The EMBO Journal vol.6 no.7 pp.1891-1897, 1987

Structure and expression of human thrombomodulin, a thrombin receptor on endothelium acting as a cofactor for protein C activation

Koji Suzuki, Hiroshi Kusumoto, Yoshihiro Deyashiki, Junji Nishioka, Ikuro Maruyama[1], Michitaka Zushi[2], Shinji Kawahara[2], Gouichi Honda[2], Shuji Yamamoto[2] and Sadayuki Horiguchi[2]

Department of Laboratory Medicine, Mie University School of Medicine, Tsu City, Mie 514, [1]Third Department of Internal Medicine, Kagoshima University School of Medicine, Kagoshima City, Kagoshima 890, and [2]Bio-Science Laboratory, Life Science Research Laboratories, Asahi Chemical Industry, Fuji City, Shizuoka 416, Japan

Communicated by S.Magnusson

We have deduced the entire 575-amino acid sequence of the human thrombomodulin precursor from cDNA clones. The precursor starts with an 18-residue signal peptide domain, followed by the NH_2-terminal domain, a domain with six epidermal growth factor-like structures, an O-glycosylation site-rich domain, a 24-residue transmembrane domain and a cytoplasmic domain. Simian COS cells transfected with the expression vector pSV2 containing thrombomodulin cDNA synthesized immunoreactive and functionally active thrombomodulin.

Key words: anti-coagulant/endothelium/protein C activation/thrombin receptor/thrombomodulin

Introduction

Thrombomodulin is a receptor for thrombin on the vascular endothelium and plays an essential role as a protein cofactor in thrombin-catalyzed activation of protein C (Esmon *et al.*, 1982b). Activated protein C is a most potent anti-coagulant in blood coagulation by inactivating the coagulation cofactors, Factors V and VIII (Marlar *et al.*, 1982; Suzuki *et al.*, 1983). Furthermore, thrombomodulin has been found to block the procoagulant activities of thrombin, such as fibrinogen clotting and Factor V activation (Esmon *et al.*, 1982a) and platelet activation (Esmon *et al.*, 1983). Recently thrombomodulin has been found to inhibit Factor Xa activity in the prothrombinase complex (Thompson and Salem, 1986). Thus, thrombomodulin is involved in functional alteration of thrombin from a procoagulant to an anti-coagulant, and also acts as an inhibitor of intravascular coagulation. Thrombomodulin has been isolated from the rabbit (Esmon *et al.*, 1982b), bovine (Suzuki *et al.*, 1986) and human (Salem *et al.*, 1984) lungs and the human placenta (Salem *et al.*, 1984) as a single-chain glycoprotein with an apparent M_r of 78 000. Thrombomodulin exists on the luminal surface of the endothelium in arteries, veins, capillaries and lymphatics, and also in the syncytiotrophoblast of placenta (Maruyama *et al.*, 1985). Once thrombin binds to thrombomodulin on the cell surface, the resulting complex is considered to be transferred into the cell where thrombin is then degraded, and the internalized thrombomodulin recycles onto the surface of the membrane (Maruyama and Majerus, 1985). The uptake and degradation of thrombin bound to thrombomodulin is very similar to the receptor-mediated endocytosis of epidermal growth factor (EGF) (Carpenter and Cohen, 1976), low density lipoprotein (LDL) (Goldstein *et al.*, 1979) and so on. Therefore, the relationship of the unqiue functions and structure of thrombomodulin should be elucidated.

We report here the isolation and nucleotide sequences of four recombinant cDNA clones, which between them contain the entire coding sequence of human thrombomodulin. The deduced amino acid sequence of human thrombomodulin is compared with that of bovine thrombomodulin which recently has been partially determined (Jackman *et al.*, 1986). We also describe the construction of an expression plasmid that directs the synthesis in simian COS-1 cells of a glycoprotein having the cofactor activity that causes thrombin-catalyzed activation of protein C.

Results

NH_2-terminal amino acid sequence

Thrombomodulin, purified from human lung endothelial membrane preparations, had an apparent M_r of 78 000 and 105 000 before and after disulfide bond reduction, respectively, on SDS-PAGE (Figure 1a). The M_r of human thrombomodulin was decreased to M_r 68 000-70 000 by treatment with neuraminidase and then O-glycanase (Umemoto *et al.*, 1977), as judged by SDS-PAGE under non-reducing conditions. The treatment with neuraminidase and then N-glycanase (Tarentino *et al.*, 1985) also lowered the M_r of the intact protein to M_r 69 000-72 000 (data not shown). These findings suggest that both O-glycosylation sites and N-glycosylation sites are present in human thrombomodulin, and that the M_r of the carbohydrate-free protein is ~60 000.

The NH_2-terminal amino acid sequence of the human thrombomodulin is shown in Figure 1b. Nine of the 14 amino acid residues in the NH_2-terminal sequence of human thrombomodulin were identical with those of the bovine protein whose residue one is blocked (unpublished data). Four 23-mer oligonucleotides which would code for amino acid residues 1-8 of the human protein were synthesized and used as probes for screening a cDNA library to identify human thrombomodulin cDNA clones (Figure 1c).

Cloning of thrombomodulin cDNA clones

The strategy used to obtain a cDNA for the human thrombomodulin is outlined in Figure 2. Polyclonal antibody obtained from the immunized rabbits was used to screen a λgt11 human lung cDNA library, and an 1104-bp cDNA containing the sequence corresponding to the probable COOH-terminal end of the thrombomodulin protein, designated TM13, was obtained from 5×10^6 clones. By using ^{32}P-labelled TM13 as a probe, re-screening of the cDNA library was performed, and a 2408-bp cDNA, designated TM137, was obtained from 7×10^6 clones. The nucleotide sequence of TM137 revealed that it contained 226 residues upstream from the 5′ end of TM13 and 1078 residues downstream from the 3′ end of TM13, but did not contain the 5′ end of the thrombomodulin cDNA. Primer extension was then performed to obtain the cDNA clone encoding the NH_2-terminal

end of thrombomodulin. A 20-mer oligonucleotide, corresponding to the nucleotide residues 94−113 of TM13, was used to construct a primer-extended cDNA library in pBR322. As a template, human umbilical cord endothelial cell poly(A)$^+$ RNA was used, since by Northern blot analysis the poly(A)$^+$ RNA prepared from the endothelial cells was found to contain the mRNA (~3.8 kb) which hybridized with the ^{32}P-labelled TM13. The library was screened with two different ^{32}P-labelled oligonucleotides corresponding to residues 21−40 and 79−93 of TM13. Of ~70 000 recombinants screened, six plasmids contained the identical length 921-bp cDNA insert, designated TMP5, which overlapped the TM13 by 113 nucleotides. As TMP5 did not contain the 5′ end of the thrombomodulin cDNA a second primer extension was performed, using a 20-mer oligonucleotide, corresponding to residues 144−163 of TMP5. Two different 20-mer oligonucleotides, corresponding to residues 22−41 and 94−113 of TMP5, were used for screening the plasmid library. Of ~50 000 recombinants screened, one plasmid was found to contain a 635-bp cDNA insert, designated TMP26. This cDNA extended 153 nucleotides upstream of the ATG codon of the thrombomodulin cDNA, as analyzed by Southern blotting using the ^{32}P-labelled oligonucleotide probes for NH_2-terminal sequence.

