Vgr-1 signal peptide corresponds to internal sequences in the OP-1 gene (Figure 6). The OP-1 sequence lacks the first ATG of the Vgr-1 sequence (Lyons *et al.*, 1989a); instead, its initiation codon is found 50 amino acids further upstream. It is part of a typical signal peptide and no other methionine is found further upstream. The nucleotide sequence of OP-1 cDNA, in the region where it diverges from Vgr-1, is based on multiple independent cDNA clones, isolated from placenta and hippocampus cDNA libraries. Among these, three independent clones from human placenta and hippocampus libraries include the first methionine and seven independent

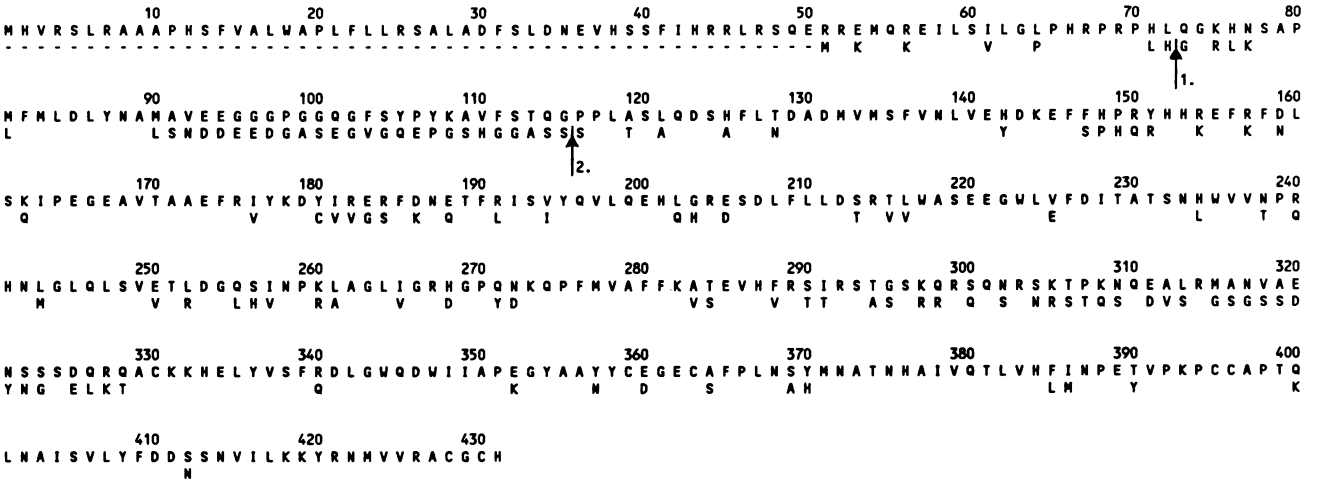

Fig. 6. Entire amino acid sequence of pre-pro OP-1 aligned with that reported for Vgr-1. The amino acid sequence of the OP-1 precursor is optimally aligned with that of Vgr-1 (Lyons *et al.*, 1989a) assuming two large inserts in the sequence of Vgr-1. While the inserts themselves are not shown in the alignment their respective locations are marked by arrows (†1 and †2). The first insert (†1) is found at ~40 amino acids following the proposed signal cleavage site of OP-1, namely L-Q-Q-P-Q-P-P-V-L-P-P-Q-Q-Q-Q-Q-Q-Q-Q-Q-Q-T-A-R-E-E-P-P-P, and the second insert (†2) is found another 40 amino acids downstream, namely S-Q-L-R-Q-P-S-P-G-Q-Q-H-S-L-N-R-K-S-L-L-A-P-G-P-G-G-G-A. The N-terminal region of OP-1 has no match in Vgr-1. Vgr-1 then seems to initiate ~20 amino acids following the putative cleavage site of OP-1, in this alignment. Note the sole cysteine in Vgr-1 that is located in the pro region and corresponds to amino acid 180 (tyrosine) of OP-1.

clones support our sequence where it diverges from the Vgr-1 sequence (see Figure 3 for the start of the cDNA clones). This rules out the possibility of a cloning artifact. The differences observed between Vgr-1 and OP-1 could also be due to alternative splicing. There is a precedent for this in TGF-β genes; a second mRNA containing an insert of 29 amino acids in the pro-region of TGF-β2 has been shown to be due to an alternative splicing event (Webb *et al.*, 1988).

