Vol. 189, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1992 Pages 356-362

MOLECULAR CLONING AND CHARACTERIZATION OF THE HUMAN AND PORCINE TRANSFORMING GROWTH FACTOR-β TYPE III RECEPTORS

Anita Morén, Hidenori Ichijo, and Kohei Miyazono*

Ludwig Institute for Cancer Research, Box 595 Biomedical Center
S-751 24 Uppsala, Sweden

Received October 12, 1992

Summary: Full-length cDNAs for the transforming growth factor-β (TGF-β) type III receptors were isolated from porcine uterus and human placenta cDNA libraries. The human TGF-β type III receptor coding region encodes a protein of 849 amino acids with a single transmembrane domain and a short stretch of the intracellular domain. Potential glycosaminoglycan attachment sites were found in the extracellular domain. The overall amino acid sequence identities with those of the porcine and rat TGF-β type III receptors were 83 % and 81 %, respectively. A high degree of sequence conservation was observed in the transmembrane and intracellular domains, which also have sequence similarity with human endoglin. In addition, two portions with 29 and 52 amino acids in the extracellular domain were found to be substantially similar with human endoglin.

Transforming growth factor-β (TGF-β) is a family of multifunctional proteins, which are involved in the cellular proliferation, differentiation, chemotaxis, and extracellular matrix production (reviewed in Refs. 1, 2). Five different isoforms (TGF-β1 to -β5) with high sequence similarity have been identified. The amino acid sequence of each isoform is highly conserved in different species; e.g. mouse TGF-β1 has only one amino acid difference from the human counterpart (3).

TGF-βs exert their effects through binding to the cell surface receptors. Three distinct receptors have been identified in most cell types, i.e. type I (53 kDa), type II (70-80 kDa) and type III (300 kDa) receptors (4). The TGF-β type III receptor may not be directly involved in the signal transduction, since TGF-β inhibits the growth of the cells which do not express the type III receptor (5, 6). Thus, the function of the TGF-β type III receptor is not fully elucidated.

The TGF-β type III receptor is a membrane proteoglycan, to which glycosaminoglycan (GAG) chains are attached (7, 8). Therefore, the type III receptor has also been denoted betaglycan (9). TGF-β binds to the 120 kDa core protein of the type III receptor. A soluble form of the TGF-β type III receptor, which may be produced through

*To whom correspondence should be addressed.

0006-291X/92 $4.00

the proteolysis of the membrane-bound form (9), exists in serum and in the extracellular matrix.

A cDNA for the rat TGF-β type III receptor has recently been cloned (10, 11). It is synthesized as a 853 amino acid protein with a single transmembrane domain. The intracellular domain of the type III receptor is composed of only 43 amino acids with a high content of serine and threonine residues (42 %) (10). The transmembrane and intracellular parts of the type III receptor have high amino acid sequence similarity (71 %) with human endoglin, an RGD-containing protein, that is observed in endothelial cells, hematopoietic progenitor cells, and mesangial cells in the kidney (12, 13).

In this communication, we show the molecular cloning of porcine and human TGF-β type III receptors and compare their structures with that of the rat receptor.

MATERIALS AND METHODS

Preparation of mRNA and Construction of cDNA Libraries. RNA was prepared from porcine uterus by using the guanidium isothiocyanate method (14). mRNA was selected by polyAT tract system (Promega, Madison, WI) as described by the manufacturer's instruction. An oligo (dT)-primed cDNA library with 1.25 x 10^6 unamplified cDNA clones was prepared by RiboClone cDNA synthesis system (Promega) and λgt10 in vitro packaging kit (Amersham, UK). The human placenta λZAP II cDNA library was provided by Hideo Toyoshima and Hisamaru Hirai, University of Tokyo, Japan.

Polymerase Chain Reaction (PCR) and Cloning of Porcine TGF-β Type III Receptor. Nucleotide sequences 2567-2580 and 2868-2889 of the rat TGF-β type III receptor (10, 11) were chosen to design the degenerated PCR primers. PCR was performed using first-strand cDNA prepared from porcine uterus and Taq polymerase (Perkin Elmer Cetus, Norwalk, CT). After 25 cycles of PCR each composed of 94°C (1 min), 55°C (2 min) and 72°C (1 min), a 3 % aliquot of the material was taken and processed to the second round of PCR amplification. Thereafter, a 0.5 % aliquot of the material was processed to the third round of PCR amplification. DNA was separated by 1.7 % agarose gel; a fragment of expected size was then recovered from the gel, and subcloned into pUC19 (15).