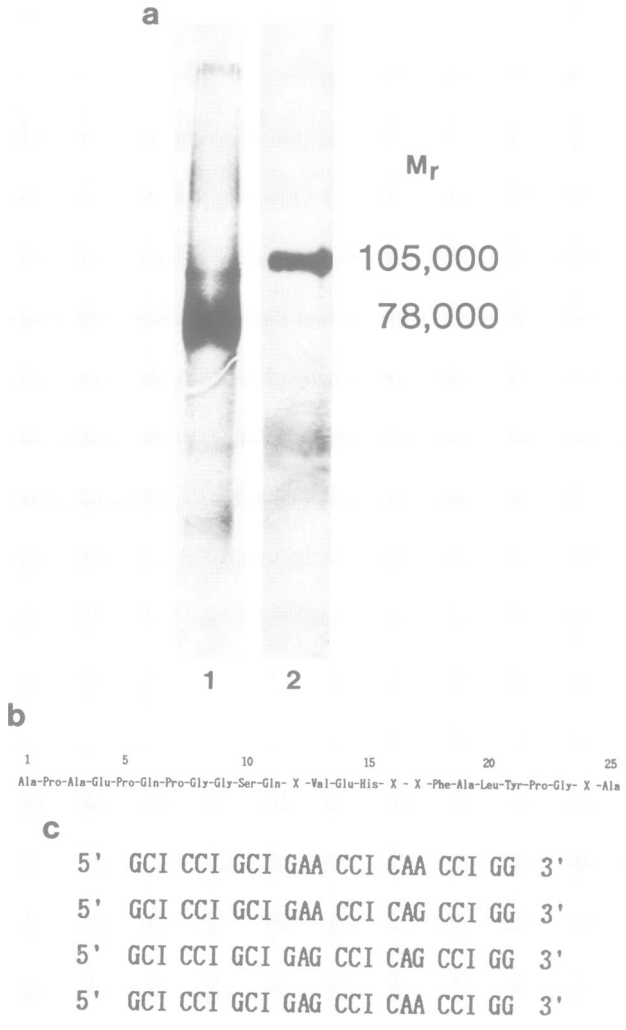

Fig. 1. Human thrombomodulin purification and NH_2-terminal sequence analysis. (**a**) SDS−PAGE of human lung thrombomodulin. **Lane 1**: non-reducing conditions; **lane 2**: reducing conditions. The M_r of thrombomodulin was determined by comparison with the mobility of commecial marker proteins. (**b**) NH_2-terminal sequence of human thrombomodulin. (**c**) Four 23-mer oligonucleotide probes corresponding to amino acid residues 1−8.

Thrombomodulin cDNA and protein sequence

The inserts of TM13, TM137, TMP5 and TMP26 were subcloned into bacteriophage vectors and employed for sequencing. Figure 3a shows the restriction endonuclease map and the entire overlapping cDNA of thrombomodulin. Figure 3b shows the nucleotide sequence of the cDNA of the precursor of human thrombomodulin and the deduced amino acid sequence. This amino acid sequence contained a stretch matching the NH_2-terminal sequence of 25 amino acid residues of human thrombomodulin. An ATG codon was found 18 amino acids upstream from the NH_2-terminal alanine of the mature thrombomodulin. According to the rule proposed by Kozak (1981) for initiation of translation in eukaryotes, this ATG codon (nucleotide positions 1−3) would be used to initiate an open reading frame of 1725 bp, coding for 575 amino acids consisting of a putative 18-residue signal peptide and the 557-residue mature thrombomodulin with a protein M_r of 58 662. The amino acid residues of this signal peptide are to a large extent hydrophobic. The present cDNA sequence possessed two characteristic polyadenylation sequences, AATAAA, at residues 2417 and 2837, but a poly(A) tail sequence was not found nearby, indicating that we may not have the complete cDNA sequence. As the length of mRNA was ~3.8 kb, the region containing the polyadenylation sequence and the poly(A) tail had presumably been removed during the preparation of the λgt11 cDNA library.

The deduced amino acid sequence revealed several structural features characteristic of the mature protein. Five potential N-glycosylation sequences (Asn-X-Ser/Thr) and seven potential O-glycosylation sequences (Ser/Thr-X-X-Pro) are distributed in the total sequence. The human thrombomodulin can be tentatively divided into five domains which are shown schematically in Figure 4.

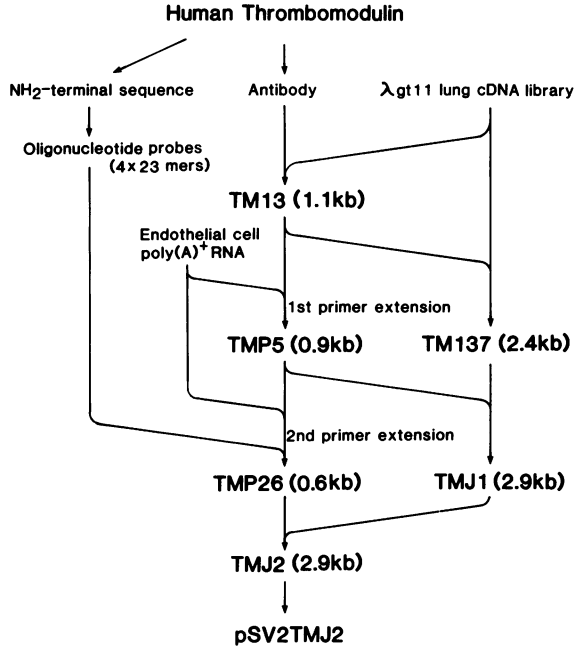

Fig. 2. cDNA cloning strategy for human thrombomodulin. For construction of the cDNA coding for the entire thrombomodulin precursor, TMP5 and TMP137 were ligated to yield a TMJ1, and this was further ligated with TMP26 to yield a TMJ2. The TMJ2, a cDNA encoding the entire thrombomodulin precursor, was cloned into an expression vector pSV2 to yield a pSV2TMJ2. Details are described in Materials and methods.

