Human Thrombomodulin: Complete cDNA Sequence and Chromosome Localization of the Gene†

Duanzhi Wen,‡,§ William A. Dittman,§ Richard D. Ye,‡,§ Larry L. Deaven,∥ Philip W. Majerus,§ and J. Evan Sadler*,‡,§

Howard Hughes Medical Institute and Departments of Medicine and Biological Chemistry, Washington University School of Medicine, St. Louis, Missouri 63110, and Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Received April 10, 1987; Revised Manuscript Received May 6, 1987

ABSTRACT: A human umbilical vein endothelial cell cDNA library in λgt11 was screened for expression of thrombomodulin antigens with affinity-purified rabbit polyclonal anti-thrombomodulin immunoglobulin G (IgG) and mouse monoclonal anti-human thrombomodulin IgG. Among 7 million recombinant clones screened, 12 were recognized by both antibodies. Two of these, λHTm10 and λHTm12, were shown to encode thrombomodulin by comparison of the amino acid sequence deduced from the nucleotide sequence to the amino acid sequence determined directly from tryptic peptides of thrombomodulin. Thrombomodulin mRNA was estimated to be 3.7 kilobases in length by Northern blot analysis of endothelial cell and placental poly(A)+ RNA. Thrombomodulin mRNA was not detected in human brain, HepG2 hepatoma cells, or the monocytic U937 cell line. Additional cDNA clones were selected by hybridization with the 1.2-kilobase insert of λHTm10. One isolate, λHTm15, contained a 3693 base pair cDNA insert with an apparent 5′-noncoding region of 146 base pairs, an open reading frame of 1725 base pairs, a stop codon, a 3′-noncoding region of 1779 base pairs, and a poly(A) tail of 40 base pairs. The cDNA sequence encodes a 60.3-kDa protein of 575 amino acids. The predicted protein sequence includes a signal peptide of ~21 amino acids, an amino-terminal ligand-binding domain of ~223 amino acids, an epidermal growth factor (EGF) homology region of 236 amino acids, a serine/threonine-rich segment of 34 amino acids, a membrane-spanning domain of 23 amino acids, and a cytoplasmic tail of 38 amino acids. The EGF-homology region consists of six tandemly repeated EGF-like domains. The organization of thrombomodulin is similar to that of the low-density lipoprotein receptor, and the protein is homologous to a large number of other proteins that also contain EGF-like domains, including factor VII, factor IX, factor X, factor XII, protein C, tissue plasminogen activator, and urokinase. The gene for thrombomodulin has been localized to chromosome 20 by hybridization of cDNA probes to purified human chromosomes.

Thrombomodulin is an endothelial cell surface glycoprotein that forms a 1:1 complex with thrombin. Binding of thrombin to this high-affinity receptor alters its specificity toward several substrates. Thrombin–thrombomodulin complexes activates protein C approximately 1000-fold faster than thrombin alone (Esmon et al., 1982). Activated protein C degrades clotting factors Va and VIIIa (Walker et al., 1979; Vehar & Davie, 1980); thus, thrombomodulin converts thrombin into a physiological anticoagulant.

Thrombomodulin has been purified from rabbit (Esmon et al., 1982), bovine (Suzuki et al., 1986; Jakubowski et al., 1986), and human lung (Maruyama et al., 1985a) and human placenta (Salem et al., 1984). The human protein has an apparent M_r of 75 000 (unreduced) that exhibits a characteristic shift to M_r 100 000 upon reduction with 2-mercaptoethanol (Salem et al., 1984). Immunohistochemical examination of tissue sections shows that thrombomodulin is widely distributed in the endothelium of arteries, veins, capillaries, and lymphatics (Maruyama et al., 1985b; DeBault et al., 1986). It is absent from the endothelium of human brain, hepatic sinusoids, and the post capillary venules of lymph nodes (Maruyama et al., 1985b; Ishii et al., 1986). In addition, human plasma and urine contain a soluble form of thrmobomodulin that appears to be smaller than cellular thrombomodulin, presumably because it lacks a membrane-binding domain (Ishii & Majerus, 1985).

Thrombomodulin is not a passive resident on the endothelial cell surface but is subject to endocytosis that is regulated by both thrombin and protein C. Binding of thrombin induces internalization of the thrombin–thrombomodulin complex, which is followed by degradation of thrombin and recycling of thrombomodulin to the plasma membrane (Maruyama & Majerus, 1985). Endocytosis of thrombin and throbomodulin may provide a mechanism to remove thrombin from the circulation and to prevent excessive protein C activation. Thrombin-stimulated endocytosis of thrombomodulin is inhibited by the substrate of the complex, protein C, but not by activated protein C (Maruyama & Majerus, 1987). Prolonged exposure to a variety of stimuli such as endotoxin (Morre et al., 1987), interleukin 1 (Nawroth et al., 1986), and tumor necrosis factor (Nawroth & Stern, 1986) decreases thrombomodulin activity and induces the expression of tissue factor on endothelial cells, suggesting another distinct mechanism to control the expression of thrmobomodulin activity in the vasculature.

The primary structure of thrombomodulin provides a foundation for understanding the biosynthesis, structure–

† This work was supported in part by Research Grants HLBI 14147 (Specialized Center for Research in Thrombosis), HLBI 16634 (to P.W.M.), and T32 HLBI 07088 from the National Institutes of Health and by the Monsanto-Washington University Biomedical Research Grant. R.D.Y. is supported by the Division of Biology and Biomedical Sciences, Washington University, St. Louis, MO.

* Address correspondence to this author.

‡ Howard Hughes Medical Institute, Washington University School of Medicine.

§ Departments of Medicine and Biological Chemistry, Washington University School of Medicine.

∥ Los Alamos National Laboratory.

0006-2960/87/0426-4350$01.50/0

function relationships, and regulation of this important natural anticoagulant. We report here the cloning and complete nucleotide sequence of essentially full-length cDNA isolates for human thrombomodulin as well as the chromosome localization of the gene. The predicted amino acid sequence indicates that thrombomodulin shares a number of structural features with another membrane receptor that undergoes endocytosis and recycling, the low-density lipoprotein (LDL)[1] receptor (Südhof et al., 1985). A preliminary report of this work has been presented (Wen et al., 1987).

Experimental Procedures

Materials. Deoxy-7-deazaguanosine 5′-triphosphate was from Boehringer Mannheim. ^{32}P-Labeled deoxyribonucleotides and deoxyadenosine 5′-([α-^{35}S]thiotriphosphate) were obtained from Amersham Radiochemicals. TPCK-trypsin was from Cooper Biomedical, Malvern, PA. Goat anti-rabbit IgG and goat anti-mouse IgG conjugated with alkaline phosphatase were from Promega Biotec. RNA-size standards were from Bethesda Research Laboratories.

