# The Structure of Human Thrombospondin, an Adhesive Glycoprotein with Multiple Calcium-binding Sites and Homologies with Several Different Proteins

**Jack Lawler and Richard O. Hynes**

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; Department of Biomedical Research and Medicine, Division of Hematology/Oncology, St. Elizabeth's Hospital of Boston, Tufts University School of Medicine, Boston, Massachusetts 02135

*Abstract.* Thrombospondin is one of a class of adhesive glycoproteins that mediate cell-to-cell and cell-to-matrix interactions. We have used two monoclonal antibodies to isolate cDNA clones of thrombospondin from a human endothelial cell cDNA library and have determined the complete nucleotide sequence of the coding region. Three regions of known amino acid sequence of human platelet thrombospondin confirm that the clones are authentic. Three types of repeating amino acid sequence are present in thrombospondin. The first is 57 amino acids long and shows homology with circumsporozoite protein from *Plasmodium falciparum*. The second is 50–60 amino acids long and shows homology with epidermal growth factor precursor. The third occurs as a continuous eightfold repeat of a 38-residue sequence; structural homology with parvalbumin and calmodulin indicates that these repeats constitute the multiple calcium-binding sites of thrombospondin. The amino acid sequence arg-gly-asp-ala is included in the last type 3 repeat. This sequence is probably the site for the association of thrombospondin with cells. In addition, localized homologies with procollagen, fibronectin, and von Willebrand factor are present in one region of the thrombospondin molecule.

THROMBOSPONDIN is a 420,000-D glycoprotein that was first identified in human blood platelets and then shown to be synthesized and secreted by various cells in culture (for a review see Lawler, 1986). Thrombospondin secreted from activated platelets becomes associated with the platelet membrane and incorporated into the developing fibrin clot (Bale et al., 1985; Murphy-Ullrich and Mosher, 1985; Wolff et al., 1986). Thrombospondin secreted by cells in culture is incorporated into the extracellular matrix (Raugi et al., 1982; Jaffe et al., 1983; McKeown-Longo et al., 1984; Majack et al., 1985). In vitro binding studies indicate that thrombospondin can bind to fibrinogen, fibronectin, laminin, and type V collagen (Leung and Nachman, 1982; Lahav et al., 1982, 1984; Mumby et al., 1984; Lawler et al., 1986*b*). These data suggest that thrombospondin is a member of a class of adhesive proteins that mediate cell-to-cell and cell-to-matrix interactions (Hynes, 1985).

Thrombospondin is composed of three polypeptide chains that are cross-linked by disulfide bonds and appear to be identical in terms of molecular weight, position of cleavage sites for thrombin, plasmin, thermolysin, and trypsin, and $NH_2$-terminal amino acid sequence (Lawler and Slayter, 1981; Dixit et al., 1984; Raugi et al., 1984; Coligan and Slayter, 1984; Lawler et al., 1985; Galvin et al., 1985). Recent immunological, biochemical, and electron microscopic data permit the formulation of models for the structure of human platelet thrombospondin (Lawler et al., 1985; Galvin et al., 1985). Electron microscopy of replicas produced by low angle rotary shadowing indicates that the thrombospondin molecule can be divided into four distinct structural regions; globular region N, a region where the chains are cross-linked, a thin, connecting region, and globular region C. Globular region N is composed of three 25,000-D segments of polypeptide which are the $NH_2$-terminal segments of each chain. This region has been shown to mediate the binding of thrombospondin to heparin (Lawler and Slayter, 1981; Dixit et al., 1984). In addition, polyclonal antibodies prepared against globular region N inhibit platelet aggregation (Gartner et al., 1984).

The thin, connecting region appears, by electron microscopy, to be flexible and its length increases by ~30% when calcium is removed from the molecule (Lawler et al., 1985; Dixit et al., 1986). Chymotryptic digestion of thrombospondin in the absence of calcium produces a 210,000-D trimeric structure composed of the thin, connecting regions from each of the three chains and the site where the chains are cross-linked (Mumby et al., 1984; Galvin et al., 1985; Lawler et al., 1986*b*). This fragment has been shown to bind type V collagen, fibronectin, fibrinogen, plasminogen, and laminin in solid-phase binding assays (Mumby et al., 1984; Lawler et al., 1986*b*).