Amino acid residues 1–226 are tentatively assigned as constituting the first domain. It is not known if the 10 cysteine residues in this domain are disulfide bonded. A search of the Protein Identification Resource (National Biomedical Research Foundation, 1986, Protein Sequence Database of the Protein Identification Resource, Release No. 11.0) failed to identify any other protein significantly homologous with this domain.

The second domain consists of residues 227–462. This domain contains 36 cysteine residues (15% of the 236 residues). Dot matrix homology analysis (Dayhoff *et al.*, 1983) showed that this cysteine-rich domain is composed of multiple repeat sequences. With the introduction of a few short gaps, six repeat

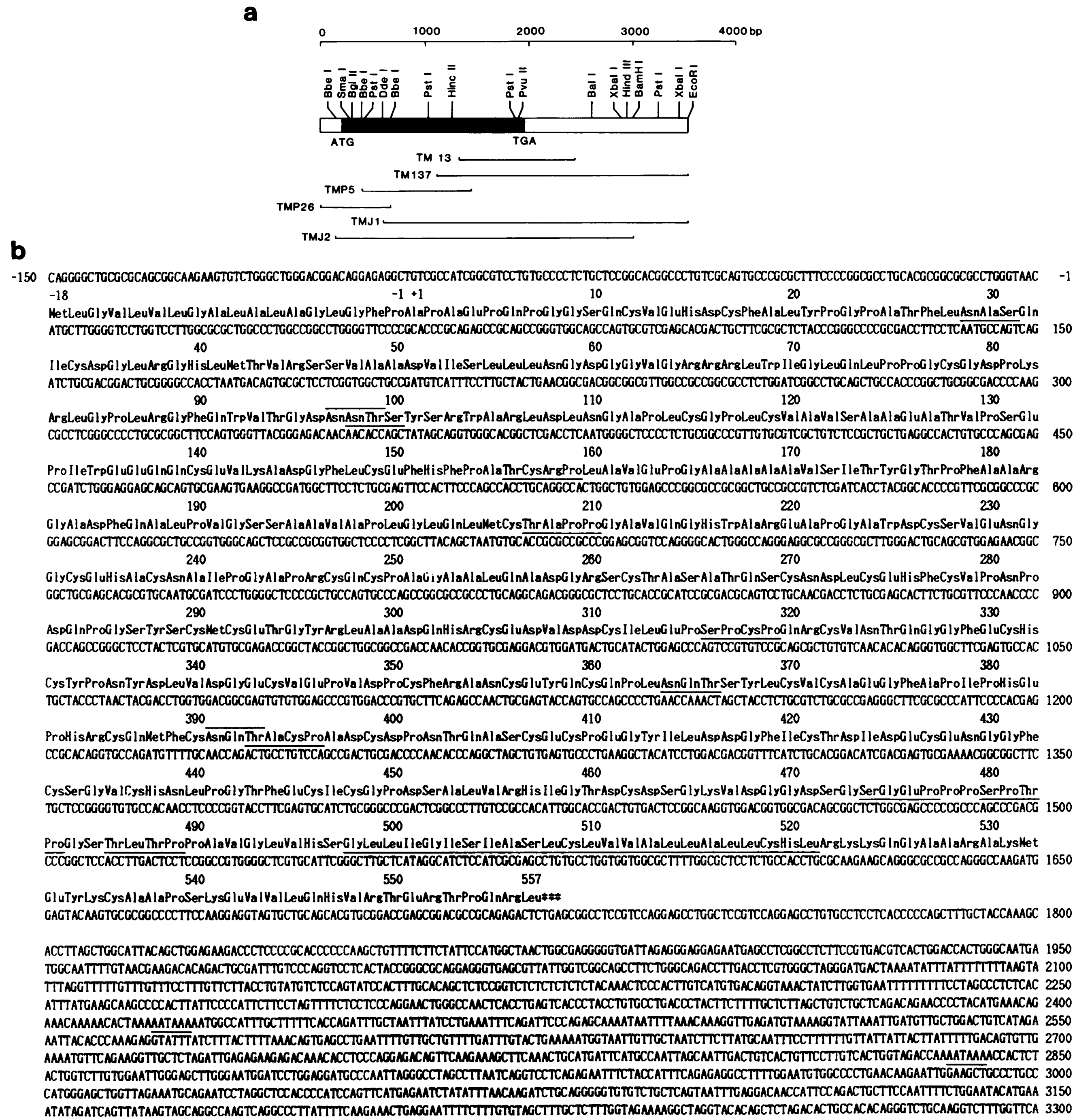

Fig. 3. Human thrombomodulin cDNA clones, nucleotide sequence and predicted amino acid sequence. (**a**) Restriction endonuclease map of the cDNA insert coding for the thrombomodulin precursor. The thick black region indicates the sequence coding for the precursor of thrombomodulin (575 amino acids) and the open regions show the 5'- and 3'- untranslated sequences. The lines below the restriction map represent the length and the position of the five partial clones, TM13, TM137, TMP5, TMP26, TMJ1 and TMJ2. (**b**) Nucleotide sequence of the human thrombomodulin precursor cDNA insert and the deduced amino acid sequence. The nucleotide numbers are shown on the right side. Number 1 is the A of the ATG codon of the initiator methionine; negative numbers refer to the 5'- untranslated region. Amino acid numbers are shown over the sequence; residue 1 is the NH_2-terminal Ala of the mature protein (557 residues); negative numbers refer to the putative signal peptide (18 residues). The transmembrane sequence (residues 498–521) is underlined. The potential N-glycosylation sites (Asn-X-Ser/Thr) and O-glycosylation sites (Ser/Thr-X-X-Pro) are under- or overlined. Two potential polyadenylation signals in the 3'-untranslated region are also underlined.