Antibodies to Human Thrombomodulin. A monoclonal antibody to human thrombomodulin was isolated as previously described (Maruyama & Majerus, 1985). Polyclonal antibodies were prepared by injecting 50 μg of human thrombomodulin into each of two male rabbits by the method of Vaitukaitus (1981). Serum and purified IgG were prepared and assayed for inhibition of thrombomodulin functional activity (Salem et al., 1984). Polyclonal IgG was affinity purified by application of 2.3 mg of purified IgG onto a thrombomodulin–Affigel 15 column (165 μg of thrombomodulin/4 mL of Affigel 15) equilibrated with 50 mM HEPES, pH 7.5, and 150 mM NaCl. After the column was washed with the same buffer, bound IgG was eluted with 200 mM glycine, pH 3.0, and immediately dialyzed against 50 mM HEPES, pH 7.5, and 500 mM NaCl. The yield of antibody was 35 μg.

Preparation of Tryptic Peptides of Thrombomodulin. Human thrombomodulin was purified as previously described (Salem et al., 1984), except that immunoaffinity chromatography on polyclonal anti-thrombomodulin IgG–Affigel 10 was substituted for ion-exchange chromatography (Ishii & Majerus, 1985). Thrombomodulin (1.0–1.3 mg) was concentrated by ultrafiltration in a Micro-ProDiCon (Bio-Molecular Dynamics, Beaverton, OR) to 1 mg/mL and precipitated at −20 °C in 90% acetone. The precipitate was dried under a stream of nitrogen and redissolved in 0.4 M Tris-HCl and 0.1% (w/v) sodium dodecyl sulfate, pH 8.8. 2-Mercaptoethanol was added to 0.15 M, and the sample was incubated at 37 °C for 30 min. Iodoacetamide was added to 0.25 M, and the sample was incubated in the dark at 25 °C for 30 min; then, 2-mercaptoethanol was added to 0.3 M. The sample was again precipitated at −20 °C with 90% acetone and dried under nitrogen and then redissolved in 0.5 M ammonium bicarbonate, pH 8.5. Alternatively, in a different preparation after the initial precipitation, thrombomodulin was redissolved in 6 M guanidine and 0.5 M Tris-HCl, pH 8.8, and a 5-fold molar excess of dithiothreitol over cystine was added followed by incubation under nitrogen at 50 °C for 30 min. The pH was adjusted to 8.0 with HCl, a 5-fold molar excess of iodoacetamide was added, and the sample was incubated at 25 °C for 40 min. Excess iodoacetamide was reacted with a 3-fold molar excess of 2-mercaptoethanol, and the sample was dialyzed against 0.5 M ammonium bicarbonate, pH 8.5. TPCK-trypsin was added (1/100 w/w), and samples were incubated at 37 °C for 24 h and then lyophylized. The tryptic peptides were applied either directly to a reverse-phase HPLC column or first to a Mono-Q column (5 × 50 mm, Pharmacia) followed by a reverse-phase HPLC column. The Mono-Q column was equilibrated on a Varian 5000 HPLC system in 20 mM Tris-HCl, pH 9.0, and eluted with a linear 0–1 M NaCl gradient at 1 mL/min over 60 min. The effluent was monitored for absorbance at 215 nm. Fractions were pooled and lyophilized as indicated in the figure legends.

Samples were applied to a reverse-phase HPLC column (Unimetrics-Knauer, Licosorb RP-8, 5 μm) equilibrated with 0.1% trifluoroacetic acid and eluted with a gradient of 0–15% (v/v) acetonitrile in 0.1% trifluoroacetic acid over 10 min and then to 40% (v/v) acetonitrile–0.1% trifluoroacetic acid over 75 min at 0.7 mL/min. The effluent was monitored for absorbance at 215 nm. Individual peaks were pooled and evaporated to near dryness under a stream of nitrogen and sequenced on an Applied Biosystems Model 470A gas-phase protein sequencer (Hunkapiller et al., 1983; Hunkapiller & Hood, 1983).

Isolation of cDNA Clones for Human Thrombomodulin. The human umbilical vein endothelial cDNA library in λgt11, procedures for screening cDNA libraries with antibodies, preparation and use of synthetic oligonucleotides and cDNA restriction fragment probes, plaque purification of λ-phage, and preparation of λ-phage DNA were as described previously (Ye et al., 1987), with the exception that the goat anti-rabbit or goat anti-mouse detecting antibody was conjugated with alkaline phosphatase (Blake et al., 1984). The affinity-purified rabbit anti-human thrombomodulin was used at a concentration of 0.1 μg/mL. The monoclonal mouse anti-human thrombomodulin was used at a concentration of 2 μg/mL.

DNA Sequence Analysis. DNA restriction fragments were subcloned into pUC18, pUC19, M13mp18, or M13mp19 as described previously (Ye et al., 1987). Nucleotide sequence was determined on both strands by the dideoxy method (Sanger et al., 1977) using deoxyadenosine 5′-([α-^{35}S]thiotriphosphate) and buffer-gradient gels (Biggin et al., 1983). Deletions were generated with exonuclease III (Henikoff, 1984). Remaining gaps were filled by sequencing with synthetic oligonucleotide primers. Deoxyguanosine 5′-triphosphate in the sequencing reaction was substituted by deoxy-7-deazaguanosine 5′-triphosphate to increase the accuracy of sequencing in G-C-rich regions (Mizusawa et al., 1986). The few persistent compressions were resolved by performing electrophoresis on 6% (w/v) acrylamide gels containing 7 M urea and 40% (v/v) formamide.

Northern Blot Analysis. Poly(A)+ RNA was prepared from human term placenta (10 μg), human umbilical vein endothelial cells (5 μg), HepG2 cells (10 μg), and U937 cells (10 μg) cultured in the presence or absence of phorbol 12-myristate 13-acetate as previously described (Ye et al., 1987). RNA prepared from human brain (10 μg) was a generous gift of Karen O'Malley (Washington University). Electrophoresis through agarose in the presence of formaldehyde, transfer to nitrocellulose, and hybridization were performed as described previously (Ye et al., 1987). For each source that did not yield a hybridization signal for human thrombomodulin, control hybridization with either human γ-actin (Gunning et al., 1983) or human tissue factor cDNA (Scarpati et al., 1987) confirmed that RNA had been transferred efficiently.