The final distinct structural region of thrombospondin is

globular region C. This region is 118–170 Å in diameter and appears at the ends of each of the thin, connecting regions (Lawler et al., 1985; Galvin et al., 1985). Dixit et al. (1985) have reported that a monoclonal antibody against this region of thrombospondin inhibits platelet aggregation. These data, in conjunction with the data of Gartner et al. (1984), suggest that multiple sites in the thrombospondin molecule are involved in platelet aggregation. In the presence of EDTA, globular region C decreases in size, concomitant with the increase in length of the thin, connecting region, suggesting that there is a redistribution of mass from globular region C to the thin, connecting region (Lawler et al., 1985). This conformational change can also be detected by changes in sedimentation coefficient, intrinsic viscosity, circular dichroism, and the peptide pattern produced by limited tryptic digestion (Lawler et al., 1982; Lawler and Simons, 1983). These methods indicate that the transition occurs at 50–120 μM calcium concentration by a cooperative mechanism that involves at least 12 calcium-binding sites.

While these biochemical and electron microscopic studies have helped to define the overall shape of the thrombospondin molecule, very little is known of the primary or secondary structure. In this paper we report the cloning of thrombospondin from a cDNA library constructed from human endothelial cell mRNA in λgtll. The complete amino acid sequence has been determined from the nucleotide sequence of the coding region. These studies permit (*a*) the identification of internal repeating sequences within the thrombospondin molecule, (*b*) the comparison of thrombospondin structures with other proteins, and (*c*) interpretation of some of the functional properties of thrombospondin on a structural level.

## Materials and Methods

### Materials

A λgtll bacteriophage library of cDNA derived from cultured human umbilical vein endothelial cells and E. coli strains Y1088, Y1089, and Y1090 were kindly provided by Dr. Robert Handin. A pool of random hexanucleotides was used as the primer for cDNA synthesis (Ginsburg et al., 1985). Two monoclonal antibodies, designated MA-I and MA-II, which were raised against human platelet thrombospondin, were used in this study (Lawler et al., 1985). The epitope for MA-II is located in the $NH_2$-terminal 25,000-D heparin-binding fragment, while the epitope for MA-I is located in a region of polypeptide that is adjacent to a 25,000-D COOH-terminal tryptic fragment (Lawler et al., 1985). All enzymes were purchased from New England Biolabs (Beverly, MA) except where otherwise noted.

### Antibody Screening of Recombinant Phage

A total of 150,000 phage were adsorbed to E. coli strain Y1090, plated at a density of 20,000 plaque-forming units/15-cm L2-ampicillin plate, and grown at 42°C for 4 h (Young and Davis, 1985). Nitrocellulose filters (0.2 μm; Schleicher & Schuell, Inc., Keene, NH) were soaked in 10 mM isopropyl-β-D-thiogalactopyranoside and dried. Filters were overlaid on plaques and β-galactosidase fusion protein synthesis was induced for 2 h at 37°C (Young and Davis, 1985). The filters were keyed to the plates, washed briefly in 0.05 M Tris HCl (pH 8.0), 0.15 M NaCl and 0.2% Tween 20 (TBST), and incubated for 1–2 h in TBST containing 5% bovine serum albumin (BSA). A second filter was overlaid on the plates and incubated at 37°C for 1 h. After incubation in TBST containing 5% BSA, the filters were washed three times with TBST during a 30-min period and were then incubated with TBST containing either MA-I or MA-II (1:250 vol/vol) overnight at 22°C with gentle mixing. The antibody solutions were saved and reused in subsequent screening rounds. The filters were washed with TBST and the antibodies were localized with biotinylated second antibody followed by avidin-biotin-peroxidase complex (Vector Laboratories, Inc., Burlingame, CA). Plaques that produced positive signals in the first high density screen were taken through four to eight successive rounds of antibody screening at progressively lower plaque densities. The resulting, repeatedly positive, and well isolated phage plaques were picked, amplified to yield high titre plate stocks, and used for the large scale preparation of phage (Maniatis et al., 1978). Recombinant phage DNA was isolated from induced lysogens by a previously described procedure (Pirrotta et al., 1971). Lysogens were prepared from each clone by infection of E. coli strain Y1089 at a multiplicity of infection of 5 (Hynes et al., 1986). Fusion proteins were identified by SDS PAGE (Laemmli, 1970). The proteins were electrophoretically transferred to nitrocellulose paper, and the replicas were probed with MA-I and MA-II as described previously (Lawler et al., 1985).

### Subcloning and Sequence Determination

All sequencing was done by the chain termination method of Sanger et al. (1977) with dideoxy sequencing reagents (Promega Biotec, Madison, WI) and standard procedures suggested by the supplier. Reactions were generally performed at 37° or 40°C; however, some regions were also sequenced at 55°C. The subcloning and nucleotide sequencing strategy is summarized in the following steps.

*Step 1.* Purified recombinant phage DNA was digested with KpnI and SacI and the nucleotide sequence of the 5′ end of the insert was determined using a λgtll primer (New England Biolabs).