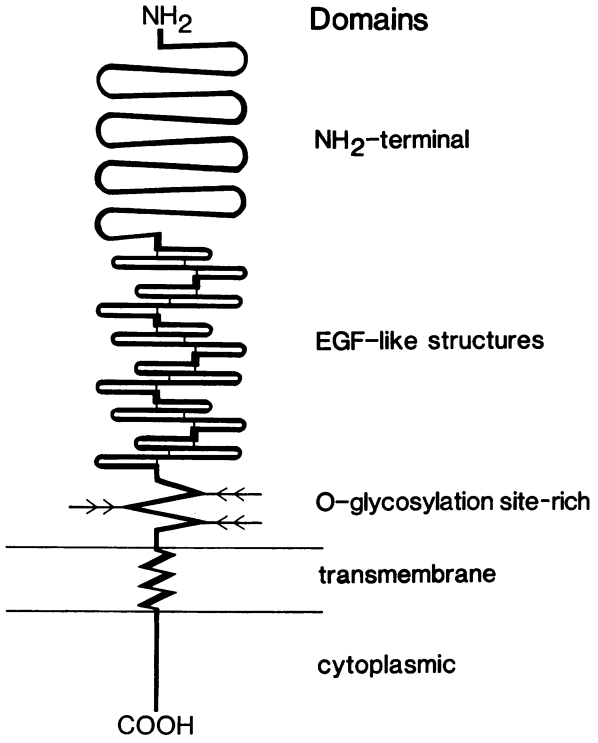

Fig. 4 A schematic model showing the five domains in the structure of the human thrombomodulin. The characteristic features of each domain are described in the text.

sequences with minor variations can be aligned so that the cysteine residues are in register (Figure 5). These sequences also had strikingly high homology to the sequence of the polyprotein precursor of mouse EGF (Gray *et al.*, 1983). Maximal homology (45%) to the precursor of mouse EGF (residues 745−786) was found particularly in the first EGF-like sequence in this domain (data not shown). These EGF-like structures also have high homology to the EGF-like structures in the human (Russell *et al.*, 1984) and bovine (Yamamoto *et al.*, 1984) LDL receptor, and those in human and bovine proteins involved in blood coagulation, fibrinolysis and complement systems, e.g. Factor IX, Factor X, protein C, tissue plasminogen activator, urokinase (Patthy, 1985), Factor VII (Hagen *et al.*, 1986), protein S (Lundwall *et al.*, 1986), protein Z (Højrup *et al.*, 1985), Factor XII (Cool *et al.*, 1985) and C1r (Leytus *et al.*, 1986).

The third domain consists of residues 463−497. Serine, threonine or proline constituted 16 of 35 amino acid residues of this domain. Three of the seven potential O-glycosylation sequences are concentrated here. As described before, the apparent M_r of the intact thrombomodulin was decreased by ~10 000 by treatment with O-glycanase. Therefore, the carbohydrate chains would be attached to some or all of these three O-glycosylation sites.

The fourth domain is a hydrophobic putative transmembrane domain of 24 amino acids (residues 498−521) containing two cysteine residues. No homology was seen between this sequence and the transmembrane sequences of the other known receptor proteins.

```
1st domain   Human (207-226)  TAPPGAVQGHVAREAPGAVD
                                  *     * *******
             Bovine(  1- 20)  RGARGETEQRVSREAPGAVA

2nd domain   Human (227-262)        CSVENGG-CEHACNAIPGAPR---CQCP-AGAALQAD-GRSC
                                    * ** ** * * *    **      * ** * ****** ****
             Bovine( 21- 56)        CGVERGG-CQHECKGSAGASN---CLCP-ADAALQAD-GRSC

             Human (263-304)  TASATQSC---NDL-CEHFCVPNPDQPGSYSCMC-ETGYRLAADQ-HRC
                                 *   *     * *****       * *  *  * * ** ***** ***
             Bovine( 57- 97)  GLPAEHPC---HQL-CEHFC--HLHGLGNYTCIC-EAGYQLAADQ-HRC

             Human (305-344)  -EDVDD-CILEPSP-CPQRCVNTQGGFE---CHCYPN-YDLV-D-GE-C
                               ***** *   *** ********* ***    ***    * ** * ** *
             Bovine( 98-136)  -EDVDD-CAQLPSP-CPQRCVNTEGGFQ---CHCDTG-YELV-D-GE-C

             Human (345-386)  VEPVDP-CFRAN---CEYQCQ--PLNQTSYLCVCAE-GFAPIPHEPHRC
                              * **** **  *   ******  *       *  *** **** *  ** *
             Bovine(137-178)  VDPVDP-CFDNN---CEYQCQ--PVGRSEHKCICAE-GFAPVPGAPHKC

             Human (387-421)  QWF----C---NQTACPADCD--PNTQAS--CECPE-GYILD-D-GFIC
                              ***     *   *** ******  *       *  *** *** *   *  *
             Bovine(179-213)  QWF----C---NQTSCPADCD--PHYPTI--CRCPE-GYIID-E-GSTC

             Human (422-462)  T-DIDE-C--ENGGFCSGVCHNLPGTFE---CICGPD-SALVRHIGTDC
                              * ** * *        * * *******  *   *******  ***   ** **
             Bovine(214-253)  T-DINE-C---DTNICPGQCHNLPGTYE---CICGPD-SALSGQIGIDC

3rd domain   Human (463-497)  DSGKV------DGGDSGSGEPPPSPTPGSTLTP-PA-VGLVHS
                              *   *           ****** *****  *  * **  *  **
             Bovine(254-296)  DPTQVNEERGTPEDYGGSGEPPVSPTPGATARPSPAPAGPLHS

4th domain   Human (498-521)  GLLIGISIASLCLVVALLALLCHL
                               * * ******* ************
             Bovine(297-320)  GVLVGISIASLSLVVALLALLCHL

5th domain   Human (522-557)  RKKQGAARAKMEYKCAAPSKEVVLQHVRTERTPQRL
                              ****** *   ****  * **  ** * ****** *
             Bovine(321-356)  RKKQGASRGELEYKCGVPAKELMLQQVKTERTPQKL
```

Fig. 5. Alignment of the sequence from the COOH-terminal end of the first domain to the fifth domain of human thrombomodulin with that of the bovine protein. Optimal alignment was made by the computer program ALIGN. Identical amino acids between two species are marked with stars. Cysteine residues of the six repeat sequences aligned in the second domain between two species are boxed. The single-letter code for amino acid residues is that recommended by the IUPAC-IUB Commission on Biochemical Nomenclature.