OP-1 and Vgr-1 are the products of two distinct genes

TGF-β like molecules are highly conserved in different vertebrate species. Alignments of the homologous proteins from different species show marginal differences both in the mature and pro regions. Mature human TGF-β1 and murine TGF-β1 differ in only one amino acid (Derynck *et al.*, 1986), while human and simian TGF-β1 are identical in the mature region with only five amino acids difference in the pro region (Sharples *et al.*, 1987). In the mature regions, human and chicken TGF-β3 differ in only one amino acid (Derynck *et al.*, 1988; Jakowlew *et al.*, 1988; ten Dijke *et al.*, 1988); human and porcine TGF-β3 differ by two amino acids (Derynck *et al.*, 1988). Human and porcine βB subunits of activin differ by one amino acid (Mason *et al.*, 1985, 1989). Human and bovine MIS is similarly conserved with only three conservative amino acid changes in the 7-cysteine C-terminal region (Cate *et al.*, 1986). In contrast, OP-1 and Vgr-1 differ by 11 amino acids in the 7-cysteine domain in addition to the differences in the pro regions, discussed earlier. Also, three glycosylation sites are found in the entire pre-pro region of OP-1 (Figure 4), as opposed to four sites in Vgr-1. One site is lost due to an asparagine to aspartic acid change (Figure 6). These observed differences would indicate that OP-1 and Vgr-1 are not the products of the same gene, in man and mouse, respectively, but related to each other much like TGF-β1, TGF-β2 and TGF-β3, or in a relationship like BMP-2a and BMP-2b. This is also suggested by the identity index (Table I).

The maturation site of the OP-1 precursor

The precursor molecules of TGF-β and related proteins have protease cleavage or maturation sites located ~20 amino acids upstream of the highly conserved domains. This region, however, is not highly conserved, hence the N-termini of the mature proteins are not conserved. As mentioned earlier, the sequence immediately upstream of the 7-cysteine domain in OP-1 and Vgr-1 is also less conserved than upstream and downstream sequences. BMP-2a and BMP-2b are similar in this respect. The N-termini of mature TGF-β1, 2 and 3 start with the sequence *ALD* after cleavage from the pro region following the sequence *HRR* (TGF-β1) or *KKR* (TGF-β2 and 3) (Derynck *et al.*, 1985, 1988; Gentry *et al.*, 1988). The βA and βB subunits of activin are processed at *RRRRR* and *RIRKR*, respectively, and both start with *GLE* (Ling *et al.*, 1985; Mason *et al.*, 1985). Recombinant BMP-2a is cleaved after *KREKR* and mature BMP-2a starts with *QAKHK* (Wang *et al.*, 1990). Vgr-1 may be processed between *RRR* and *QQS* (Lyons *et al.*, 1989a). While OP-1 does not reveal a stack of basic residues, it contains the sequence *KQR* at a location corresponding to *RRR* in Vgr-1, and another potential site, *RSK*, three residues downstream. It is apparent that the N-terminal regions of mature TGF-β like proteins display a striking diversity, as seen with OP-1 and Vgr-1 or with BMP-2a and BMP-2b. The biological significance of this heterogeneity remains a challenging question.

No cysteines in the pro region of OP-1

The precursors of TGF-β1, 2, 3 and of TGF-β related proteins contain between one and five cysteines in their pro regions (Derynck *et al.*, 1988). Some of these cysteines are thought to be responsible for the dimer formation of the TGF-β precursor into a 125−160 kd protein complex (Miyazano *et al.*, 1988). Cysteines in the pro region have

also been implicated in the association with TGF-β binding protein, thus producing the latent form of TGF-β (Miyazano *et al.*, 1988; Okada *et al.*, 1989). While, in their pro regions, TGF-β1, 2, 3 and BMP-3 contain up to six cysteines, BMP-2 contains none; although the latter has one cysteine located in the signal peptide. OP-1 lacks cysteine residues in the pro region or pre region. Vgr-1, however, as an additional distinction from OP-1, contains one cysteine in the pro region. Therefore, OP-1 is the first example in the TGF-β family with a pre-pro region that is entirely devoid of cysteines. The lack of cysteines in the pro regions of OP-1 and BMP-2 precursors supports the relatedness of OP-1 and BMP-2. BMP-3, which is more distantly related, contains four cysteines in the pre region.