*Step 2.* Phage DNA was subjected to EcoRI or PstI endonuclease digestion and the inserts or fragments were separated by agarose gel electrophoresis (Maniatis et al., 1982) (Fig. 1, *b* and *c*). DNA was eluted by the glass bead method of Vogelstein and Gillespie (1979) and subcloned into the appropriate sites of the Riboprobe Gemini transcription vector pGEM-2 (Promega Biotec). The nucleotide sequences of the 5′ and 3′ ends of each of the inserts were determined using oligonucleotide primers to the SP6 and T7 promoters of the vector (Promega Biotec).

*Step 3.* Ordered sets of deletion clones of the major PstI fragments were generated by exonuclease III deletion essentially as described by Henikoff (1984) (Fig. 1 *c*). Briefly, 10 μg of purified plasmid DNA was sequentially digested with SacI and BamHI. The sample was extracted with an equal volume of phenol/chloroform (1:1 vol/vol) and then precipitated with ethanol. The pellet was dissolved in 60 μl of 66 mM Tris-HCl (pH 8.0) and 0.66 mM $MgCl_2$, then 500 U of exonuclease III was added. The sample was incubated at 35°C. At 20-s intervals 2.5-μl samples were removed and mixed with 7.5 μl of 0.3 M KOAC (pH 4.6), 0.5 M NaCl, 4.5% glycerol, 16.7 mM, $ZnSO_4$, and 220 U/ml S1 nuclease (Boehringer-Mannheim Biochemicals, Indianapolis, IN) and incubated at 0°C. After all of the time points were collected, the samples were incubated at 22°C for 30 min, 1 μl of 0.3 M Tris and 0.05 M EDTA was added to each, and the samples were heated to 70°C for 10 min. A 2-μl aliquot of each sample was removed for agarose gel electrophoresis. 1 μl of 20 mM Tris-HCl (pH 8.0), 100 mM $MgCl_2$, and 100 U/ml large fragment DNA polymerase I were added and the samples were incubated at 37°C. After 2 min, 1 μl of a mixture containing 0.125 mM dGTP, 0.125 mM dATP, 0.125 mM dTTP, and 0.125 mM dCTP was added and the samples were incubated at 37°C for 5 min. The samples were heated to 70°C for 10 min, a 1/10 vol of 0.66 M Tris-HCl (pH 7.6), 50 mM $MgCl_2$, 50 mM dithiothreitol, and 10 mM ATP was added, and they were incubated with 100–400 U of T4 DNA ligase at 22°C for 18 h.

*Step 4.* Additional PvuII, XmnI, BamHI, SmaI, and AccI fragments were subcloned and sequenced to resolve ambiguities, to confirm the termination codon, and to cross the EcoRI and PstI sites (Fig. 1 *d*).

### Northern Blot Hybridization

Total cellular RNA was isolated as described previously (Schwarzbauer et al., 1983) from cultured human umbilical vein endothelial cells, which were kindly provided by Dr. Robert Weinstein. RNA was separated in 1% agarose-formaldehyde gels and transferred to nitrocellulose paper (0.45 μm; Schleicher & Schuell, Inc.). A hybridization probe was prepared from the M5 insert by EcoRI digestion and preparative agarose gel electrophoresis followed by nick-translation to a specific activity of $10^8$ cpm/μg. Blots were prehybridized at 42°C for 3 h in a solution of 5× Denhardt's, 5× standard saline citrate (SSC), 50% formamide, 100 mM Na phosphate, and 0.1% SDS containing 100 μg/ml E. coli DNA and 5% dextran sulfate. Hybridization was done in the same buffer at 42°C for 12–16 h with 10% dextran sulfate and 0.5–1.0 × $10^7$ cpm/lane of nick-translated probe. Filters were washed at 68°C in several changes of 2× SSC, 0.1% SDS (moderate stringency) followed by 0.1× SSC, 0.1% SDS (high stringency). Molecular