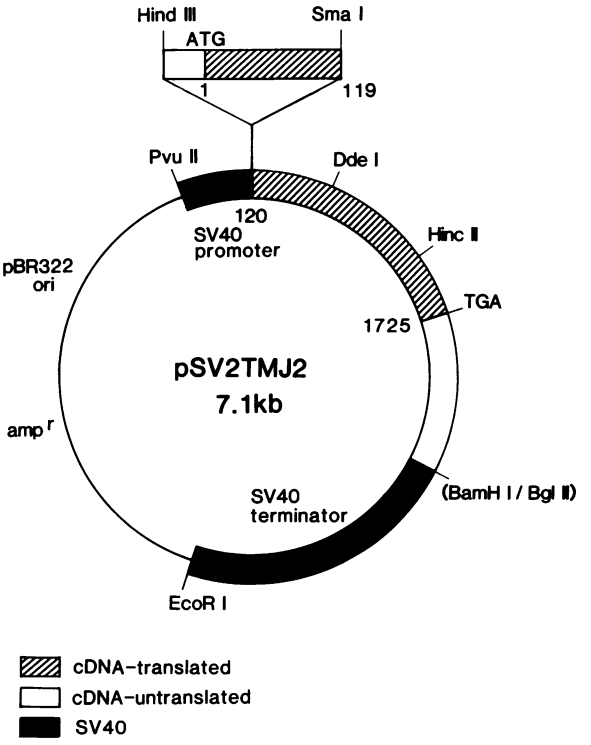

Fig. 6. Structure of pSV2TMJ2 containing a cDNA for the entire thrombomodulin precursor. The coding region (hatched area) encompasses nucleotides 1–1725. The solid areas in the cloning vector denote regions containing the SV40 promoter and terminator.

Table I. Expression of thrombomodulin cDNA in COS-1 cells

COS-1 cells transfected with	APC formed (pmol/ml/min)
pSV2TMJ2 (20 μg)	16.7
pSV2TMJ2 (10 μg)	7.4
pSV2DHFR (20 μg)	0
Thrombomodulin (2 nM)	4.1

The cells transfected with pSV2TMJ2 or pSV2DHFR were cultured in Dulbecco's MEM containing 10% fetal calf serum for 48 h, and the collected cells were assayed for thrombomodulin activity as described in Materials and methods. The activity of thrombomodulin purified from human lungs was assayed as a control. APC, activated protein C.

The fifth domain starts with the basic amino acids, Arg-Lys-Lys (Figure 3b), and consists of 36 amino acids (residues 522–557) at the COOH-terminal end of the protein. It would be located on the cytoplasmic side of the plasma membrane. Like the LDL receptor (Russell *et al.*, 1984; Yamamoto *et al.*, 1984) and unlike the receptors for EGF (Ullrich *et al.*, 1984) and insulin (Ullrich *et al.*, 1985), it has no tyrosine kinase domain sequence.

Expression of thrombomodulin cDNA in COS cells

pSV2TMJ2, which contains a cDNA for the entire precursor of thrombomodulin and the SV40 translation promoter and terminator (Figure 6), was transfected in SV40-transformed simian kidney-derived COS-1 cells by an electroporation technique. In controls, COS cells transfected with a cDNA of mouse dehydrofolate reductase (DHFR) were used. Table I shows the expression of the cofactor activity in the COS cells transfected with pSV2TMJ2, whereas no activity was detected in the cells transfected with pSV2DHFR. The cofactor activity was detected on the cell, but not in the culture medium. The COS cells transfected with pSV2TMJ2 were estimated to express an average of 2 400 000 molecules of thrombomodulin per cell, based on the cofactor activity of the cells in comparison with that of the purified thrombomodulin. The thrombomodulin expressed could also be visualized with the monoclonal antibody for thrombomodulin (Figure 7a and b). About 25% of the transfected COS cells adsorbed large amounts of the monoclonal antibody on the surface of the cells. No fluorescence was seen in the COS cells transfected with pSV2DHFR (Figure 7c). These results indicate that pSV2TMJ2 directs the synthesis of the immunoreactive and functionally active human thrombomodulin.

The M_r of the products expressed in the COS cells, which were solubilized with 0.5% Triton X-100, was estimated using SDS–PAGE followed by immunoblotting using the monoclonal antibody for thrombomodulin. The thrombomodulin expressed in the COS cells appeared to have the same M_r as the protein that was purified from the human lung and placenta (data not shown). Therefore, the size of carbohydrate chains of the recombinant thrombomodulin is presumably very similar to that of the purified protein.

Discussion

We have determined the nucleotide sequence of the cloned cDNA of human thrombomodulin, and its amino acid sequence was deduced. To confirm that the cDNA codes for thrombomodulin,

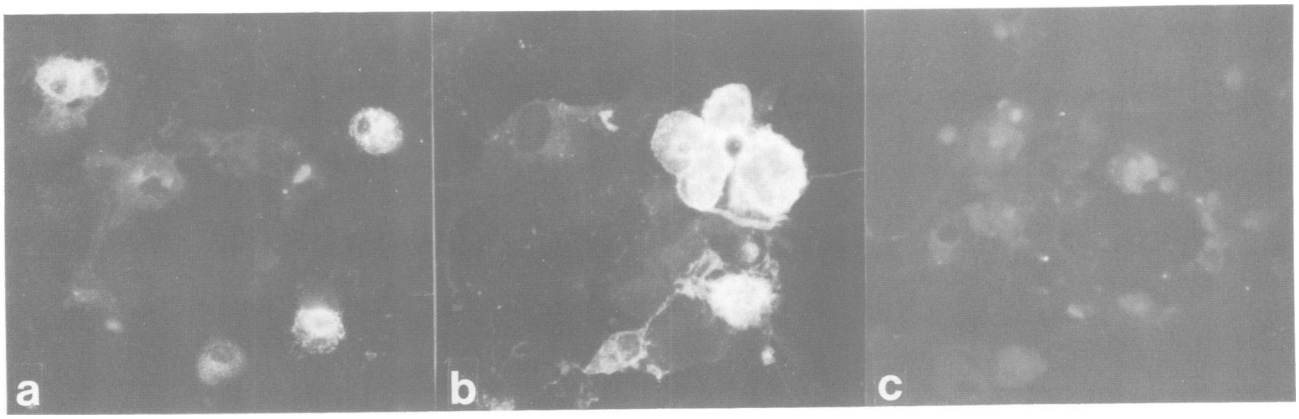

Fig. 7. Visualization of thrombomodulin present on the pSV2TMJ2-transfected COS-1 cells with monoclonal anti-thrombomodulin IgG. (**a**) and (**b**) transfected with pSV2TMJ2; (**c**) transfected with pSV2DHFR.

transfection of simian COS-1 cells with the plasmid containing the cDNA was performed and expression of functionally active thrombomodulin on the surface of the cells was obtained. To prepare a cDNA coding for the entire thrombomodulin protein, we performed primer extension twice. As described before, the extension was extremely inefficient. From the nucleotide sequence of the cDNA, the coding region of mRNA was estimated to contain G + C-rich regions, particularly over 70% of residues −50 to +200 and +450 to +850. At least 10 stem (consisting of >10 bp) and loop structures (Yanofsky, 1981) could be formed in this region. The inefficiency of the extension might have been due to this characteristic feature of the mRNA.