A porcine uterus λgt10 cDNA library was screened with the PCR probe labeled by the Megaprime DNA labeling system (Amersham). Hybridization to nitrocellulose replica filters and purification of positive bacteriophages were performed as described previously (16). A clone with a 4.0 kb insert denoted PU-5 was identified and subcloned into pUC19. Nucleotide sequencing (17) was performed on both strands in the translated region using Sequenase (United States Biochemical Corporation, Cleveland, OH).

Cloning of the Human TGF-β Type III Receptor. PU-5 was digested by *Bam*HI and *Eco*RI, and a 1.0 kb fragment which corresponded to the N-terminal part of the porcine TGF-β type III receptor was obtained. A human placenta λZAP II cDNA library with 5 x 10^5 independent clones was screened with the PU-5 fragment labeled by the Megaprime DNA labeling system. A phagemid pBluescript SK(-) with a 4.2 kb insert, denoted TIIIR-2, was obtained by excision from the lambda vector with helper phage R408 as described by Short et al. (18). Nucleotide sequencing was performed on both strands in the translated region.

RESULTS AND DISCUSSION

Cloning of the Porcine and Human TGF-β Type III Receptor cDNAs. Porcine TGF-β type III receptor cDNA was obtained from a porcine uterus λgt10 cDNA

library probed by a PCR product corresponding to the C-terminal part of the rat TGF-β type III receptor (nucleotides 2567-2889). The longest clone with a 4.0 kb insert, designated PU-5, was isolated and sequenced (the deduced amino acid sequence is shown in Fig. 2). For the cloning of the human TGF-β type III receptor, a *Bam*HI and *Eco*RI digested fragment of PU-5, which corresponded to the N-terminal part of the porcine TGF-β type III receptor, was used as a probe. From a human placenta λZAP II cDNA library, a clone with a 4.2 kb insert, denoted TIIIR-2, was obtained and analyzed further.

Sequence of the Human TGF-β Type III Receptor. Sequencing of TIIIR-2 yielded a 4213 bp fragment consisting of an open reading frame of 2547 bp (849 amino acids), flanked by a 5' untranslated sequence of 621 bp, and a 3' untranslated sequence of 1045 bp. The nucleotide sequence and predicted amino acid sequence are shown in Fig. 1. No polyadenylation signal was found in the 3' untranslated region of TIIIR-2, indicating that the cDNA was incomplete in the 3' end. Two ATG codons were found in the 5' part of the open reading frame (nucleotides 622-624 and 658-660). Only the first ATG codon fulfilled the rules for the translational initiation (19), and was followed by a hydrophobic leader sequence. An in-frame stop codon was observed at nucleotides 583-585. Therefore, the first ATG codon is likely to be used as a translation initiation site. Amino acids 6-21 have the characteristics of a hydrophobic leader sequence (20). The calculated molecular weight of the primary translated product of the human TGF-β type III receptor without signal sequence is 91,154.

In the extracellular domain, 6 potential N-glycosylation sites and 5 potential GAG attachment sites (Ser-Gly sequences) could be found. Two of the GAG attachment sites at Ser532-Gly and Ser543-Gly are surrounded by acidic amino acids; thus, they are the most preferable sites for GAG attachment (21).

The human TGF-β type III receptor has a single hydrophobic transmembrane domain at amino acids 782-806. The intracellular part is composed of only 43 amino acids. Similar with the rat TGF-β type III receptor (10, 11), the intracellular part of the human type III receptor is rich in serine and threonine residues (42 %).

