Human cDNA Encoding a Novel TGF-β Superfamily Protein Highly Expressed in Placenta[1]

Midori Yokoyama-Kobayashi, Mihoro Saeki, Shingo Sekine, and Seishi Kato[2]

Sagami Chemical Research Center, 4-4-1 Nishi-Ohnuma, Sagamihara, Kanagawa 229

Received for publication, April 28, 1997

Recently, we developed a simple method for detecting a secretory signal sequence encoded by a cDNA fragment. In this study, we used this method to select cDNA clones encoding secretory proteins from a human full-length cDNA library. Full-sequencing analysis of the candidate clones revealed that one clone encoded a novel TGF-β superfamily protein. The clone encodes a protein of 308 amino acids of which the C-terminal region shows a characteristic feature of TGF-β superfamily proteins: seven conserved cysteine residues at the C-terminal preceded by a putative processing site composed of a basic amino acid repeat. The corresponding transcripts are highly expressed in the placenta, so the novel protein may play an important role in reproduction.

Key words: cDNA, cloning, secretory protein, TGF-β superfamily.

We are constructing a full-length cDNA bank from an entire set of human genes, called the Homo•Protein Bank (*1*). The first step is to select full-length cDNA clones from various human cDNA libraries by means of partial 5′-terminal sequencing followed by *in vitro* translation. The next step is to analyze the function of the protein encoded by each cDNA. There are several ways to assess the function of an unknown protein, one of which is to examine the localization of the protein.

Various sorting signals for the targeting of a protein to its working site have been reported: import signals to the endoplasmic reticulum, mitochondria, nucleus, peroxisomes, and so on. Since the signal sequence of a secreted protein or a type I membrane protein is located at the N-terminal of the protein chain and has the characteristic feature of a hydrophobic core region, we can search for these types of proteins in the bank by analyzing the N-terminal amino acid sequence encoded by an open reading frame determined by partial sequencing of the 5′ terminal of the full-length cDNA. In spite of the hydrophobic feature of the signal sequence, these proteins do not have any specific consensus sequence. Thus, it is impossible to determine whether the putative signal peptide can actually function as a secretory signal or not.

Recently, we established a system by which a cDNA clone encoding a signal peptide can be detected (*2*). In this system, the 5′-terminal fragment of a cDNA clone is ligated to a cDNA fragment of the urokinase-type plasminogen activator (u-PA) protease domain, and the resulting fusion gene is expressed in an adequate mammalian cell. If the cDNA fragment encodes a signal peptide and is ligated to the u-PA gene in frame, the fusion protein could be secreted into the medium and its u-PA activity could be easily detected in the medium. Using cDNA fragments encoding signal peptides of known secreted or type I membrane proteins, we showed that the system works well (*2*).

In this study, we applied this system to unknown cDNA clones obtained from a cDNA library of human fibrosarcoma cell line HT-1080, which was one of the cDNA libraries prepared from various human tissues and cell lines for the bank construction. As a result, we found several clones of which the signal sequences are functionally active. The entire sequences of these cDNA clones showed that one clone encodes a novel transforming growth factor-β (TGF-β) superfamily protein.

The TGF-β superfamily proteins show diverse biological activities involved in the control of cell proliferation, cell adhesion, phenotype, and so on (*3, 4*). Many members of the human TGF-β superfamily have been reported: TGF-β1 (*5*), TGF-β2 (*6, 7*), TGF-β3 (*8, 9*), MIS (*9*), inhibin α (*10*), inhibin βA (*11, 12*), inhibin βB (*11*), activin βC (*13*), BMP-2, BMP-3, BMP-4 (*14*), BMP-5, BMP-6 (*15*), BMP-7 (*15, 16*), BMP-8 (*17*), GDF-1 (*18*), and GDF-5 (*19*).

The TGF-β superfamily proteins have a common structural feature. The proteins are synthesized as a form of precursor which contains a signal sequence for secretion and a processing site preceded by a cluster of basic amino acids. An active domain composed of a carboxyl-terminal segment of 110-140 amino acids is released on cleavage at the processing site. The seven cysteine residues in the active domain are well conserved in the members of the TGF-β superfamily, but the identity of the remaining region of the active domain is low, 20-40%. The cDNA reported in this paper encodes a protein showing the above structural feature characteristic of the TGF-β superfamily and seems to be a novel member of the family because the amino acid identity of the putative C-terminal active domain is less than 35% compared with every member of the family.