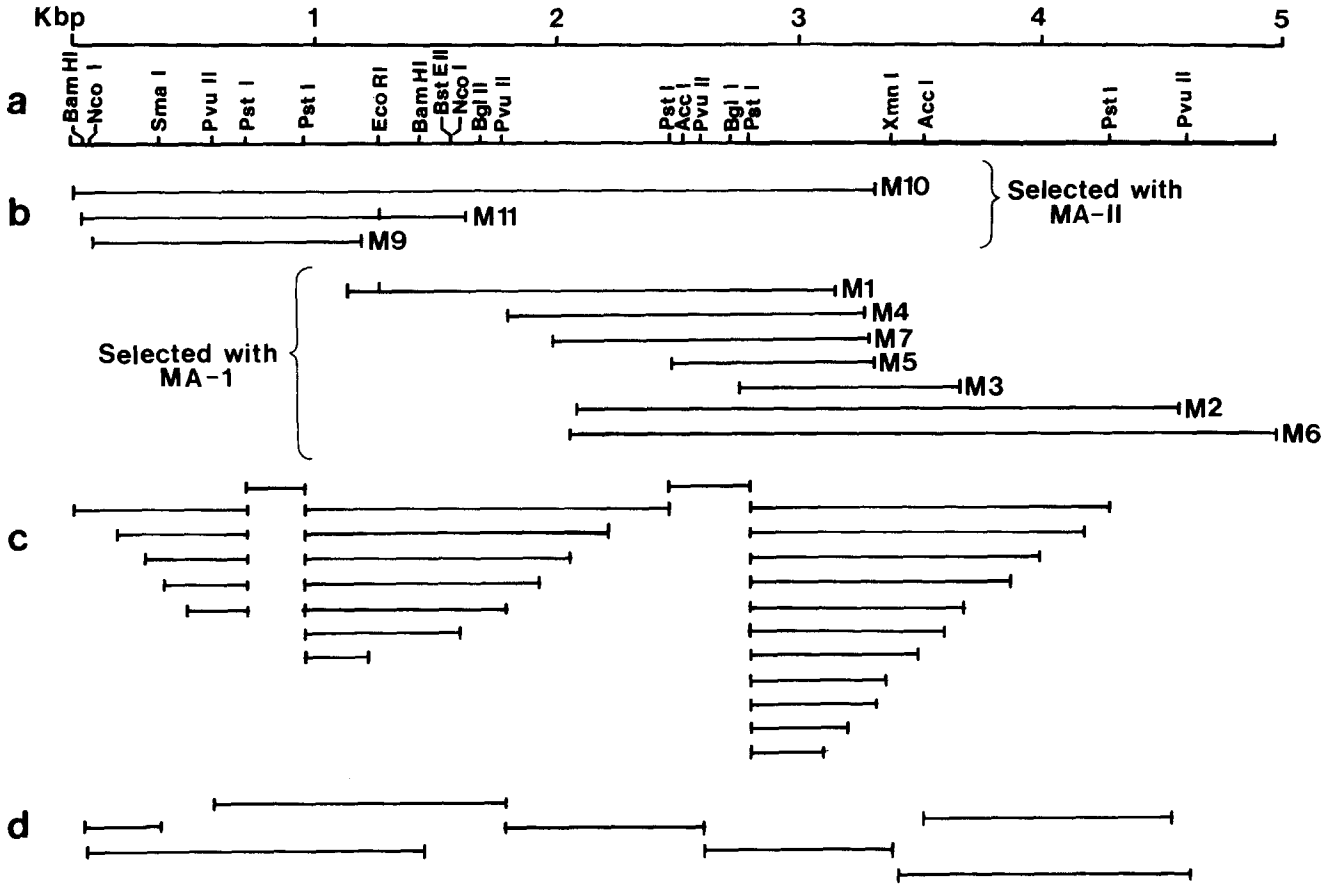



*Figure 1.* Alignment of human thrombospondin clones. (*a*) Restriction map showing endonuclease sites used to align the original cDNA clones and to subclone fragments for sequencing. (*b*) The original cDNA clones selected by MA-I and MA-II from the human endothelial cell cDNA library. (*c*) cDNA subclones produced for nucleotide sequencing by exonuclease III deletion of the major PstI fragments. (*d*) Additional subclones produced by BamHI, SmaI, PvuII, XmnI, and AccI digestion to complete the determination of the nucleotide sequence. Both ends of all clones were sequenced to generate a complete sequence.

weights were determined by comparison with rRNA and RNA size markers (Bethesda Research Laboratories, Gaithersburg, MD).