Recently, more extensive purification of bovine osteogenic protein resulted in separation of BMP-3 from the other components. Also, sequencing of these highly purified preparations of the bovine osteogenic protein revealed that the 18 kd subunit of osteogenic protein is the product of the OP-1 gene while the 16 kd subunit is encoded by the BMP-2 gene. Hence, the subunits could potentially exist as homo- and heterodimers (Sampath *et al.*, 1990).

Differential transcription of OP-1 mRNA

As mentioned earlier, we were able to find numerous OP-1 clones in several human cDNA libraries. Libraries made from placenta, osteosarcoma and hippocampus contained multiple OP-1 clones whereas a fetal liver library lacked it, indicating that OP-1 may not be expressed in liver. Similarly, other (Lyons *et al.*, 1989a) showed the absence of Vgr-1 message in murine liver. OP-1 mRNA was also detected in a human osteosarcoma cell line (U-2 OS) by Northern blot analysis. The size of the specific OP-1 transcript in this cell line is ~2.4 kb (Figure 7a and b). BMP-2b specific transcript was found to be slightly shorter, ~2.2 kb (Figure 7c). BMP-3 specific message was not detected, indicating that that particular cell line does not express BMP-3 at detectable levels (Figure 7d). The other members of this family have comparable size messages, the message of TGF-β1 is ~2.5 kb and TGF-β3 is ~3 kb (Derynck *et al.*, 1988; Jakowlew *et al.*, 1988), Vg-1 is ~2.7 kb (Weeks and Melton, 1987), and MIS is ~2.0 kb (Forage *et al.*, 1986).

It has been shown recently (Lyons *et al.*, 1989b) that high levels of *Vgr*-1 transcripts are seen in hypertrophic cartilage in the limb. Vgr-1 was also found in various epithelial tissues. The *Vgr*-1 gene was mapped to chromosome 13, near the congenital hydrocephalus (*ch*) locus. Homozygous ch^- embryos die shortly after birth, partly due to a failure to develop the subarachnoid space but these mutants have other defects, which include delayed growth of long bones and cartilaginous skull. This suggests the possibility that the *ch* mutation may indeed be a defect in the *Vgr*-1 gene. Although less likely, it cannot be excluded that the *ch* mutation may be a defect in the *Vgr*-1 receptor. An intriguing model for the temporal sequence of developmental signals by BMP-2, TGF-β, Vgr-1 and related molecules has been proposed in which BMP represents an early signal while Vgr-1 acts at a later stage (Lyons *et al.*, 1989b). By analogy, such a relationship may apply to BMP-2 and OP-1. Expression of the OP-1 gene and related genes in homologous and heterologous expression systems and evaluation of the expressed protein in the *in vivo* bone induction assay will shed new light on the detailed role of this protein in osteogenesis. In summary, the findings presented here, in concert with data on the biological activity of recombinant OP-1 (unpublished results) and the work of Lyons *et al.* on Vgr-1 (1989a, b), suggest that OP-1 and perhaps Vgr-1 play an important role in bone induction.