Chromosome Localization of the Human Thrombomodulin Gene. Human chromosome suspensions were prepared (Sillar

[1] Abbreviations: HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; EGF, epidermal growth factor; LDL, low-density lipoprotein; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride; kb, kilobase(s).

& Young, 1981), stained with Hoechst 33258 and chromomycin A3, sorted with a dual-laser flow cytometer, and hybridized to cDNA probes, and signals were detected as previously described (Bartholdi et al., 1987; Murray et al., 1987). The cDNA insert of λHTm10 was labeled with Klenow fragment (Feinberg & Vogelstein, 1983) to a specific activity of $\geq 10^9$ cpm/μg. Two complete filter sets of the 22 autosomes and both sex chromosomes were examined, as well as Southern blots of human genomic DNA digested with *Eco*RI (Chomczynski & Qasba, 1984).

Computer Analysis of Sequences. The human thrombomodulin protein sequence was compared to all entries in the NBRF Protein Sequence Database (Georgetown University, Washington, DC, release 11.0, December 4, 1986) with the computer programs SEARCH (Dayhoff et al., 1983) and FASTP (Lipman & Pearson, 1985). The nucleotide sequence of cDNA isolated λHTm15 was compared to all entries in the Genbank genetic sequence data bank (BBN Laboratories Inc., Cambridge, MA, release 48.0, February 16, 1987) with the program FASTN (Lipman & Pearson, 1985). The alignment of EGF-like domains in thrombomodulin and other proteins, including a partial sequence of bovine thrombomodulin (Jackman et al., 1986), was performed with the programs RELATE and ALIGN (Dayhoff et al., 1983). Nucleotide sequences of human and bovine thrombomodulin were aligned with the program NUCALN (Wilbur & Lipman, 1983). Hydropathy or hydrophilicity profiles of the human thrombomodulin precursor were computed by the methods of Hopp and Woods (1981) and Kyte and Doolittle (1982).

RESULTS

Preparation and Sequencing of Tryptic Peptides of Human Thrombomodulin. Thrombomodulin was purified to homogeneity from human placenta, reduced and carboxyamidomethylated, and digested with bovine trypsin. The resultant complex mixture of peptides was resolved by reverse-phase HPLC (Figure 1a), yielding peptide T-R1. Subsequent tryptic digests were first separated by anion-exchange chromatography, and pools of the individual peaks were further resolved by reverse-phase chromatography. For example, pool T-M1 from a Mono-Q column (Figure 1b) yielded homogeneous peptides T-M1-R1, T-M1-R2, and T-M1-R3 after reversed-phase chromatography (Figure 1c). One other peptide was isolated by similar methods as described under Experimental Procedures. The partial sequences of five peptides containing a total of 62 amino acid residues were determined.

Screening of an Endothelial Cell λgt11 cDNA Library and Chacterization of Recombinant Proteins. A human endothelial cell cDNA library in phage λgt11 was screened with affinity-purified rabbit anti-human thrombomodulin IgG. Among 7×10^6 independent recombinant clones screened, 12 expressed a fusion protein recognized both by the polyclonal antibody and also by a monoclonal antibody to human thrombomodulin (Maruyama & Majerus, 1985). The cDNA inserts of the 12 isolates were subcloned into pUC19 for further characterization. The length of the inserts ranged from 1.2 to 1.7 kb. Although all 12 isolates were initially selected by the same polyclonal and monoclonal antibodies, by cross-hybridization they fell into three unrelated groups of 1, 2, or 9 members. Representatives of each group were subsequenced for comparison with independently determined protein sequence. The 5′-sequence of one isolate from the two-member group, λHTm10, exactly encoded the sequence of peptide T-R1. Similarly, the 5′-sequence of the second member of this hybridization group, λHTm12, encoded peptides T-M1-R1 and T-MO-R1, and the 3′-sequences of both isolates overlapped,

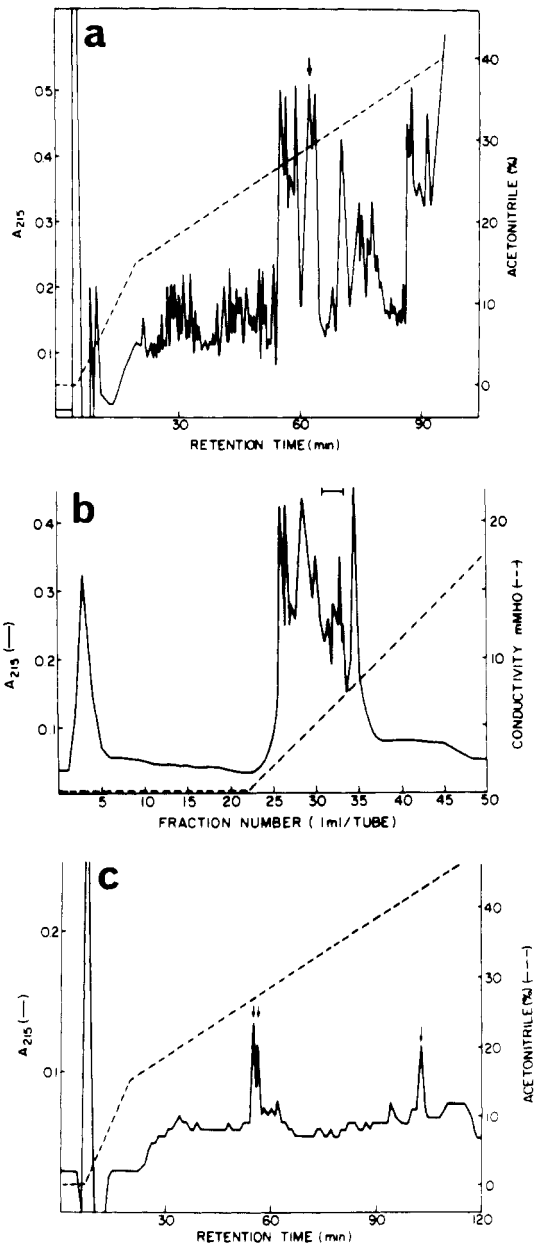

FIGURE 1: Separation of tryptic peptides of thrombomodulin. (a) A total tryptic digest of reduced, alkylated thrombomodulin was prepared and chromatographed as described under Experimental Procedures on a Licosorb RP-8 column. A total of 700 μg of protein was injected, and 1-min (0.7 mL) fractions were collected. The solid line (—) indicates absorbance at 215 nm, and the dashed line (- -) indicates the percent acetonitrile in the column buffer. The peak marked with the arrow (↓) corresponds to peptide T-R1. (b) A total tryptic digest of thrombomodulin was prepared and chromatographed on a Mono-Q column as described under Experimental Procedures. After injection of 700 μg of protein, 1-mL fractions were collected. The solid line (—) indicates absorbance at 215 nm, and the dashed line (- -) indicates conductivity of the column buffer. The limit buffer had a conductivity of 41 mΩ. The fractions indicated by the bar were pooled (T-M1) for further purification on a reversed-phase column as shown in panel c. (c) Pooled sample T-M1 from panel b was chromatographed on a Licosorb RP-8 column as in panel a. The doublet eluting at approximately 55 min yielded peptides T-M1-R1 and T-M1-R2, and the single peak at approximately 102 min yielded T-M1-R3.

confirming that they encoded human thrombomodulin.