Recently, after our sequencing had been accomplished, the amino acid sequence of the COOH-terminal half of bovine thrombomodulin was revealed (Jackman *et al.*, 1986). The amino acid sequence of human thrombomodulin and the partial sequence of bovine are compared in Figure 5. As for the first domain, although only the COOH-terminal 20 amino acid residues were revealed in the bovine protein, eight of the 20 residues are identical in the two species. Jackman *et al.* (1986) stated that the NH_2-terminal half of the bovine thrombomodulin is rich in lysine, methionine and phenylalanine and low in cysteine content, based on their amino acid analysis. However, our present results show that the first domain contains two lysines, two methionines, seven phenylalanines and 10 cysteines, and also that the first 14 amino acid residues of the mature human protein have a high degree of homology (64%) with those of our bovine protein (noted before). These results suggest that the NH_2-terminal and the COOH-terminal ends of the first domain are relativley conserved between both proteins, but most of the middle is not. In the second domain, both proteins contain 36 cysteine residues and six EGF-like structures and the six repeated sequences show a high degree of homology (67−77% of the amino acid sequences and 74−80% of the nucleotide sequences). The markedly retarded mobility of the reduced human (Figure 1a) and bovine thrombomodulin (Suzuki *et al.*, 1986) on SDS−PAGE is presumably due to the unfolding of the extensively cross-linked structures in this domain as seen in the LDL receptor (Russell *et al.*, 1984). This degree of structural conservation between the two species suggests that this domain has some important function such as binding of thrombin. In the third domain consisting of 35 amino acid residues in the human protein and 43 in the bovine, a 45% homology was found between the two. Although the human protein had three potential O-glycosylation sites, the bovine had four. In the fourth domain, a putative transmembrane domain consisting of 24 amino acid residues in both species is remarkably homologous (87%). The difference between the two species is that the human protein has two cysteine residues, the bovine protein only one. In the fifth domain containing 36 amino acid residues starting with the basic residues, Arg-Lys-Lys, in both proteins, they are also highly homologous (67%). This comparison shows that the sequences from the second to the fifth domain of human and bovine thrombomodulins are highly homologous and an overall feature of these four COOH-terminal domains is quite similar to that of the LDL receptor.

Thrombomodulin has been demonstrated to have the following functional activities: (i) cofactor activity for thrombin-catalyzed activation of protein C (Esmon *et al.*, 1982b), (ii) direct anti-coagulant activity for thrombin (Esmon *et al.*, 1982a), (iii) inhibitory activity for prothrombinase complex (Factor Xa + Factor Va + phospholipid + calcium) (Thompson and Salem, 1986), and (iv) anti-thrombin III-dependent heparin-like activity (Bourin *et al.*, 1986). Thrombomodulin is unusually stable against detergents, but its cofactor activity is rapidly decreased by reducing agents such as 2-mercaptoethanol (Esmon *et al.*, 1982b). This suggests that the disulfide-bonded conformation, e.g. in the domain containing EGF-like structures, is indispensable in expressing its cofactor activity. Bovine thrombomodulin is rather sensitive to proteolysis during purification (Suzuki *et al.*, 1986). An intact form with M_r 78 000 is often degraded into a form with M_r 65 000−70 000. These products, however, retain their cofactor activity, and their NH_2-terminal sequence is the same as that of the intact form (unpublished data), suggesting that the COOH-terminal region (at least M_r 10 000) possibly involving the third to fifth domains of thrombomodulin is not required for the expression of cofactor activity in protein C activation. Recently, in human plasma and urine a soluble and active thrombomodulin was found (Ishii and Majrus, 1985). It is presumably derived from the membrane form by proteolysis outside the transmembrane domain, as in the case of our bovine thrombomodulin.

Materials and methods

Purification and amino acid sequence determination of human thrombomodulin

Human thrombomodulin was purified from the lung membrane preparations by chromatography on di-isopropyl phosphate (DIP)−thrombin−agarose and gel filtration on an Ultrogel AcA-44 (LKB) column by the method used for bovine thrombomodulin (Suzuki *et al.*, 1986), except that three cycles of chromatography on DIP−thrombin−agarose were performed. From 1 kg of starting material, 0.4 mg purified protein was finally obtained with a specific activity >8000-fold that of the solubilized membrane extract. Samples were electrophoresed on gradient polyacrylamide slab gels (5−15%) in 0.1% SDS according to Blobel and Dobberstein (1975), and protein visualized by silver staining (Oakley *et al.*, 1980).

To determine the NH_2-terminal sequence purified thrombomodulin was applied to the vapor phase protein sequencer (Model 470A, Applied Biosystems), (Hewick *et al.*, 1981) and the resulting amino acid derivatives determined by reversed-phase h.p.l.c. on a Microsorb C8 column (4.6 × 250 mm, Rainin) using an aqueous phosphate buffer with acetonitrile as the organic solvent.

Oligonucleotide synthesis

Oligonucleotide probes were synthesized using an automatic DNA synthesizer (Model 380A, Applied Biosystems). Probes corresponding to the NH_2-terminal amino acid sequence of thrombomodulin were designed using deoxyinosine according to Ohtsuka *et al.* (1985). The probes were phosphorylated using a 3-fold excess of [^{32}P]ATP (Amersham) and T_4 DNA kinase (Takara) (Maniatis *et al.*, 1982).

Nucleotide sequence determination

The dideoxy termination method (Sanger *et al.*, 1977) was used with M13mp18 and M13mp19 vectors (Amersham and Takara) (Mizusawa *et al.*, 1986).

Screening and construction of cDNA libraries

A λgt11 human lung cDNA library (8.2 × 10^5 independent clones) was purchased from Clontech Laboratories (Palo Alto, CA). It was screened first with polyclonal anti-human thrombomodulin IgG according to Young and Davis (1983), then with ^{32}P-labelled TM13 cDNA. A primer-extended cDNA library in plasmid pBR322 (Sutcliffe, 1979) was constructed by a modification of the method of Ragg (1986). Poly(A)$^+$ RNA, as a template, was prepared by the method of Chirgwin *et al.* (1979), from cultured venous endothelial cells of human umbilical cord (Mano *et al.*, 1983). The plasmid cDNA library was screened essentially as in Maniatis *et al.* (1982), using ^{32}P-labelled oligonucleotide probes. Northern and Southern blot analyses were performed according to Lehrach *et al.* (1977) and Silhavy *et al.* (1984), respectively.