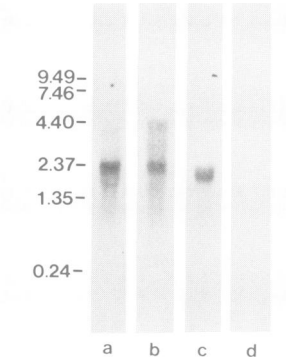

Fig. 7. Expression of OP-1, BMP-2b and BMP-3 in human osteosarcoma cells (U-2 OS line). The same filter was used for sequential hybridizations with labeled DNA probes. An RNA ladder (0.24–9.5 kb; BRL, Inc.) was used as size standard. **Lane a**: the OP-1 specific probe was the 0.79 kb *Eco*RI fragment from OP-1 cDNA (including the TGF-β like domain); **lane b**: the OP-1 specific probe was the 0.55 kb *Eco*RI fragment from OP-1 cDNA (5′ flanking, putative signal sequence and part of the pro region); **lane c**: the BMP-2b specific probe, derived from a genomic clone, was a fragment containing the last 350 nucleotides of the coding region, flanked by artifical *Bam*HI and *Pst*I restriction sites; **lane d**: the BMP-3 specific probe was the 0.31 kb *Pst*I–*Sph*I genomic fragment (part of the pro region and part of the TGF-β like domain).

Materials and methods

Peptide sequencing

Enriched bovine osteogenic protein with M_r of 30 kd was obtained from demineralized bovine bone (Sampath *et al.*, 1987, 1990). The peptides generated from the tryptic digestion of the reduced and carboxymethylated 30 kd bovine protein were purified by narrowbore C8 reverse phase HPLC and subjected to automated sequence analysis by Edman degradation using an Applied Biosystems 470A protein sequencer, equipped with an AB 120A on-line PTH analyzer, and programs supplied by the manufacturer. T2 and T4 indicate the assigned peak numbers of the HPLC fractions and the corresponding sequences are shown in the text.

Gene synthesis

The consensus gene was assembled from four pairs of synthetic oligonucleotides spanning strategically positioned assembly sites, flanked by cloning ends (*Bam*HI and *Eco*RI) (Huston *et al.*, 1988). The block assembly sites were chosen using the reverse translation and search program of Compugen, Inc., St Louis, MO. The complementary oligomers, ~80 bp in length, were prepared on a Milligen DNA synthesizer, Model 7500, by the phosphoramidite method and purified by electrophoresis on 10% acrylamide gels. The blocks were individually cloned into Bluescript vectors (Stratagene, La Jolla, CA) and thereafter assembled in three ligation steps.

Isolation of OP-1 genomic DNA

Approximately 5×10^5 plaques from a human genomic library (Clontech, Palo Alto, CA) were screened under non-stringent conditions (in 5 × SSPE, 10 × Denhardt's, 0.5% SDS at 50°C, overnight) with the consensus probe, shown in Figure 2 and described below, that had been labeled with ^{32}P (Feinberg and Vogelstein, 1984). The non-specific counts were washed off the filters in 1 × SSPE, 0.5% SDS at 50°C. This screening yielded 24 positive clones all of which were subsequently purified to homogeneity. Phage DNA from 24 clones was purified from isolated plaques using established procedures (Ausubel *et al.*, 1987). The genomic DNA fragments were subcloned into Bluescript vectors and sequenced by the chain termination method (Sanger *et al.*, 1977).

Isolation of OP-1 cDNA

Approximately 3×10^5 plaques from a human placenta '5'-Stretch' library based on λ gt11 (Clontech, Palo Alto, CA) were screened by hybridization in 40% formamide, 5 × SSPE, 5 × Denhardt's, 0.1% SDS at 37°C, overnight with an OP-1 genomic probe in which the introns were deleted by oligonucleotide mutagenesis (Kunkel, 1985). The respective exons had been identified after translation, on the basis of highly conserved amino acid patterns (e.g. cysteines) which were flanked by consensus splice signals. The filters were washed in 0.1 × SSPE, 0.1% SDS at 50°C. Phage DNA from isolated clones was prepared as described above and the inserts were subcloned into Bluescript vectors for sequence analysis. Approximately 4×10^5 plaques from a human hippocampus library based on 'Lambda ZAP' (Stratagene, La Jolla, CA) were screened with an OP-1 human placenta cDNA probe and positive clones were isolated for sequence analysis.

DNA sequencing

To obtain sequence information, a number of strategies, including exonuclease III mediated unidirectional deletion (Özkaynak and Putney, 1987), subcloning of restriction fragments, and synthetic oligonucleotide primers were used in the dideoxy chain termination method (Sanger *et al.*, 1977). The OP-1 cDNA was sequenced on both strands, using 7-deazaGTP in regions of initial ambiguity.