Comparison of the TGF-β Type III Receptors from Human, Porcine and Rat. The amino acid sequence of the human TGF-β type III receptor was compared with those of the porcine (clone PU-5) and rat (10, 11) counterparts (Fig. 2). The overall amino acid sequence identities of the human sequence with those of porcine and rat were 83 % and 81 %, respectively. Of the 6 potential N-linked glycosylation sites in the human clone, 5 are common in all three species. The most preferable GAG attachment sites (Ser532-Gly and Ser543-Gly in human) are well conserved in all three species. Among the 3 other possible GAG attachment sites, only one was found in all three species (Ser50-Gly in human). The 17 cysteine residues are all conserved; the porcine type III receptor has an additional cysteine residue at position 5. López-Casillas et al. (10)

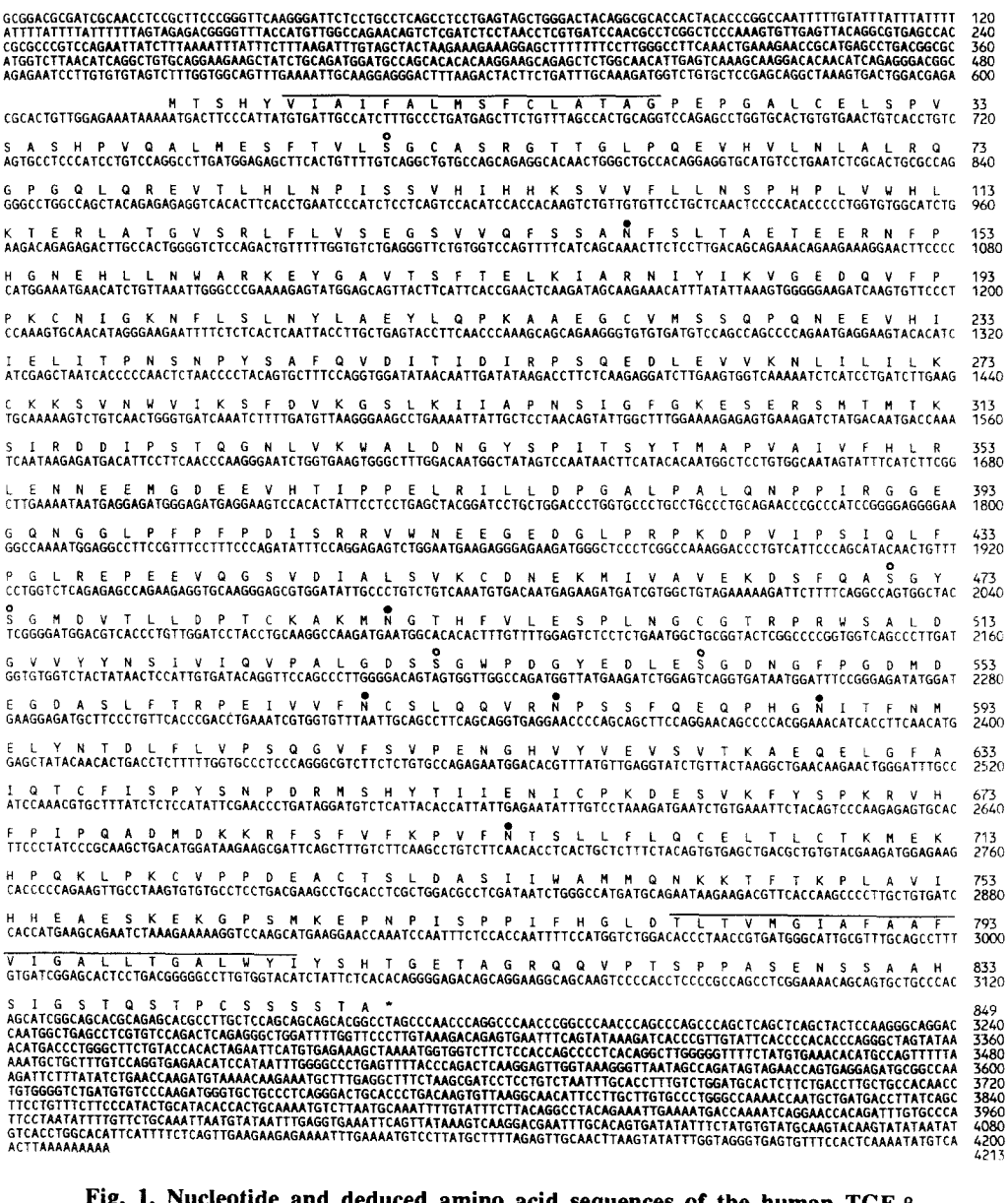

Fig. 1. Nucleotide and deduced amino acid sequences of the human TGF-β type III receptor. The hydrophobic leader sequence and putative transmembrane domain are overlined. Possible sites for N-glycosylation (●) and GAG attachment (o) are indicated. The stop codon which ends the open reading frame is marked by an asterisk.