The cDNA insert of clone λHTm10 was used to screen 1×10^6 recombinants from the endothelial cell cDNA library by plaque hybridization. Ninety positive clones were detected, and half of these were plaque purified. The 45 clones were then rescreened with an oligonucleotide probe corresponding

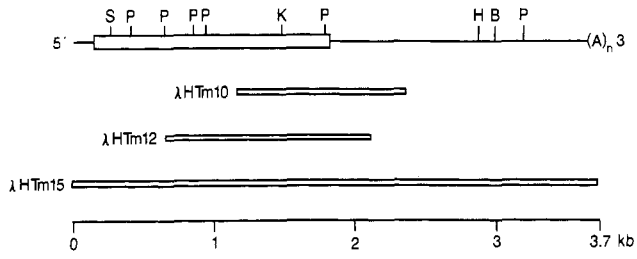

FIGURE 2: Restriction map of thrombomodulin cDNA isolates. The 5′ and 3′ ends of the restriction map are labeled. Selected restriction sites that have been useful in subcloning the cDNA inserts for sequencing are shown: *Bam*HI, B; *Hin*dIII, H; *Kpn*I, K; *Pst*I, P; *Sma*I, S. The thin segments indicate noncoding sequences, and the thick segment indicates the open reading frame that encodes thrombomodulin. The portion of the sequence contained in each of the cDNA isolates, λHTm10, λHTm12, and λHTm15, is shown by the thin open bars. The scale is in kilobases (kb).

to the 5′-end of clone λHTm12. Nine positive clones were detected, and among them four clones contained a poly(A) tail as determined by hybridization to an oligo(dT) probe. Restriction analysis showed that of these four clones λHTm15 contained the largest cDNA insert of 3.7 kb. The relationship of the three cDNA isolates, λHTm10, λHTm12, and λHTm15, to the restriction map of the full-length thrombomodulin cDNA is shown in Figure 2.

Nucleotide Sequence of Thrombomodulin cDNA Isolates. The cDNA insert of λHTm12 (1.5 kb) was sequenced completely on both strands. The cDNA insert of λHTm15 (3.7 kb) was sequenced on at least one strand, and those regions at the 5′ and 3′ ends that did not overlap with λHTm12 were sequenced on both strands. Thus, the complete sequence was determined at least once on both strands. The coding sequence of thrombomodulin has an extremely high G+C content of 68% that made sequencing difficult due to frequent compressions. All of these were resolved by the use of deoxy-7-deazaguanosine 5′-triphosphate in place of dGTP (Mizusawa et al., 1986) except for a short sequence between nucleotides 395 and 405. This sequence was resolved unambiguosly on both strands by the use of sequencing gels containing 40% (v/v) formamide in addition to urea. In contrast to the high G+C content of the coding sequence, the 3′-noncoding sequence is only ~43% G+C.

The nucleotide and translated amino acid sequence of λHTm15 is shown in Figure 3. The first ATG codon occurs at nucleotide 147, embedded in a sequence that agrees well with the proposed optimal sequence for initiation by eukaryotic ribosomes, ACCATGG (Kozak, 1986). The preceding 146 nucleotides of the proposed 5′-noncoding sequence do not have a termination codon in the same reading frame as the ATG codon. This proposed initiator codon begins an open reading frame of 1725 nucleotides, followed by a TGA termination codon and 1779 additional nucleotides of 3′-noncoding sequence before a poly(A) tail of 40 nucleotides. There are four potential polyadenylation or processing signals with the sequence AATAAA (Proudfoot & Brownlee, 1981), the last of which begins 21 nucleotides before the poly(A) tail.

There is a single nucleotide difference between λHTm12 and λHTm15 at nucleotide 1564, which is a T in λHTm12 and a C in λHTm15. This alters the encoded amino acid sequence from Ala-473 to Val. This could be due to nucleotide sequence polymorphism or the result of an error in DNA replication during cDNA cloning.

Amino Acid Sequence of Thrombomodulin and Homology to Other Proteins. The cDNA sequence encodes a protein of 575 amino acids with a calculated M_r of 60 328. There are five potential N-glycosylation sites with the sequence Asn-X-Ser/Thr (Figure 3). The amino-terminal ~21 residues are hydrophobic with the characteristics of a typical signal peptide (von Heijne, 1983, 1985). The site of cleavage by signal peptidase is difficult to predict. The potential site indicated in Figure 3 between Ala-21 and Glu-22 gives the highest score (14.0) by the weighted matrix method of von Heijne (1983) but suffers from the presence of proline at position −2. The only other site that scores nearly as highly (12.7) is between Ala-19 and Pro-21, but this site also places a proline residue at the unfavorable +1 position. The remainder of the protein sequence contains the sequence of five tryptic peptides isolated from human placental thrombomodulin (Figure 3).

The signal peptide is followed by a relatively cysteine-poor domain of ~223 amino acids and a cysteine-rich region of 236 residues composed of six tandem EGF-like repeats of ~40 residues each. The alignment of these EGF repeats with a representative domain from the human EGF precursor (Bell et al., 1986) is shown in Figure 4. The amino-terminal and EGF-homology regions are followed sequentially by a serine/threonine-rich domain of 34 amino acids, a hydrophobic segment of 23 amino acids that may span the plasma membrane, and a proposed cytoplasmic tail of 38 amino acids.

The amino acid sequence of thrombomodulin was compared to all sequences in the NBRF Protein Sequence Database, and the nucleotide sequence of λHTm15 was compared to all entries in the Genbank Genetic Sequence Data Bank. Aside from proteins containing EGF-like domains, no protein or DNA sequences showed significantly similar to human thrombomodulin.