[1] The nucleotide sequence reported in this paper has been submitted to the GenBank™/EMBL/DDBJ data bank under accession number AB000584.

[2] To whom correspondence should be addressed. Phone: +81-427-42-5091, Fax: +81-427-42-5091, e-mail: seishi@sagami.or.jp

Abbreviations: TGF, transforming growth factor; u-PA, urokinase-type plasminogen activator.

Downloaded from http://jb.oxfordjournals.org/ at Peking University on October 2, 2012

MATERIALS AND METHODS

cDNA Library—A cDNA library constructed from human fibrosarcoma cell line HT-1080 (*1*) was used for random sequencing to find cDNA clones encoding a signal sequence.

Sequencing—The sequencing reaction was performed by the dideoxy method using a fluorescent dye-labeled primer kit (Applied Biosystems). The reaction mixture was analyzed with a 373A automated DNA sequencer (Applied Biosystems). The deleted clones for sequencing of the entire region of the cDNA were prepared using a deletion kit (Takara Shuzo).

Signal Sequence Detection Assay—Functional analysis of a signal sequence was performed according to the previous report (*2*). A plasmid carrying a cDNA encoding a putative secretory protein was digested with *Not*I. The resulting fragment was deleted with exonuclease III and then S1 nuclease. After blunting with T4 polymerase, the fragments were further digested with *Hin*dIII. The cDNA fragments of 600-900 bp were isolated on a 1.5% agarose gel, and then inserted between the *Hin*dIII and *Eco*RV sites of pSSD1 which carries a cDNA fragment encoding the protease domain of u-PA preceded by the SV40 promoter (*2*). This ligation mixture was used for the transformation of *E. coli* JM109. f1 phage particles were prepared from the transformants and mixed with lipopolyamine, Transfectam (Sepracor). The liposomes were put into contact with monkey COS7 cells to introduce and express the chimeric gene (*20*). The fibrinolytic activity of the medium of the transfected cells was examined using the fibrin plate assay (*21*).

Northern Blot Hybridization—A cDNA fragment was labeled with [α-^{32}P]dCTP (Amersham) using a random primer labeling kit (Takara Shuzo). The Northern blots of human tissues (Clontech) were used as mRNA sources.

RESULTS

Screening of an Open Reading Frame Containing the Putative Signal Sequence—We determined the 5′-terminal 400 bp sequences of ~1,000 clones randomly selected from the HT-1080 cDNA library. Since the library was constructed using a capping method developed by us (*1*), more than 70% of the clones contained a full-length cDNA insert starting from a cap site. As found on a BLAST search using GenBank/EMBL/DDBJ, about 50% of the clones showed no matching with the sequences registered in the databases. Novel nucleotide sequences were converted into amino acid sequences. If there was an open reading frame starting with an initiation codon that fitted Kozak's rule (*22*), the hydrophobicity of the N-terminal amino acid sequence encoded by this open reading frame was examined using the Kyte and Doolittle method (*23*). Figure 1A shows the hydrophobicity profile of clone HP00269, one of the 10 candidates obtained using the above protocol.

Functional Analysis of the Putative Signal Sequence—To determine whether the N-terminal hydrophobic region encoded by this clone functions as a signal for secretion, 5′-terminal fragments of about 50-350 bp were prepared by deletion from the 3′-end of the cDNA, and inserted into the pSSD1 vector to express the chimeric gene between the target cDNA and the u-PA cDNA (Fig. 1B). Seven derivatives carrying the chimeric gene were introduced into COS7 cells and the fibrinolytic activity of the expression product in these cells was assayed. Two derivatives showed fibrinolytic activity (Fig. 1C). These results are consistent with the expectation that a fragment will be inserted into the vector in frame with a probability of 1/3. The above analysis suggested that the N-terminal part encoded by this clone functions as a signal peptide, and thus clone HP00269 must encode a secretory or type I membrane protein.