reported two potential proteolytic cleavage sites, which may generate a soluble-form of the TGF-β type III receptor. The dibasic sequence (Lys743-Lys in human) was found in all three species. However, another cleavage site found in rat (Leu-Ala-Val-Val sequence), which is cleavable by an elastase-like activity (22), is not conserved in the human and porcine sequences. The proline-rich region (Pro-Ile-Pro-Pro-Pro-Pro) in rat is also less conserved.

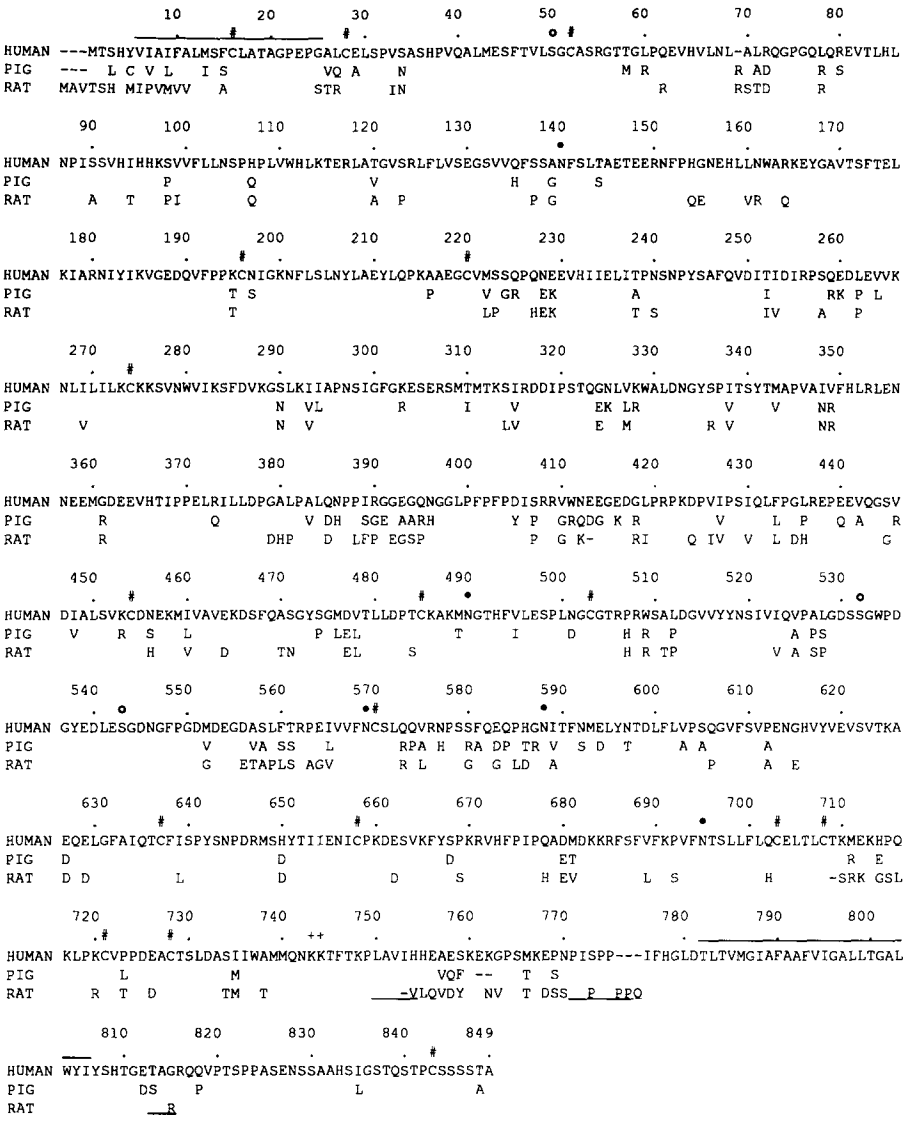

Fig. 2. Alignment of the amino acid sequences of the human, porcine and rat TGF-β type III receptors. In the porcine and rat (10, 11) sequences, only the amino acid residues which are different from those of human are shown. Numbers indicate the amino acid residues of the human sequence. (-) indicates gaps introduced to align the sequences. The hydrophobic leader sequence and transmembrane domain are overlined. N-glycosylation sites (•), potential GAG attachment sites (o), cysteine residues (#), and the dibasic potential cleavage site (+), which are common in all three species, are indicated. Structures found only in the rat receptor, i.e.,the putative cleave site by an elastase-like activity, the proline-rich region, and the putative protein kinase C phosphorylation site, are underlined.

The transmembrane and intracellular domains have a high degree of sequence conservation; only one amino acid difference was observed between human and rat in this region. However, the consensus site for phosphorylation by protein kinase C (Thr-Ala-Arg) (23) is not conserved in the human and porcine sequences.

```
Type III Rec   67   LNLALRQGPGQLQREVTLHLNPISSVHIH   95
                    : :      :   ::: : :   ::: .:
Endoglin       67   LTLQASKQNGTWPREVLLVLSVNSSVFLH   95

                     #                          #   •          #
Type III Rec  453   KCDNEKMIVAVEKDSFQASGYSGMDVTLLDPTCKAKMNGTHFVLESPLNGCG   504
                    ::  . : . . :.      .  .:  ::.: :   :  ::: :   ::
Endoglin      349   KCADDAMTLVLKKELVAHLKCTITGLTFWDPSCEAEDRGDKFVLRSAYSSCG   400
                     #                  #          #          #
                                                                        #
Type III Rec  779   GLDTLTVMGIAFAAFVIGALLTGALWYIYSHTGETAGRQQV--PTSPPASENSSAAHSIGSTQSTPCSSSSTA   849
                    ::    : :: :.::.:::::.:::::::::     :  :     :  :: ::  :::::::::::::.:: :
Endoglin      573   GLVLPAVLGITFGAFLIGALLTAALWYIYSHTRSPSKREPVVAVAAPASSESSSTNHSIGSTQSTPCSTSSMA   645
                                                                        #
```