Size and Occurrence of Thrombomodulin mRNA in Tissues and Cultured Cells. The distribution of mRNA for human thrombomodulin was studied by Northern blotting. A single mRNA species of 3.7 kb was detected in human placenta and endothelial cell poly(A)+ RNA. Thrombomodulin mRNA was not detected in poly(A)+ RNA from human hepatoma HepG2 cells or the monocytic U937 cell line. In addition, no hybridization was detected with 10 μg of human brain poly(A)+ RNA (data not shown).

Chromosome Localization of the Thrombomodulin Gene. The insert of λHTm10 was hybridized to human chromosomes purified by fluorescence-activated flow sorting. Two complete sets of 22 autosomes and the X and Y chromosomes were tested, both of which gave signals only with chromosome 20 (Figure 5).

Comparison of Human and Partial Bovine Thrombomodulin cDNA Sequences. A partial cDNA sequence encoding the carboxy-terminal 356 amino acids of bovine thrombomodulin has been reported (Jackman et al., 1986), and this sequence shares ~229 identical amino acids (64%) with the homologous segment of human thrombomodulin. The bovine protein appears to have two insertions totaling eight amino acids in the serine/threonine-rich domain that are absent in the human sequence. The human protein has two additional residues in the second EGF repeat and one in the sixth EGF repeat relative to the bovine sequence. The nucleotide sequences in the coding region of the bovine and human cDNAs are ~75% identical, and the gaps introduced to optimize the alignment with the program NUCALN (Wilbur & Lipman, 1983) match those in the protein sequences with ALIGN (Dayhoff et al., 1983). The 3′-noncoding sequences for bovine and human thrombomodulin are ~68% identical.

DISCUSSION

We have characterized an essentially full-length cDNA isolate for human thrombomodulin. The longest cDNA insert contains a poly(A) tail and is 3.7 kb in length, compared to

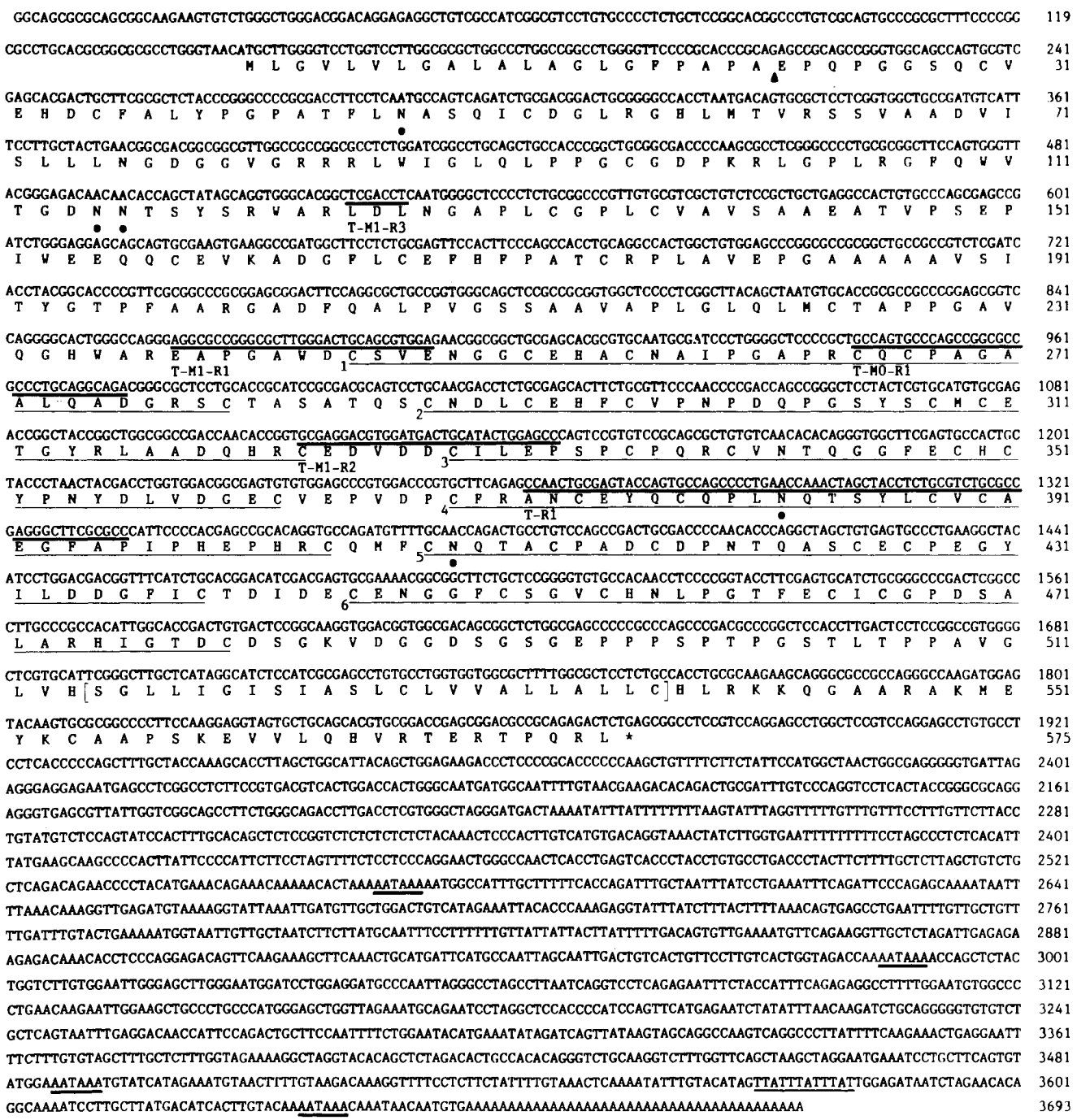

FIGURE 3: Nucleotide and translated amino acid sequence of human thrombomodulin cDNA isolated λHTm15. Nucleotides and amino acids are numbered on the right. Nucleotide 1 was assigned to the first residue of the cDNA insert, and amino acid 1 was assigned to the first methionine of the open reading frame that encodes thrombomodulin peptide sequences. A potential signal peptidase cleavage site is indicated by the triangle (▲). Potential N-linked glycosylation sites are marked by filled circles (●). Sequences that match that determined for tryptic peptides of thrombomodulin are marked by bold overlining. The six EGF-like repeats are underlined and numbered. The transmembrane domain is enclosed in brackets ([]). Potential polyadenylation or processing signals AATAAA are marked with bold underlining. The pair of overlapping TTATTTAT sequences (Caput et al., 1986) are marked by thin underlining. The sequence represented in λHTm10 includes nucleotides 1208–2403, and λHTm12 includes nucleotides 671–2142.

the single 3.7-kb mRNA species identified by Northern blotting. There is no stop codon 5′ of the first methionine encoded by the cDNA sequence, but this ATG is followed by an amino acid sequence that is typical of cleavable signal peptides (Figure 3). The open reading frame that includes this signal peptide encodes a protein of 575 amino acids that contains the sequence of five tryptic peptides determined directly by protein sequencing (Figure 3). The distribution of thrombomodulin mRNA in the cultured cells and tissues that have been examined is consistent with the previously reported restriction of thrombomodulin antigen and activity to endothelial cells and the absence of the protein from brain (Maruyama et al., 1985b; Ishii et al., 1986). Southern blots of human genomic DNA digested with *Eco*RI probed with the cDNA insert of λHTm10 show a single hybridizing band. The human genome appears to contain a single copy of the thrombomodulin gene, which has been localized to chromosome 20 (Figure 5).