Northern hybridization

RNA from the cultured human osteosarcoma cells (U-2 OS cell line, ATCC HTB 96) was prepared using the guanidine thiocyanate procedure, essentially as described (Chirgwin *et al.*, 1979). Polyadenylated RNA (20 μg) was separated by electrophoresis on 1% agarose gels containing formaldehyde (Lehrach *et al.*, 1977) and blotted onto nitrocellulose (Thomas, 1980). Hybridizations were carried out in lucite cylinders in a roller bottle apparatus at 1 rev/min for 15 h at 37°C. The hybridization solution contained 40% formamide, 5 × SSPE, 5 × Denhardt's, 0.1% SDS, 100 μg/ml sonicated salmon testes DNA. The filters were washed in 0.1 × SSPE, 0.1% SDS, at 50°C. For additional hybridizations using the same filter, the annealed probe was removed by incubating the filter in 1 mM Tris–HCl, pH 7.5, 1 mM Na–EDTA, 0.1% SDS at 90°C.

Acknowledgements

We thank Charles Cohen for his interest and encouragement, John E.Coughlin, Robert M.Whetstone and Gail Clifford for excellent technical assistance and James Maliakal for the cell culture work. This work was supported by Stryker Corporation (Kalamazoo, Michigan).

References

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (eds) (1987) *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Interscience, New York.

Brunner,A.M., Gentry,L.E., Cooper,J.A. and Purchio,A.F. (1988) *Mol. Cell. Biol.*, **8**, 2229–2232.

Cate,R.L. *et al.* (1986) *Cell*, **45**, 685–698.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.F. (1979) *Biochemistry*, **18**, 5294–5299.

Derynck,R., Jarret,J.A., Chen,E.Y., Eaton,D.H., Bell,J.R., Assoian,R.K., Roberts,A.B., Sporn,M.B. and Goeddel,D.V. (1985) *Nature*, **316**, 701–705.

Derynck,R., Jarrett,J.A., Chen,E.Y. and Goeddel,D.V. (1986) *J. Biol. Chem.*, **261**, 4377–4379.

Derynck,R., Lindquist,P.B., Lee,A., Wen,D., Tamm,J., Graycar,J.L., Rhee,L., Mason,A.J., Miller,D.A., Coffey,R.J., Moses,H.L. and Chen,E.Y. (1988) *EMBO J.*, **7**, 3737–3743.

Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266–267.

Forage,R.G. *et al.* (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3091–3095.

Gentry,L.E., Lioubin,M.N., Purchio,A.F. and Marquardt,H. (1988) *Mol. Cell. Biol.*, **8**, 4162–4168.

Gray,A. and Mason,A. (1990) *Science*, **247**, 1328–1330.

von Heijne,G. (1986) *Nucleic Acids Res.*, **14**, 4683–4690.

Huston,J.S. *et al.* (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5879–5883.

Jakowlew,S.B., Dillard,P.J., Kondaiah,P., Sporn,M.B. and Roberts,A.B. (1988) *Mol. Endocrinol.*, **2**, 747–755.

Kondaiah,P., Sands,M.J., Smith,J.M., Fields,A., Roberts,A.B., Sporn,M.B. and Melton,D.A. (1990) *J. Biol. Chem.*, **15**, 1089–1093.

Kunkel,T.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 488–492.

Lehrach,H., Diamond,D., Wozney,J.M. and Boedtker,H. (1977) *Biochemistry*, **16**, 4743–4751.

Ling,N., Ying,S.-Y., Ueno,N., Esch,F., Denoroy,L. and Guillemin,R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7217–7221.

Ling,N., Ying,S.-Y., Ueno,N., Shimasaki,S., Esch,F., Hotta,M. and Guillemin,R. (1986) *Nature*, **321**, 779–782.

Lyons,K., Graycar,J.L., Lee,A., Hashmi,S., Linquist,P.B., Chen,E.Y., Hogan,B.L. and Derynck,R. (1989a) *Proc. Natl. Acad. Sci. USA*, **86**, 4554–4558.