## Results

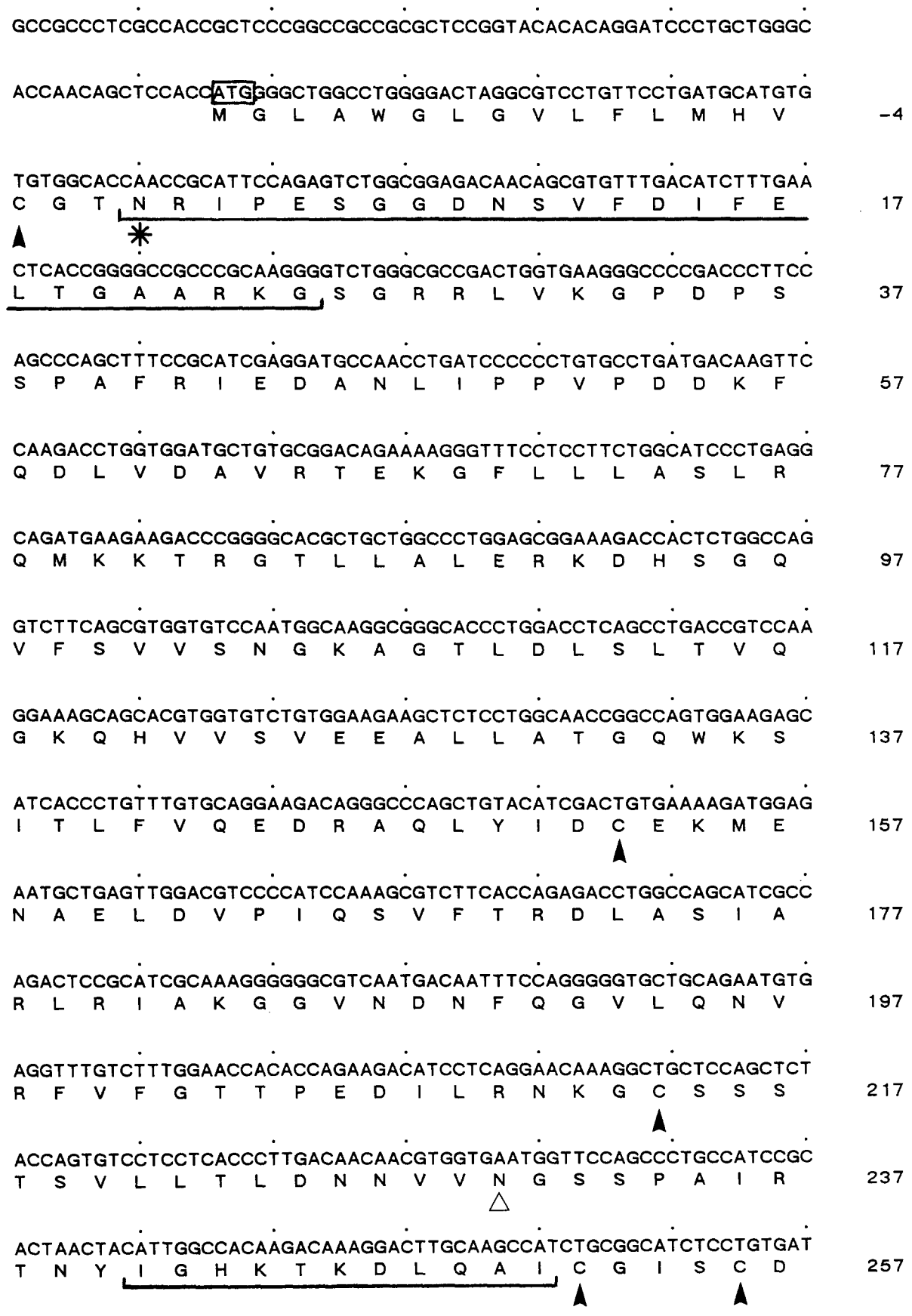
### Isolation of cDNA Clones

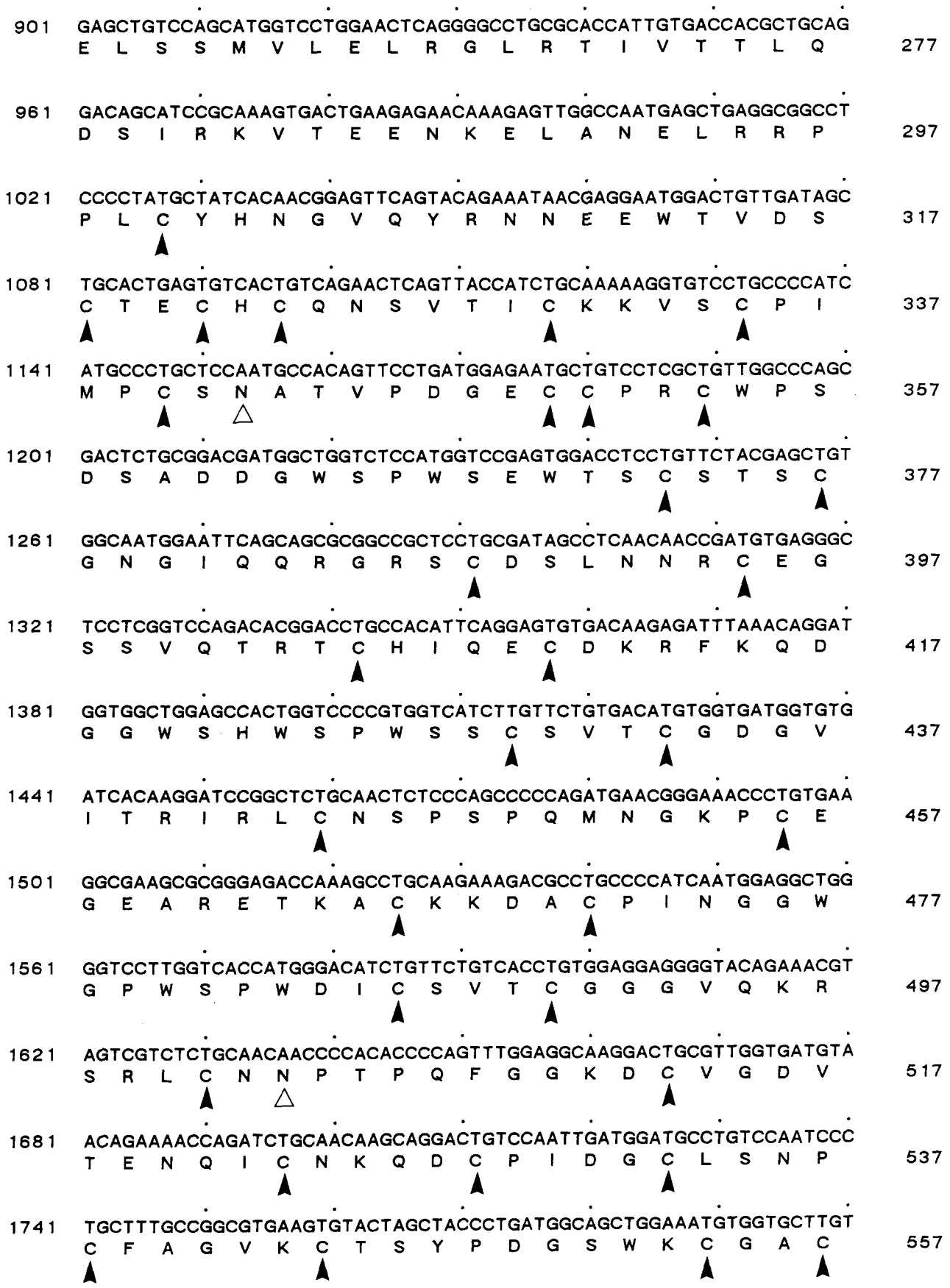
Seven clones, designated M1–M7, were selected and plaque-purified using immunoscreening with the monoclonal antibody MA-I (Fig. 1 *b*). Three clones, designated M9–M11, were selected and plaque-purified with MA-II (Fig. 1 *b*). Inspection of the filters revealed that the plaques produced by the M10 clone stained with MA-I in addition to MA-II, suggesting that this 3.3-kbp insert encoded a fusion protein that contained the epitopes for both antibodies. The cDNA inserts were oriented and aligned based on restriction endonuclease sites and the known positions of the epitopes for the monoclonal antibodies. The antibody results indicate that M9 and M11 must be near the 5′ end of the map, whereas M1–M7 must be near the 3′ end and M10 should overlap both sets of clones. In addition, the nucleotide sequence of the 5′ end of each insert was determined by the chain termination method of Sanger et al. (1977) with a λgt11 primer. These sequence data facilitated the subsequent subcloning and sequence determination by identifying the 5′ end of each insert and the correct reading frame.