Two additional classes of cDNA clones that do not correspond to thrombomodulin were identified by polyclonal and monoclonal antibodies to thrombomodulin. Partial sequences have been obtained for representatives of these groups, and

Tm	#1	C S V E N G G C E H A C - - - N A I P G A P R C Q C P A G A A L Q A - D G R S C
	#2	C - - - N D L C E H F C V P N P D Q P G S Y S C M C E T G Y R L A A - D Q H R C
	#3	C I L E P S P C P Q R C V N T - - - Q G G F E C H C Y P N Y D L V - - D G E - C
	#4	C F R A N - - C E Y Q C Q P L N Q T - - S Y L C V C A E G F A P I P H E P H R C
	#5	C - - N Q T A C P A D C D P N T Q A - - S - - C E C P E G Y I L - - D D G F I C
	#6	C - E N G G F C S G V C - - - H N L P G T F E C I C G P D S A L A R H I G T D C
EGFpre	#3	C P R N V S E C S H D C V L T S E - - - G P L C F C P E G S V L E R - D G K T C

FIGURE 4: Alignment of the EGF-like repeats of human thrombomodulin. The EGF-like repeats of thrombomodulin are numbered 1–6 as in Figure 3. The bottom line is the sequence of the third EGF-like domain from the human EGF precursor, residues 401–436 (Bell et al., 1986), a representative sequence for comparison. Dashes (–) represent gaps introduced to optimize the alignment. Residues identical in two or more of the aligned sequences are enclosed in the boxed outlines.

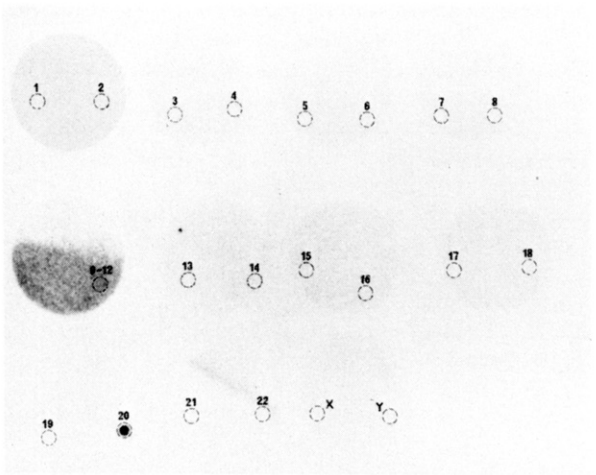

FIGURE 5: Chromosome localization of the human thrombomodulin gene. The characters 1–22, X, and Y on the nitrocellulose filters indicate the human chromosomes present in each spot that is outlined in the dashed circles.

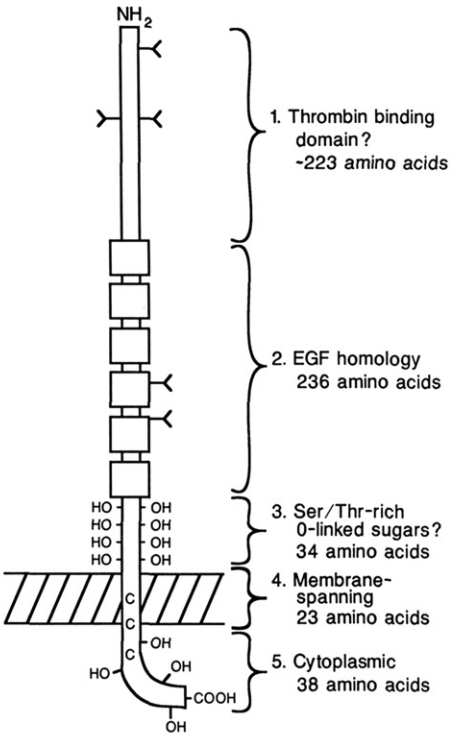

FIGURE 6: Structural domains of human thrombomodulin. The organization of thrombomodulin is depicted schematically, with the amino terminus (NH_2) and carboxy terminus (COOH) of the protein labeled. Potential N-glycosylation sites are indicated (Y). Hydroxy amino acids in the serine/threonine-rich domain and the cytoplasmic tail are shown (–OH). Cysteine residues in the transmembrane and cytoplasmic domains are also indicated (C).

they do not appear to be related to thrombomodulin, nor are they homologous to known protein or nucleic acid sequences (unpublished data). At present we cannot explain why these apparently specific immunlogical reagents have identified three distinct proteins that are expressed in a human endothelial cell cDNA library.

The protein encoded by the thrombomodulin cDNA sequence has a calculated mass of 60.3 kDa including the signal sequence, compared to 75 kDa (unreduced) and 100 kDa (reduced) for human thrombomodulin purified from placenta (Salem et al., 1984). This suggests that the natural protein may be extensively glycosylated and is consistent with a report that the apparent size of a bovine thrombomodulin fragment decreases from 45 to 19 kDa upon removal of carbohydrate with trifluoromethanesulfonic acid (Jackman et al., 1986). The human sequence contains five potential sites of N-glycosylation (Figure 3).

The thrombomodulin precursor is organized into domains (Figure 6) that resemble those of the LDL receptor (Südhof et al., 1985). Many of these features were apparent from the partial cDNA sequence reported recently for bovine thrombomodulin (Jackman et al., 1986). The signal peptide is followed by a proposed amino-terminal ligand-binding domain. The boundary between the signal peptide and the mature protein is not known. Our repeated efforts to sequence intact human thrombomodulin have been unsuccessful, suggesting that the amino terminus is blocked.