Fig. 3. Sequence similarity between the human TGF-β type III receptor and human endoglin. Numbers indicate the amino acid residues. Identical amino acids are indicated by (:), and chemically similar amino acids by (.). Cysteine residues (#), an N-glycosylation site (•), and an RGD sequence (underlined) are indicated. The putative transmembrane domain of the human TGF-β type III receptor is overlined.

The overall sequence conservation of the TGF-β type III receptors is not high compared to that of their ligands. This was also the case for the TGF-β type II receptor. Lin et al. (24) reported that the sequence identity between the human and porcine type II receptor was 88 % in the N-terminal 297 amino acids. This is a clear contrast with the activin type II receptor, which has an overall 99 % amino acid sequence identity between mouse and human (25, 26).

Structural Similarity between the Human TGF-β Type III Receptor and Human Endoglin. The cDNA cloning of the rat TGF-β type III receptor revealed a high sequence similarity with human endoglin in their transmembrane and intracellular domains (10, 11). Similar with the rat type III receptor, the transmembrane and intracellular domains (amino acids 779-849) of the human type III receptor have high sequence similarity with human endoglin (12) (Fig. 3). The amino acid sequence of endoglin in this region is also highly conserved between human and porcine (our unpublished observation). Moreover, it has recently been shown that TGF-βs bind to endoglin in an isoform specific manner (27). These results suggest that the transmembrane and the intracellular domains have an important function, which may be common for both the TGF-β type III receptor and endoglin.

In the extracellular domain, regions with lower sequence similarity between the rat TGF-β type III receptor and human endoglin have been reported by other investigators (27, 28). Here we show two stretches of amino acid sequences, which are substantially similar between the human TGF-β type III receptor and human endoglin. In the proximity of the N-terminus of the molecule, a 29 amino acid sequence (amino acids 67-95) was found to have 41 % sequence identity with endoglin. In the middle part of the molecule, a stretch of 52 amino acids including three cysteine residues (amino acids 453-504) was found to be similar with endoglin; most profound (48 % identity) in the C-terminal 23 amino acids. Whether these homology regions in the extracellular domains are important for the binding of TGF-βs, is an interesting question to be elucidated.