Full-Sequence Analysis—We have determined the entire sequence of clone HP00269. The nucleotide sequence and deduced amino acid sequence are shown in Fig. 2. This clone contains a 5′-noncoding region of 32 bp, a 3′-noncoding region of 242 bp and a poly(A) tail of 40 bp. The open

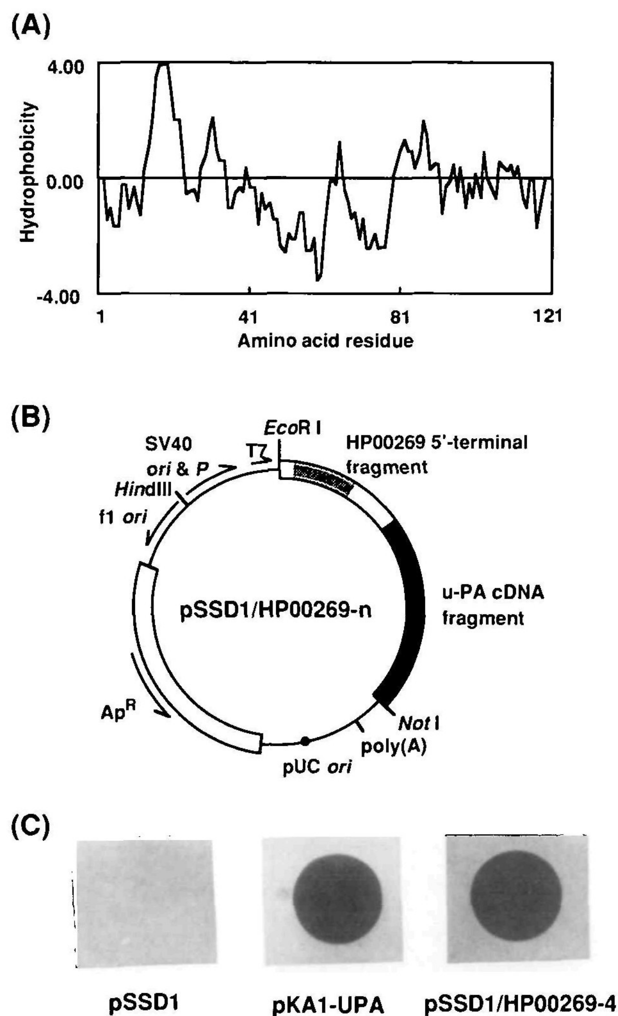

Fig. 1. **Functional analysis of the putative signal sequence.** (A) The hydrophobicity profile of the N-terminal encoded by clone HP00269. (B) The structure of the chimeric gene between the target cDNA fragment and the u-PA cDNA fragment. The shadowed box in the target cDNA represents the region encoding a putative signal peptide. (C) The fibrinolytic activity of the expression products of the chimeric genes. The pKA1-UPA contains the full-length cDNA of u-PA (*2*). pSSD1/HP00269-4 is one of the 7 chimeric clones in which the 3′ terminal region of HP00269 cDNA was deleted. The black circle represents the lytic part produced by the fibrinolytic activity of the medium placed on the fibrin plate.

Downloaded from http://jb.oxfordjournals.org/ at Peking University on October 2, 2012

reading frame encodes a protein of 308 amino acid residues with a calculated M_r of 34,167. The region of the initiation codon is in good agreement with the consensus sequence (*22*). The putative poly(A) addition signal is located at position 1,080. The coding region has a unique *Not*I site surrounded by a GC-rich region, which made it difficult to sequence this region. The nucleotide sequence database search showed that there are many human truncated ESTs but not a full-sequence corresponding to this cDNA.