The proposed ligand binding domain is succeeded by an EGF-homology domain composed of six tandemly arranged EGF-like sequences (Figure 3). Similar EGF-like domains are found in many other proteins, including the LDL receptor (Russell et al., 1984), factor VII (Hagen et al., 1986), factor IX (Katayama et al., 1979), factor X (Enfield et al., 1980), factor XII (McMullen & Fujikawa, 1985), tissue plasminogen activator (Pennica et al., 1983), urokinase (Günzler et al., 1982), tumor-derived growth factor α (Marquardt et al., 1984), protein C (Fernlund & Stenflo, 1982), protein S (Dahlbäck et al., 1986), protein Z (Hojrup et al., 1985), vaccinia virus 19-kDa protein (Blomquist et al., 1984), and complement component C1r (Gagnon & Arlaud, 1985). In addition, several neurogenic genes of *Drosophila melanogaster* (Knust et al., 1987; Wharton et al., 1985) encode tandem arrays of up to 36 EGF-like sequences, and the *Lin-12* locus of *Caenorhabditis elegans* encodes at least 11 EGF-like repeats (Greenwald, 1985). For several of these proteins, the EGF-like sequences are part of a domain that binds to a specific receptor. The function of these structural elements in other proteins is not known.

The next domain is a serine/threonine-rich segment that contains 8 hydroxy amino acids among 34 residues. The corresponding domaine of the LDL receptor has many clustered O-linked carbohydrate chains (Russell et al., 1984), and this domain of thrombomodulin may be glycosylated as well.

This segment shows the most variation between human and bovine thrombomodulin. The bovine sequence contains one fewer serine/threonine residue but eight additional amino acids in two insertions (Jackman et al., 1986).

The fourth domain (Figure 6) is a potential transmembrane domain of 23 hydrophobic amino acids flanked by positively charged histidine residues. The hydrophobic character of this domain and the signal peptide was confirmed by examination of hydrophilicity or hydropathy profiles, and no other candidate membrane-spanning segments were identified (data not shown). This is one of the most highly conserved segments of the human and bovine sequences, with only three amino acid differences. This degree of conservation is not seen for the transmembrane domain of the bovine and human LDL receptors (Goldstein et al., 1985). Both the LDL receptor and thrombomodulin are subject to endocytosis. The LDL receptor is continuously internalized and recycled, but thrombomodulin has the additional property that endocytosis is regulated by its ligands. Thrombin stimulates (Maruyama & Majerus, 1985) and protein C inhibits (Maruyama & Majerus, 1987) the internalization of thrombomodulin. The lack of variation in the sequence of the bovine and human transmembrane domains suggests that this sequence may be important for transmitting the signals necessary for the regulation of thrombomodulin endocytosis by the binding of ligands.

The carboxy-terminal cytoplasmic domain is composed of only 38 amino acids. Notable features of this sequence include one tyrosine, one serine, and two threonine residues that could be substrates for cellular protein kinases and a cysteine residue that might be modified by acylation. The functional importance of a tyrosine residue in the cytoplasmic tail of the LDL receptor has been demonstrated by the characterization of a missense Tyr → Cys mutant that fails to cluster in coated pits and is not internalized (Davis et al., 1986). This result suggests that the cytoplasmic domain may direct thrombomodulin to coated pits and that specific residues may participate in the regulation of this process by thrombin and protein C.

Comparison of the 3′-noncoding sequences of bovine (Jackman et al., 1986) and human thrombomodulin demonstrates that this region is highly conserved, in contrast to many other interspecies comparisons of noncoding sequences. There is a strikingly conserved segment of 168 nucleotides just 5′ of the polyadenylation site (bovine 2675–2842, human 3475–3642) in which there is 87% identity (22 differences) between the sequences. These 3′-noncoding sequences are exceptionally rich in AT content and show superficial similarity to the 3′-noncoding regions of several lymphokine and protonocogene mRNAs (Shaw & Kamen, 1986; Caput et al., 1986). Sequences rich in AT residues and poor in GC residues have been shown to promote the degradation of certain mRNA species (Shaw & Kamen, 1986). A specific octanucleotide consensus sequence (TTATTTAT) has been proposed to be characteristic of the 3′-untranslated region of mRNAs encoding several inflammatory mediators (Caput et al., 1986), and this sequence is present in two overlapping copies in the human thrombomodulin cDNA isolate λHTm15 between nucleotides 3568 and 3579 (Figure 3) and also in the corresponding segment of the bovine thrombomodulin cDNA (Jackman et al., 1986). Thrombomodulin activity disappears within a few hours from the endothelial cell surface after exposure of cells to tumor necrosis factor (Nawroth & Stern, 1986), interleukin 1 (Nawroth et al., 1986), or endotoxin (Moore et al., 1987), suggesting that thrombomodulin mRNA might have a comparably short half-life. Since the 3′-noncoding regions appear to influence the rate of degradation of several mRNAs, the conserved segments of the 3′-noncoding region of thrombomodulin may be important for the rapid modulation of thrombomodulin activity in response to these stimuli.

ACKNOWLEDGMENTS

We thank Gregory A. Grant for performing amino acid analyses and for the protein sequencing of tryptic peptides of thrombomodulin and Karen O'Malley for providing human poly(A)+ RNA.

REFERENCES

Bartholdi, M., Meyne, J., Albright, K., Luedermann, M., Campbell, E., Chritton, D., Deaven, L., vanDilla, M., & Cram, S. (1987) *Methods Enzymol.* (in press).

Bell, G. I., Fong, N. M., Stempien, M. M., Wormsted, M. A., Caput, D., Ku, L., Urdea, M. S., Rall, L. B., & Sanchez-Pescador, R. (1986) *Nucleic Acids Res. 14*, 8427–8446.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Blake, M. S., Johnston, K. H., Russell-Jones, G. J., & Gotschlich, E. C. (1984) *Anal. Biochem. 136*, 175–179.

Blomquist, M. C., Hunt, L. T., & Barker, W. C. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 7363–7367.

Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S., & Cerami, A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 1670–1674.

Chomczynski, P., & Qasba, P. K. (1984) *Biochem. Biophys. Res. Commun. 122*, 340–344.

Dahlbäck, B., Lundwall, Å., & Stenflo, J. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4199–4203.

Davis, C. G., Lehrman, M. A., Russell, D. W., Anderson, R. G. W., Brown, M. S., & Goldstein, J. L. (1986) *Cell (Cambridge, Mass.) 45*, 15–24.

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

DeBault, L. E., Esmon, N. L., Olson, J. R., & Esmon, C. T. (1986) *Lab. Invest. 54*, 172–178.