Lyons,K.M., Pelton,R.W. and Hogan,B.L.M. (1989b) *Genes Dev.*, **3**, 1657–1668.

Mason,A.J., Hayflick,J.S., Ling,N., Esch,F., Ueno,N., Ying,S.-Y., Guillemin,R., Niall,H. and Seeburg,P.H. (1985) *Nature*, **318**, 659–663.

Mason,A.J., Berkemeier,L.M., Schmelzer,C.H. and Schwall,R.H. (1989) *Mol. Endocrinol.*, **3**, 1352–1358.

Mayo,K.E., Cerelli,G.M., Spiess,J., Rivier,J., Rosenfeld,M., Evans,R.M. and Vale,W. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5849–5853.

Miyazano,K., Hellman,U., Wernstedt,C. and Heldin,C.-H. (1988) *J. Mol. Biol.*, **263**, 6407–6415.

Needleman,S. and Wunsch,C. (1970) *J. Mol. Biol.*, **48**, 443–453.

Okada,F., Yamaguchi,K., Ichihara,A. and Nakamura,T. (1989) *FEBS Lett.*, **242**, 240–244.

Özkaynak,E. and Putney,S.D. (1987) *BioTechniques*, **5**, 770–773.

Padgett,R.W., St. Johnston,R.D. and Gelbart,W.M. (1987) *Nature*, **325**, 81–84.

Pepinsky,R.B., Sinclair,L.K., Chow,E.P., Mattaliano,R.J., Manganaro,T.F., Donahoe,P.K. and Cate,R.L. (1988) *J. Biol. Chem.*, **263**, 18961–18964.

Rebagliati,M.R., Weeks,D.L., Harvey,R.P. and Melton,D.A. (1985) *Cell*, **42**, 769–777.

Sampath,T.K. and Reddi,A.H. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7599–7603.

Sampath,T.K. and Reddi,A.H. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6591–6595.

Sampath,T.K., Muthukumaran,N. and Reddi,A.H. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 7109–7113.

Sampath,T,K., Coughlin,J.E., Whetstone,R.M., Banach,D., Corbett,C., Ridge,R.J., Özkaynak,E., Oppermann,H. and Rueger,D.C. (1990) *J. Biol. Chem.*, in press.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Segal,D. and Gelbart,W.M. (1985) *Genetics*, **109**, 119–143.

Sharples,K., Plowman,G.D., Rose,T.M., Twardzik,D.R. and Purchio,A.F. (1987) *DNA*, **6**, 239–244.

Tannahill,D. and Melton,D.A. (1989) *Development*, **106**, 775–785.

ten Dijke,P., Hansen,P., Iwata,K.K., Pieler,C. and Foulkes,J.G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 4715–4719.

Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201–5205.

Urist,M.R. (1965) *Science*, **150**, 893–899.

Vale,W., Rivier,J., Vaughan,J., McClintock,R., Corrigan,A., Woo,W., Karr,D. and Spiess,J. (1986) *Nature*, **321**, 776–779.

Wang,E.A., Rosen,V., Cordes,P., Hewick,R.M., Kriz,M.J., Luxemberg,D.P., Sibley,B.S. and Wozney,J.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 9484–9488.

Wang,E.A. *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2220–2224.

Webb,N.R., Madisen,L., Rose,T.M. and Purchio,A.F. (1988) *DNA*, **7**, 493–497.

Weeks,D.L. and Melton,D.A. (1987) *Cell*, **51**, 861–867.

Wozney,J.M., Rosen,V., Celeste,A.J., Mitsock,L.M., Whitters,M.J., Kriz,R.W., Hewick,R.M. and Wang,E.A. (1988) *Science*, **242**, 1528–1533.

Received on February 13, 1990; revised on April 19, 1990

Note added in proof

We have recently cloned and sequenced the murine OP-1 gene which proves that OP-1 and Vgr-1 are related but separate genes (Schnegelsberg,P.N.J. *et al.*, in preparation). The sequence data for the human OP-1 cDNA has been deposited in the EMBL data library under the accession number X51801.