Amino Acid Sequence Analysis—The protein database search revealed that this protein is novel and exhibits similarity to members of the TGF-β superfamily, as shown in Fig. 3. The similarity is in the C-terminal half of the protein, while the N-terminal half showed no similarity with not only TGF-β family members but also any other known proteins. The positions of seven cysteines in the C-terminal half were highly conserved. This protein contains a putative proteolytic processing site (Arg-Arg-Arg)

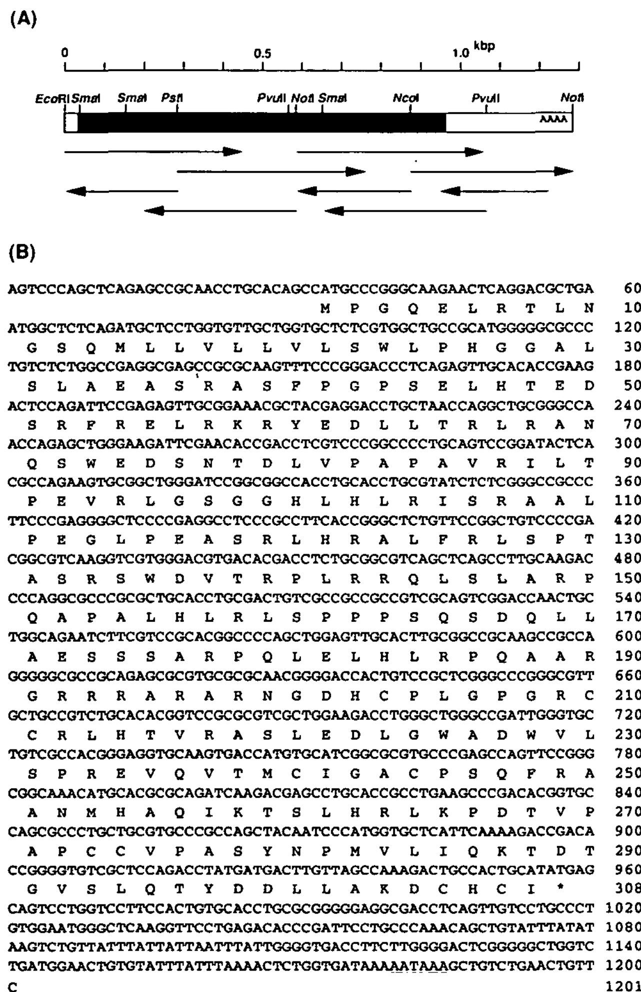

Fig. 2. **Structure of the cDNA.** (A) Restriction map of the cDNA and the sequencing strategy. The solid box represents the open reading frame. (B) The nucleotide sequence of the cDNA and the deduced amino acid sequence. The putative poly (A) addition signal is underlined.