Construction of an expression plasmid

To construct the cDNA coding for the entire thrombomodulin precursor, the cDNA inserts of TM137, TMP5 and TMP26 were joined as in Maniatis *et al.* (1982). In brief, TM137 cloned into the *Eco*RI site of plasmid pUC18 (pUCTM137) was digested with *Hin*cII and *Eco*RI, and a 2289-bp *Hin*cII−*Eco*RI fragment isolated by polyacrylamide gel (4%) electrophoresis. TMP5 in plasmid pBR322 was digested with *Dde*I and then repaired wth *Escherichia coli* DNA polymerase (Klenow fragment), and a 1069-bp cDNA fragment was cloned into the *Sma*I site of plasmid pUC18 (pUCTMP5). This clone was digested completely with *Bam*HI and *Hin*cII to yield a 599-bp *Bam*HI−*Hin*cII fragment. This and the *Hin*cII−*Eco*RI fragment were then ligated with T_4 DNA ligase and cloned into the *Bam*HI−*Eco*RI-digested pUC18, designated pUCTMJ1, which contained a

2882-bp cDNA insert of TMJ1. The pUCTMJ1 was digested completely with *Dde*I, *Kpn*I and *Bam*HI, giving a 950-bp *Dde*I−*Kpn*I fragment and a 1504-bp *Kpn*I−*Bam*HI fragment. TMP26 cloned into the *Pst*I site of plasmid pUC13 (pUCTMP26) was digested completely with *Bbe*I and then repaired with T_4 DNA polymerase. It was further digested completely with *Bgl*II to yield a 173-bp fragment. The pUCTMP26 was further digested completely with *Bgl*II and *Dde*I to yield a 279-bp fragment. The 177-bp *Bbe*I−*Bgl*II fragment, the 279-bp *Bgl*II−*Dde*I fragment and the 950-bp *Dde*I−*Kpn*I fragment were ligated followed by digestion with *Kpn*I to yield a 1406-bp *Bbe*I−*Kpn*I fragment. This 1406-bp fragment and the 1504-bp *Kpn*I−*Bam*HI fragment were ligated and cloned into the pUC18, which had been digested with *Sph*I, repaired with T_4 DNA polymerase and then digested with *Bam*HI. This clone was designated pUCTMJ2 and contained a 2906-bp TMJ2 cDNA insert that codes for the complete nucleotide sequence of the precursor of thrombomodulin. Cloning of the TMJ2 cDNA into an expression vector pSV2 (Mulligan and Berg, 1981) was performed as follows. The pUCTMJ2 was partially digested with *Hin*dIII and then digested completely with *Bam*HI to yield a 2912-bp fragment. A pSV2DHFR (ATCC 37146), an expression vector containing a DHFR cDNA insert, was digested completely with *Hin*dIII and *Bgl*II to remove the DHFR cDNA. The 2912-bp *Hin*dIII−*Bam*HI fragment was then ligated and cloned into the *Hin*dIII−*Bgl*II site of pSV2. This clone was designated pSV2TMJ2 (Figure 6).

DNA transfection and fluorescence analysis

Plasmid DNA was transfected into COS-1 cells by an electroporation technique as follows: 5 μl plasmid DNA, pSV2TMJ2 (10 and 20 μg) or pSV2DHFR (20 μg) in 1 mM Tris−HCl, 0.1 M NaCl pH 8.0 and 200 μl of the suspension of COS cells (ATCC CRL 1650) in Hank's balanced salt solution were mixed in a 1.5-ml plastic tube, and the cells were transfected with the plasmid by the electroporation technique (1.2 kV, 40 μs, twice) (Potter *et al.*, 1984). The cells were cultured in Dulbecco's MEM containing 10% fetal calf serum for 48 h in a 10-cm dish and rinsed with 5 mM Tris−HCl, 0.1 M NaCl, pH 7.5.

Expressed thrombomodulin on the cell surface was visualized by indirect immunofluorescence (Basu *et al.*, 1981) using a monoclonal antibody for thrombomodulin (Maruyama *et al.*, 1985). A coverslip of transfected cells was incubated with 50 μg antibody/ml phosphate buffer pH 7.5 at 37°C for 1 h. It was washed with the same buffer, fixed with 3% paraformaldehdye for 15 min, washed, permeabilized with 0.1% Triton X-100, washed, stained with 40 μg/ml of affinity-purified goat anti-mouse IgG coupled to FITC (Cappel), washed, and then viewed under a fluorescence microscope.

Assay of thrombomodulin activity

Before the assay of thrombomodulin activity of the transfected cells, the rinsed cells were collected using a spatula and concentrated by centrifugation at 800 r.p.m. for 10 min. The activity of thrombomodulin was assayed by measuring the increase in protein C activation (Suzuki *et al.*, 1986). To 180 μl of cell suspension in 50 mM Tris−HCl, 2 mM $CaCl_2$, 0.1 M NaCl, pH 8.0 containing 0.1% bovine serum albumin, 18 μl human protein C (10 μg) and 2 μl human thrombin (0.1 μg) were added and the mixture incubated at 37°C for 30 min. The protein C activation was terminated by adding 15 μl of a mixture of anti-thrombin III (5 μg) and heparin (5 U), and the resulting mixture was centrifuged at 10 000 r.p.m. for 5 min. The amount of activated protein C (APC) in the supernatant was then measured using a chromogenic substrate (200 μM S-2366, Kabi) in 50 mM Tris−HCl, pH 8.0 containing 50 mM CsCl. In control, the activity of the purified human lung thrombomodulin was measured.

Acknowledgements

We thank Dr Senichiro Hashimoto, Mr Masaaki Kato and Mr Yasushi Ichikawa for their general support and encouragement on this study, Dr Sadao Horiguchi for technical advice in endothelial cell culture, Mr Kazunori Toma for computer analysis, Mr Tetsuo Hitouji for NH_2-terminal sequence analysis and Mr Shuichi Kuroda for technical advice in gene expression. We are grateful to Mr Yasuyuki Takahashi, Mr Michiaki Akira, Mr Kohei Ogawa and Mr Yoichi Kato for their technical assistance. This study was supported in part by grants from the Research Foundation for Cancer and Cardiovascular Disease, The Mochida Memorial Foundation on Medical and Pharmaceutical Research and The Naito Foundation.