Enfield, D. L., Ericsson, L. H., Fujikawa, K., Walsh, K. A., Neurath, H., & Titani, K. (1980) *Biochemistry 19*, 659–667.

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

Feinberg, A. P., & Vogelstein, B. (1983) *Anal. Biochem. 132*, 6–10.

Fernlund, P., & Stenflo, J. (1982) *J. Biol. Chem. 257*, 12170–12179.

Gagnon, J., & Arlaud, G. J. (1985) *Biochem. J. 225*, 135–142.

Goldstein, J. L., Brown, M. S., Anderson, R. G. W., Russell, D. W., & Schneider, W. J. (1985) *Annu. Rev. Cell Biol. 1*, 1–39.

Greenwald, I. (1985) *Cell (Cambridge, Mass.) 43*, 583–590.

Gunning, P., Ponte, P., Okayama, H., Engel, J., Blau, H., & Kedes, L. (1983) *Mol. Cell Biol. 3*, 787–795.

Günzler, W. A., Steffens, G. J., Ötting, F., Kim, S.-M. A., Frankus, E., & Flohé, L. (1982) *Hoppe-Seyler's Z. Physiol. Chem. 363*, 1155–1165.

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., & Davie, E. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2412–2416.

Henikoff, S. (1984) *Gene 28*, 351–359.

Hojrup, P., Jensen, M. S., & Petersen, T. E. (1985) *FEBS Lett. 184*, 333–338.

Hopp, T. P., & Woods, K. R. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 3824–3828.

Hunkapiller, M. W., & Hood, L. E. (1983) *Methods Enzymol. 91*, 486–493.
Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J., & Hood, L. E. (1983) *Methods Enzymol. 91*, 399–413.
Ishii, H., & Majerus, P. W. (1985) *J. Clin. Invest. 76*, 2178–2181.
Ishii, H., Salem, H. H., Bell, C. E., Laposata, E. A., & Majerus, P. W. (1986) *Blood 67*, 362–365.
Jackman, R. W., Beeler, D. L., VanDeWater, L., & Rosenberg, R. D. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8834–8838.
Jakubowski, H. V., Kline, M. D., & Owen, W. G. (1986) *J. Biol. Chem. 261*, 3876–3882.
Katayama, K., Ericsson, L. H., Enfield, D. L., Walsh, D. A., Neurath, H., Davie, E. W., & Titani, K. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 4990–4994.
Knust, E., Dietrich, U., Tepass, U., Bremer, K. A., Weigel, D., Vässin, H., & Campos-Ortega, J. A. (1987) *EMBO J. 6*, 761–766.
Kozak, M. (1986) *Cell (Cambridge, Mass.) 44*, 283–292.
Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.
Lipman, D. J., & Pearson, W. R. (1985) *Science (Washington, D.C.) 227*, 1435–1441.
Marquardt, H., Hunkapiller, M. W., Hood, L. E., Twardzik, D. R., De Larco, J. E., Stephenson, J. R., & Todaro, G. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4684–4688.
Maruyama, I., & Majerus, P. W. (1985) *J. Biol. Chem. 260*, 15432–15438.
Maruyama, I., & Majerus, P. W. (1987) *Blood 69*, 1481–1484.
Maruyama, I., Salem, H. H., Ishii, H., & Majerus, P. W. (1985a) *J. Clin. Invest. 75*, 987–991.
Maruyama, I., Bell, C. E., & Majerus, P. W. (1985b) *J. Cell Biol. 101*, 363–371.
McMullen, B. A., & Fujikawa, K. (1985) *J. Biol. Chem. 260*, 5328–5341.
Mizusawa, S., Nishimura, S., & Seela, F. (1986) *Nucleic Acids Res. 14*, 1319–1324.
Moore, K. L., Anderoli, S. P., Esmon, N. L., Esmon, C. T., & Bang, N. U. (1987) *J. Clin. Invest. 79*, 124–130.
Murray, S. S., Deaven, L. L., Burton, D. W., O'Connor, D. T., Mellon, P. L., & Deftos, L. J. (1987) *Biochem. Biophys. Res. Commun. 142*, 141–146.
Nawroth, P. P., & Stern, D. M. (1986) *J. Exp. Med. 163*, 740–745.
Nawroth, P. P., Handley, D. A., Esmon, C. T., & Stern, D. M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3460–3464.
Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., & Collen, D. (1983) *Nature (London) 301*, 214–221.
Proudfoot, N., & Brownlee, G. (1981) *Nature (London) 252*, 359–362.
Russell, D. W., Schneider, W. J., Yamamoto, T., Luskey, K. L., Brown, M. S., & Goldstein, J. L. (1984) *Cell (Cambridge, Mass.) 37*, 577–585.
Salem, H. H., Maruyama, I., Ishii, H., & Majerus, P. W. (1984) *J. Biol. Chem. 259*, 12246–12251.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Scarpati, E. M., Broze, G. J., Miletich, J. P., Lebo, R. V., & Sadler, J. E. (1987) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 46*, 2242.
Shaw, G., & Kamen, R. (1986) *Cell (Cambridge, Mass.) 46*, 659–667.
Sillar, R., & Young, B. D. (1981) *J. Histochem. Cytochem. 29*, 74.
Südhof, T. C., Goldstein, J. C., Brown, M. S., & Russell, D. W. (1985) *Science (Washington, D.C.) 228*, 815–822.
Suzuki, K., Kusumoto, H., & Hashimoto, S. (1986) *Biochim. Biophys. Acta 882*, 343–352.
Vaitukaitus, J. L. (1981) *Methods Enzymol. 73*, 46–52.
Vehar, G. A., & Davie, E. W. (1980) *Biochemistry 19*, 401–409.
von Heijne, G. (1983) *Eur. J. Biochem. 133*, 17–21.
von Heijne, G. (1985) *J. Mol. Biol. 184*, 99–105.
Walker, J. J., Sexton, P. W., & Esmon, C. T. (1979) *Biochim. Biophys. Acta 571*, 333–342.
Wen, D., Dittman, W. A., Ye, R. D., Majerus, P. W., & Sadler, J. E. (1987) *Clin. Res. 35*, 603A.
Wharton, K. A., Johansen, K. M., Xu, T., & Artavanis-Tsakonas, S. (1985) *Cell (Cambridge, Mass.) 43*, 567–581.
Wilbur, W. J., & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 726–730.
Ye, R. D., Wun, T.-C., & Sadler, J. E. (1987) *J. Biol. Chem. 262*, 3718–3725.