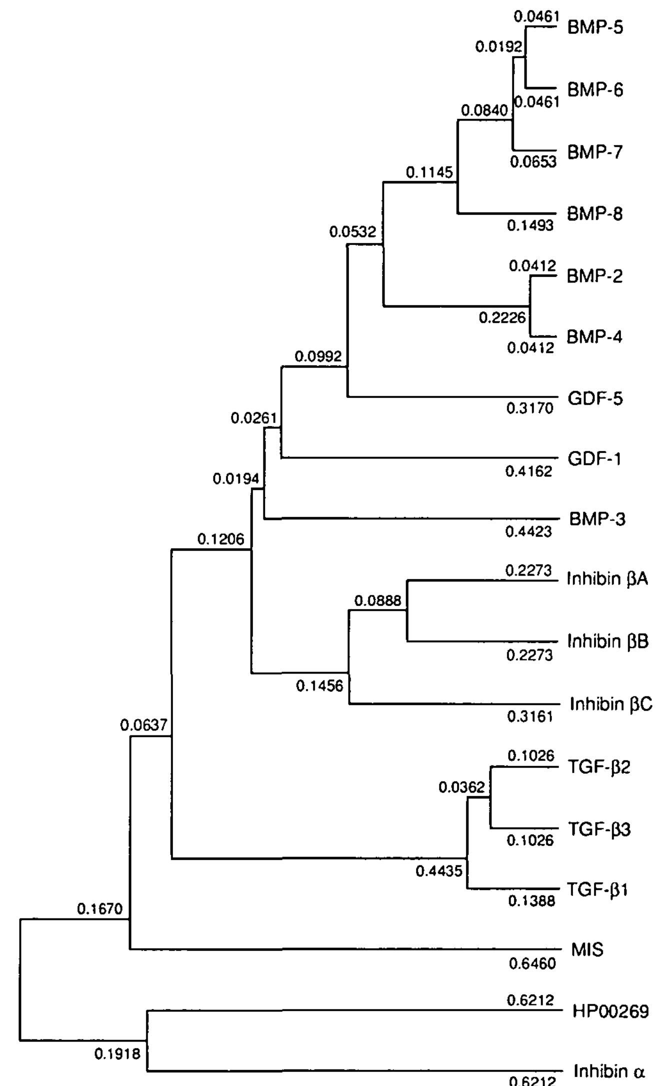

Fig. 4. **Phylogenetic tree obtained for human members of the TGF-β superfamily protein.** The analysis was performed on the C-terminal sequences in Fig. 3 using the software package, GENETYX-MAC (Software Development). The numbers at the branches indicate the distance.

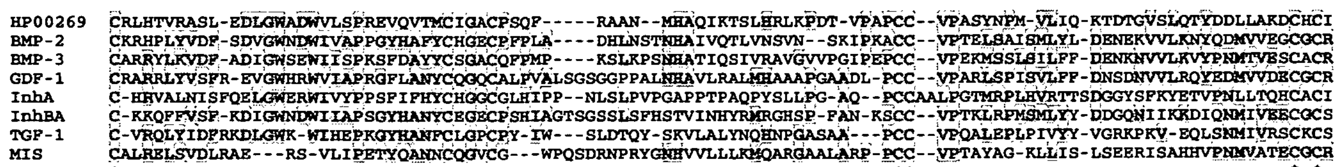

Fig. 3. **Alignment of TGF-β superfamily protein amino acid sequences.** Identical amino acids shared by more than three sequences are shaded. Asterisks under the sequences indicate cysteine residues conserved in all proteins.

Downloaded from http://jb.oxfordjournals.org/ at Peking University on October 2, 2012

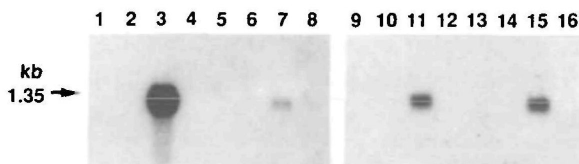

Fig. 5. **Northern blot hybridization.** 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas; 9, spleen; 10, thymus; 11, prostate; 12, testis; 13, ovary; 14, small intestine; 15, colon; 16, peripheral blood leukocytes.

at position 192 and a potential *N*-linked glycosylation site (Asn-Gln-Ser) at position 70. Figure 4 shows the results of phylogenetic analysis of the human TGF-β family members. It indicates that HP00269 diverged at an early stage of evolution and falls into a novel subgroup distinct from the decapentaplegic-Vg-related, activin, and TGF-β subfamilies.

Expression Pattern—We compared the mRNA level of this novel protein in various human tissues by Northern analysis. A hybridization signal with a size of about 1.3 kb was found in prostate, colon, and kidney, and an extremely strong signal in placenta, as shown in Fig. 5.

DISCUSSION

We developed a novel method for assessing a functional signal peptide and applied it to the selection of a novel secretory protein from a full-length cDNA bank. This is the first report that novel secretory protein cDNAs can be cloned by the above strategy. In order to screen cDNAs encoding a secretory protein, similar methods involving a fusion protein with a reporter protein such as the IL-2 receptor or invertase have been reported (*24, 25*). These methods were adopted to screen a cDNA fragment encoding a signal peptide region as a first step. The success of these methods is affected by probability, but this disadvantage was overcome by means of positive selection by the invertase fusion method (*25*). Even though a positive clone was obtained, a full-length cDNA corresponding to the cloned cDNA fragment remains to be screened. This process is time- and labor-consuming. When the Homo•Protein cDNA Bank is completed in the near future, our method would increase its merit.

The obtained cDNA is assumed to encode a novel TGF-β superfamily member based on the sequence similarity with known members. The encoded protein shows the characteristic feature of the TGF-β superfamily: a signal sequence for secretion and seven conserved cysteine residues at the C-terminal preceded by a putative processing site composed of a basic amino acid repeat. Phylogenetic analysis indicated that this protein may represent a new subgroup within the superfamily. Taking the high expression in placenta into account, this protein has been named TGF-βPL.