References

Basu,S.K., Goldstein,J.L., Anderson,R.G.W. and Brown,M.S. (1981) *Cell*, **24**, 493−502.

Blobel,G.B. and Dobberstein,B. (1975) *J. Cell Biol.*, **17**, 835−851.

Bourin,M.-C., Boffa,M.-C., Bjork,I. and Lindahl,U. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5924−5928.

Carpenter,G. and Cohen,S.J. (1976) *J. Cell Biol.*, **71**, 154−171.

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

Cool,D.E., Edgell,C.-J.S., Louie,G.V., Zoller,M.J., Brayer,G.D. and MacGillivray,R.T.A. (1985) *J. Biol. Chem.*, **260**, 13666−13676.

Dayhoff,M.O., Barker,W.C. and Hunt,L.T. (1983) *Methods Enzymol.*, **91**, 524−545.

Esmon,C.T., Esmon,N.L. and Harris,K.W. (1982a) *J. Biol. Chem.*, **257**, 7944−7947.

Esmon,N.L., Owen,W.G. and Esmon,C.T. (1982b) *J. Biol. Chem.*, **257**, 859−864.

Esmon,N.L., Carroll,R.C. and Esmon,C.T. (1983) *J. Biol. Chem.*, **258**, 12238−12242.

Goldstein,J.L., Anderson,R.G.W. and Brown,M.S. (1979) *Nature*, **279**, 679−685.

Gray,A., Dull,T.J. and Ullrich,A. (1983) *Nature*, **303**, 722−725.

Hagen,F.S., Gray,C.L., O'Hara,P., Grant,F.J., Saari,G.C., Woodbury,R.G., Hart,C.E., Insley,M., Kisiel,W., Kurachi,K. and Davie,E.W. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2412−2416.

Hewick,R.M., Hunkapiller,M.W., Hood,I.E. and Dreyer,W.J. (1981) *J. Biol. Chem.*, **256**, 7990−7997.

Højrup,P., Jensen,M.S. and Petersen,T.E. (1985) *FEBS Lett.*, **184**, 333−338.

Ishii,H. and Majerus,P.W. (1985) *J. Clin. Invest.*, **76**, 2178−2181.

Jackman,R.W, Beeler,D.L., VanderWater,L. and Rosenberg,R.D. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8834−8838.

Kozak,M. (1981) *Nucleic Acids Res.*, **9**, 5233−5252.

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

Leytus,S.P., Kurachi,K., Sakariassen,K.S. and Davie,E.W. (1986) *Biochemistry*, **25**, 4855−4863.

Lundwall,A., Dackowski,W., Cohen,E., Shaffer,M., Mahr,A., Dahlback,B., Stenflo,J. and Wydro,R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6716−6720.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) In *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, New York Press.

Mano,Y., Sawasaki,Y., Takahashi,K. and Goto,T. (1983) *Experientia*, **39**, 1144−1146.

Marlar,R.A., Kleiss,A.J. and Griffin,J.H. (1982) *Blood*, **59**, 1067−1072.

Maruyama,I. and Majerus,P.W. (1985) *J. Biol. Chem.*, **260**, 15432−15438.

Maruyama,I., Bell,C.E. and Majerus,P.W. (1985) *J. Cell Biol.*, **101**, 363−371.

Mizusawa,S., Nishimura,S. and Seela,F. (1986) *Nucleic Acids Res.*, **14**, 1319−1324.

Muligan,R.C. and Berg,P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 2072−2076.

Oakley,B.R., Kirsch,D.D. and Morris,N.R. (1980) *Anal. Biochem.*, **105**, 361−363.

Ohtsuka,E., Matsuki,S., Ikehara,M., Takahashi,Y. and Matsubara,K. (1985) *J. Biol. Chem.*, **260**, 2605−2608.

Patthy,L. (1986) *Cell*, **41**, 675−683.

Potter,H., Weir,L. and Leder,P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7161−7165.

Ragg,H. (1986) *Nucleic Acids Res.*, **14**, 1073−1088.

Russell,D.W., Schneider,W.J., Yamamoto,T., Luskey,K.L., Brown,M.S. and Goldstein,J.L. (1984) *Cell*, **37**, 577−585.

Salem,H.H., Maruyama,I., Ishii,H. and Majerus,P.W. (1984) *J. Biol. Chem.*, **259**, 12246−12251.

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

Silhavy,T.J., Berman,M.L. and Enquist,L.W. (1984) In *Experiments with Gene Fusions*. Cold Spring Harbor Laboratory, New York, pp. 191−195.

Sutcliffe,J.G. (1979) *Cold Spring Harbor Symposium*, **43**, pp. 77−90.

Suzuki,K., Stenflo,J., Dahlback,B. and Teodorsson,B. (1983) *J. Biol. Chem.*, **258**, 1914−1920.

Suzuki,K., Kusumoto,H. and Hashimoto,S. (1986) *Biochim. Biophys. Acta*, **882**, 343−352.

Tarentino,A.L., Gomez,C.M. and Plummer,T.H.,Jr (1985) *Biochemistry*, **24**, 4665−4671.

Thompson,E.A. and Salem,H.H. (1986) *J. Clin. Invest.*, **78**, 13−17.

Ullrich,A., Coussens,L., Hayflick,J.S., Dull,T.J., Gray,A., Tam,A.W., Lee,J., Yarden,Y., Libermann,T.A., Schlessinger,J., Downward,J., Mayes,E.L.V., Whittle,N., Waterfield,M.D. and Seeburg,P.H. (1984) *Nature*, **309**, 418−425.

Ullrich,A., Bell,J.R., Chen,E.Y., Herrera,R., Petruzzeli,L.M., Dull,T.J., Gray,A., Coussens,L., Liao,Y.C., Tsubokawa,M., Mason,A., Seeburg,P.H., Grunfeld,C., Rosen,O.M. and Ramachandran,J. (1985) *Nature*, **313**, 756−761.

Umemoto,H., Bhavanandan,V.P. and Davidson,E.A. (1977) *J. Biol. Chem.*, **252**, 8609−8614.

Yamamoto,T., Davis,G.C., Brown,M.S., Schneider,W.J., Casey,M.L., Goldstein,J.L. and Russell,D.W. (1984) *Cell*, **39**, 27−38.

Yanofsky,C. (1981) *Nature*, **289**, 751−758.

Young,R.A. and Davis,R.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 1194−1198.

Received on March 2, 1987; revised on April 6, 1987