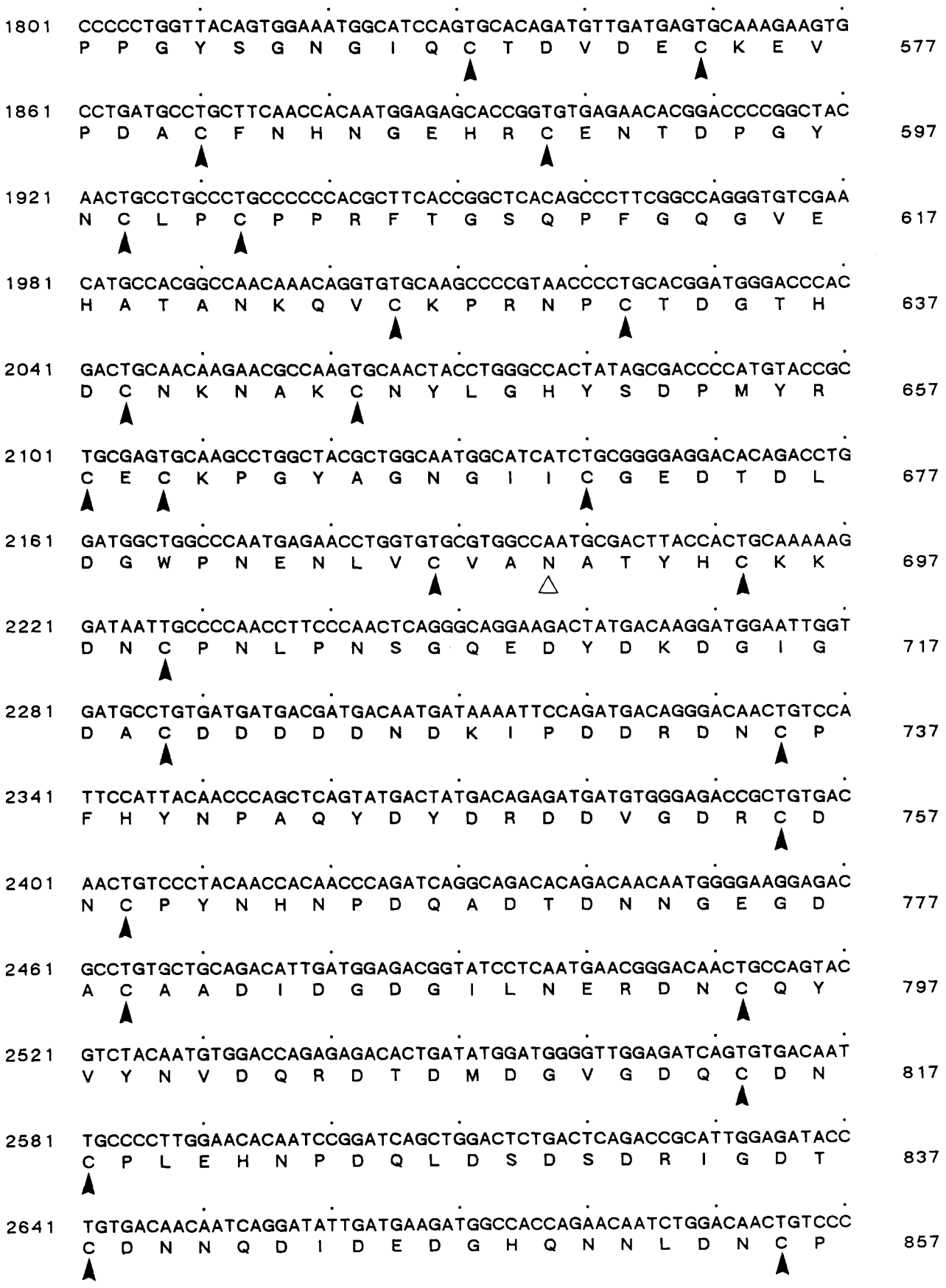
The restriction map indicates that the 10 clones correspond to ~5 kbp of nucleotide sequence. Northern blot analysis of human endothelial cell RNA using clone M5 gave a major band of 6.1 kbp, indicating that the mRNA is ~1.1 kbp larger than the cloned region (data not shown). Minor bands were also observed at 4.6, 4.2, and 3.8 kbp on the Northern blot, however we have not investigated this heterogeneity in detail (data not shown).

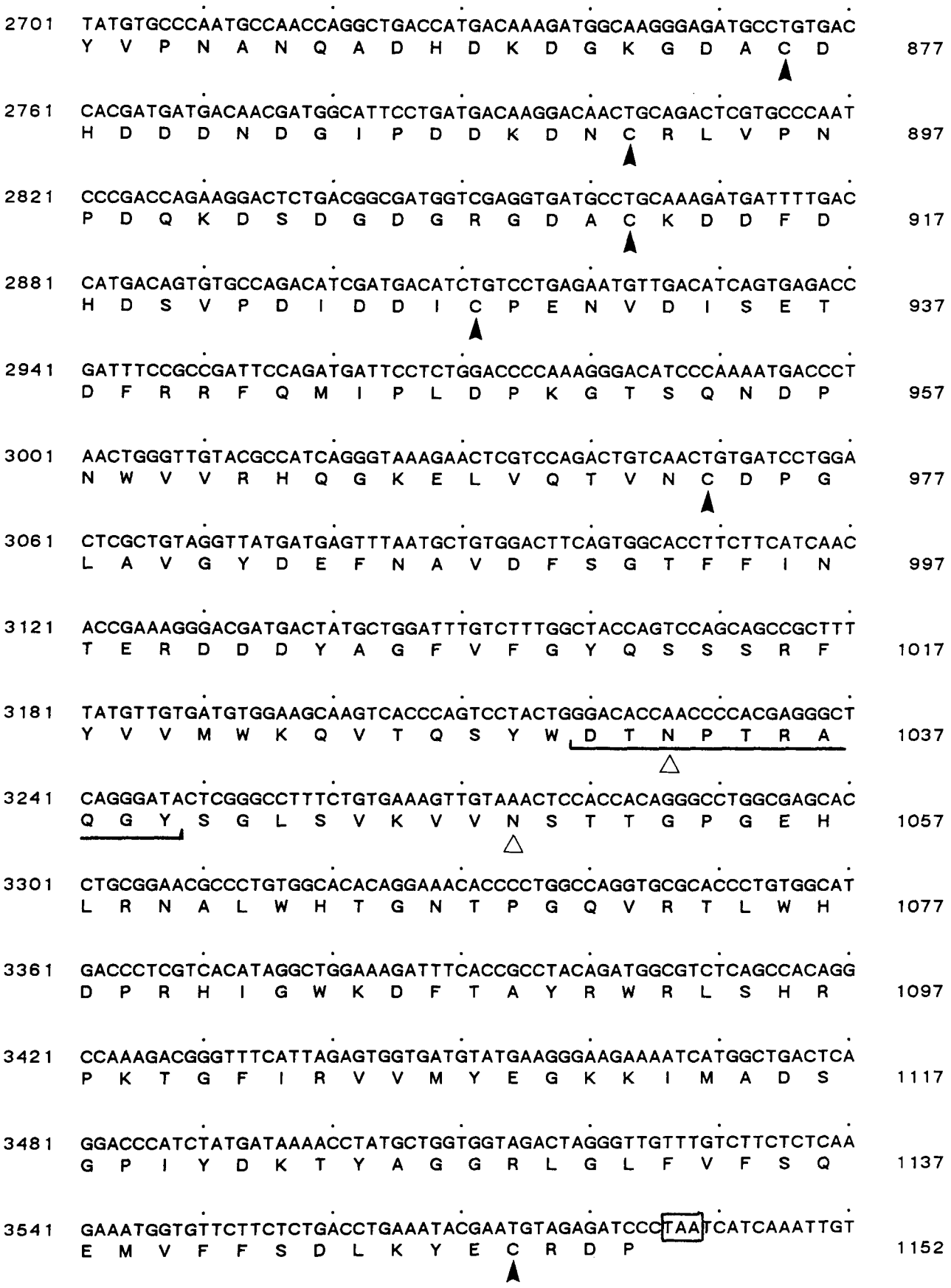
### Determination of the Nucleotide Sequence