The expression pattern of this protein also suggests that it is different from the other members. Strong expression of this mRNA was observed in placenta, with weak expression in prostate, colon, and kidney. It is unknown what roles this protein plays in these tissues. Some members of the TGF-β family are closely related to some gonadotropic and sex steroid hormones. For example, activin and inhibin modulate the release of follicle-stimulating hormone (*26, 27*). TGF-β1, whose secretion by immature sertoli cells is dependent on follicle-stimulating hormone, might control the synthesis of some steroid hormones (*28*). Similarly, this novel protein might have some biological activity related to reproduction.

The results of Northern analysis explained why most of the hundreds of ESTs registered in the database were derived from placenta cDNA libraries. Nucleotide sequence comparison between our cDNA and these ESTs showed that half of the ESTs are derived from the 5′-region ending at the *Not*I site located at the middle of the cDNA and the remaining half from the 3′-region starting from the *Not*I site to poly(A). This kind of truncated cDNAs might be due to the *Not*I digestion of the synthesized cDNA for subcloning at the *Not*I site of the vector. *Not*I sites are thought to be rare in cDNAs, but we observed many cDNAs possessing *Not*I sites.

We have shown the effectiveness of the signal detection system for assessing the secretory function of a signal peptide encoded by a novel cDNA. During the selection of cDNA clones encoding a secretory protein from the cDNA library using this system, we found a novel TGF-β superfamily protein that is highly expressed in placenta. The obtained cDNA can be used to produce a recombinant protein. Using this protein we can determine its biological activity, which will help to elucidate the role of this protein in the placenta.

Downloaded from http://jb.oxfordjournals.org/ at Peking University on October 2, 2012

REFERENCES

1. Kato, S., Sekine, S., Oh, S.-W., Kim, N.-S., Umezawa, Y., Abe, N., Yokoyama-Kobayashi, M., and Aoki, T. (1994) Construction of a human full-length cDNA bank. *Gene* **150**, 243-250
2. Yokoyama-Kobayashi, M., Sugano, S., Kato, T., and Kato, S. (1995) A signal sequence detection system using secreted protease activity as an indicator. *Gene* **163**, 193-196
3. Massague, J. (1990) The transforming growth factor-β family. *Annu. Rev. Cell Biol.* **6**, 597-641
4. Kingsley, D.M. (1994) The TGF-β superfamily: new members, new receptors, and new genetic tests of function in different organisms. *Genes Dev.* **8**, 133-146
5. Derynck, R., Jarrett, 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) Human transforming growth factor-β complementary DNA sequence and expression in normal and transformed cells. *Nature* **316**, 701-705
6. Madisen, L., Webb, N.R., Rose, T.M., Marquardt, H., Ikeda, T., Twardzik, D., Seyedin, S., and Purchio, A.F. (1988) Transforming growth factor-β2: cDNA cloning and sequence analysis. *DNA* **7**, 1-8
7. de Martin, R., Haendler, B., Hofer-Wabinek, R., Gaugitsch, H., Wrann, M., Schlüsener, H., Seifert, J.M., Bodmer, S., Fontana, A., and Hofer, E. (1987) Complementary DNA for human glioblastoma-derived T cell suppressor factor, a novel member of the transforming growth factor-β gene family. *EMBO J.* **6**, 3673-3677
8. 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) A new type of transforming growth factor-β, TGF-β3. *EMBO J.* **7**, 3737-3743
9. Cate, R.L., Mattaliano, R.J., Hession, C., Tizard, R., Farber, N.M., Cheung, A., Ninfa, E.G., Frey, A.Z., Gash, D.J., Chow, E.P., Fisher, R.A., Bertonis, J.M., Torres, G., Wallner, B.P., Ramachandran, K.L., Ragin, R.C., Manganaro, T.F., MacLaughlin, D.T., and Donahoe, P.K. (1986) Isolation of the bovine and human genes for Mullerian inhibiting substance and expression of