Acknowledgments: We thank Drs. Hideo Toyoshima and Hisamaru Hirai for a human placenta cDNA library. We also thank Carl-Henrik Heldin, Peter ten Dijke and Hidetoshi Yamashita for valuable discussion, Hans Ronne for the help in computer work, and Christer Wernstedt for preparing oligonucleotides.

REFERENCES

1. Roberts, A.B. and Sporn, M.B. (1990) In *Peptide Growth Factors and Their Receptors, part I* (M.B. Sporn and A.B. Roberts, eds.) pp. 419-472, Springer-Verlag, Berlin.
2. Moses, H.L., Yang, E.Y. and Pietenpol, J.A. (1990) *Cell* **63**, 245-247.
3. Derynck, R., Jarrett, J.A., Chen, E.Y. and Goeddel, D.V. (1988) *J. Biol. Chem.* **261**, 4377-4379.
4. Massagué, J. (1992) *Cell* **69**, 1067-1070.
5. Ohta, M., Greenberger, J.S., Anklesaria, P., Bassols, A. and Massagué, J. (1987) *Nature* **329**, 539-541.
6. Segarini, P.R., Rosen, D.M. and Seyedin, S.M. (1989) *Mol. Endocrinol.* **3**, 261-272.
7. Segarini, P.R. and Seyedin, S.M. (1988) *J. Biol. Chem.* **263**, 8366-8370.
8. Cheifetz, S., Andres, J.L. and Massagué, J. (1988) *J. Biol. Chem.* **263**, 16984-16991.
9. Andres, J.L., Stanley, K., Cheifetz, S. and Massagué, J. (1989) *J. Cell Biol.* **109**, 3137-3145.
10. López-Casillas, F., Cheifetz, S., Doody, J., Andres, J.L., Lane, W.S. and Massagué, J. (1991) *Cell* **67**, 785-795.
11. Wang, X.-F., Lin, H.Y., Ng-Eaton, E., Downward, J., Lodish, H.F. and Weinberg, R.A. (1991) *Cell* **67**, 797-805.
12. Gougos, A. and Letarte, M. (1990) *J. Biol. Chem.* **265**, 8361-8364.
13. Gougos, A. and Letarte, M. (1988) *J. Immunol.* **141**, 1925-1933.
14. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry* **18**, 5294-5299.
15. Yannisch-Perron, C., Vieira, C.J. and Messing, J. (1985) *Gene* **33**, 103-119.
16. Kanzaki, T., Olofsson, A., Morén, A., Wernstedt, C., Hellman, U., Miyazono, K., Claesson-Welsh, L. and Heldin, C.-H. (1990) *Cell* **61**, 1051-1061.
17. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
18. Short, J.M., Fernandez, J.M., Sorge, J.A. and Huse, W.D. (1988) *Nucleic Acids Res.* **16**, 7583-7600.
19. Kozak, M. (1989) *J. Cell Biol.* **108**, 229-241.
20. von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243-251.
21. Saunders, S., Jalkanen, M., O'Farrell, S. and Bernfield, M. (1989) *J. Cell Biol.* **108**, 1547-1556.
22. Pandiella, A. and Massagué, J. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1726-1730.
23. Pearson, R.B. and Kemp, B.E. (1991) *Meth. Enzymol.* **200**, 62-81.
24 Lin, H.Y., Wang, X.-F., Ng-Eaton, E., Weinberg, R.A. and Lodish, H.F. (1992) *Cell* **68**, 775-785.
25. Donaldson, C.J., Mathews, L.S. and Vale, W.W. (1992) *Biochem. Biophys. Res. Commun.* **184**, 310-316.
26. Matzuk, M.M. and Bradley, A. (1992) *Biochim. Biophys. Acta* **1130**, 105-108.
27. Cheifetz, S., Bellón, T., Calés, C., Vera, S., Bernabeu, C., Massagué, J. and Letarte, M. (1992) *J. Biol. Chem.* **267**, 19027-19030.
28. Bork, P. and Sander, C. (1992) *FEBS Lett.* **300**, 237-240.