the human gene in animal cells. *Cell* **45**, 685-698
10. Mayo, K.E., Cerelli, G.M., Spiess, J., Rivier, J., Rosenfeld, M.G., Evans, R.M., and Vale, W. (1986) Inhibin A-subunit cDNAs from porcine ovary and human placenta. *Proc. Natl. Acad. Sci. USA* **83**, 5849-5853
11. Mason, A.J., Niall, H.D., and Seeburg, P.H. (1986) Structure of two human ovarian inhibins. *Biochem. Biophys. Res. Commun.* **135**, 957-964
12. Murata, M., Eto, Y., Shibai, H., Sakai, M., and Muramatsu, M. (1988) Erythroid differentiation factor is encoded by the same mRNA as that of the inhibin β_A chain. *Proc. Natl. Acad. Sci. USA* **85**, 2434-2438
13. Hötten, G., Neidhardt, H., Schneider, C., and Pohl, J. (1995) Cloning of a new member of the TGF-β family: a putative new activin β_C chain. *Biochem. Biophys. Res. Commun.* **206**, 608-613
14. 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) Novel regulators of bone formation: molecular clones and activities. *Science* **242**, 1528-1534
15. Celeste, A.J., Iannazzi, J.A., Taylor, R.C., Hewick, R.M., Rosen, V., Wang, E.A., and Wozney, J.M. (1990) Identification of transforming growth factor β family members present in bone-inductive protein purified from bovine bone. *Proc. Natl. Acad. Sci. USA* **87**, 9843-9847
16. Özkaynak, E., Rueger, D.C., Drier, E.A., Corbett, C., Ridge, R.J., Sampath, T.K., and Oppermann, H. (1990) OP-1 cDNA encodes an osteogenic protein in the TGF-β family. *EMBO J.* **9**, 2085-2093
17. Özkaynak, E., Schnegelsberg, P.N.J., Jin, D.F., Clifford, G.M., Warren, F.D., Drier, E.A., and Oppermann, H. (1992) Osteogenic protein-2. A new member of the transforming growth factor-β superfamily expressed early in embryogenesis. *J. Biol. Chem.* **267**, 25220-25227
18. Lee, S.-J. (1991) Expression of growth/differentiation factor 1 in the nervous system: conservation of a bicistronic structure. *Proc. Natl. Acad. Sci. USA* **88**, 4250-4254
19. Hötten, G., Neidhardt, H., Jacobowsky, B., and Pohl, J. (1994) Cloning and expression of recombinant human growth/differentiation factor 5. *Biochem. Biophys. Res. Commun.* **204**, 646-652
20. Yokoyama-Kobayashi, M. and Kato, S. (1994) Recombinant f1 phage-mediated transfection of mammalian cells using lipopolyamine technique. *Anal. Biochem.* **223**, 130-134
21. Astrup, T. and Müllertz, S. (1952) The fibrin plate method for estimating fibrinolytic activity. *Arch. Biochem. Biophys.* **40**, 346-351
22. Kozak, M. (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125-8148
23. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132
24. Tashiro, K., Tada, H., Heilker, R., Shirozu, M., Nakano, T., and Honjo, T. (1993) Signal sequence trap: a cloning strategy for secreted proteins and type I membrane proteins. *Science* **261**, 600-603
25. Klein, R.D., Gu, Q., Goddard, A., and Rosenthal, A. (1996) Selection for genes encoding secreted proteins and receptors. *Proc. Natl. Acad. Sci. USA* **93**, 7108-7113
26. Vale, W., Rivier, J., Vaughan, J., McClintock, R., Corrigan, A., Woo, W., Karr, D., and Spiess, J. (1986) Purification and characterization of an FSH releasing protein from porcine ovarian follicular fluid. *Nature* **321**, 776-779
27. Ling, N., Ying, S.-Y., Ueno, N., Shimasaki, S., Esch, F., Hotta, M., and Guillemin, R. (1986) Pituitary FSH is released by a heterodimer of the β-subunits from the two forms of inhibin. *Nature* **321**, 779-782
28. Benahmed, M., Cochet, C., Keramidas, M., Chauvin, M.A., and Morera, A.M. (1988) Evidence for a FSH dependent secretion of a receptor reactive transforming growth factor β-like material by immature sertoli cells in primary culture. *Biochem. Biophys. Res. Commun.* **154**, 1222-1231

Downloaded from http://jb.oxfordjournals.org/ at Peking University on October 2, 2012