Nested sets of overlapping clones of the major PstI fragments were generated by exonuclease III digestion (Fig. 1 *c*). These clones and the original inserts were sequenced by the method of Sanger et al. (1977). The 5′ end of each insert, as well as three regions of known amino acid sequence (see below), provided thirteen checks of the reading frame. The final ambiguities and the position of the termination codon were resolved by subcloning fragments produced by BamHI, SmaI, PvuII, XmnI, and AccI digestion (Fig. 1 *d*). The nucleotide sequence was determined in multiple overlapping clones, and ~70% of the coding region (see below) was de-


1 GCCGCCCTCGCCACCGCTCCCGGCCGCCGCGCTCCGGTACACACAGGATCCCTGCTGGGC
61 ACCAACAGCTCCACCATGGGGCTGGCCTGGGGACTAGGCGTCCTGTTCCTGATGCATGTG
M G L A W G L G V L F L M H V -4
121 TGTGGCACCAACCGCATTCCAGAGTCTGGCGGAGACAACAGCGTGTTTGACATCTTTGAA
C G T N R I P E S G G D N S V F D I F E 17
181 CTCACCGGGGCCGCCCGCAAGGGGTCTGGGCGCCGACTGGTGAAGGGCCCCGACCCTTCC
L T G A A R K G S G R R L V K G P D P S 37
241 AGCCCAGCTTTCCGCATCGAGGATGCCAACCTGATCCCCCCTGTGCCTGATGACAAGTTC
S P A F R I E D A N L I P P V P D D K F 57
301 CAAGACCTGGTGGATGCTGTGCGGACAGAAAAGGGTTTCCTCCTTCTGGCATCCCTGAGG
Q D L V D A V R T E K G F L L L A S L R 77
361 CAGATGAAGAAGACCCGGGGCACGCTGCTGGCCCTGGAGCGGAAAGACCACTCTGGCCAG
Q M K K T R G T L L A L E R K D H S G Q 97
421 GTCTTCAGCGTGGTGTCCAATGGCAAGGCGGGCACCCTGGACCTCAGCCTGACCGTCCAA
V F S V V S N G K A G T L D L S L T V Q 117
481 GGAAAGCAGCACGTGGTGTCTGTGGAAGAAGCTCTCCTGGCAACCGGCCAGTGGAAGAGC
G K Q H V V S V E E A L L A T G Q W K S 137
541 ATCACCCTGTTTGTGCAGGAAGACAGGGCCCAGCTGTACATCGACTGTGAAAAGATGGAG
I T L F V Q E D R A Q L Y I D C E K M E 157
601 AATGCTGAGTTGGACGTCCCCATCCAAAGCGTCTTCACCAGAGACCTGGCCAGCATCGCC
N A E L D V P I Q S V F T R D L A S I A 177
661 AGACTCCGCATCGCAAAGGGGGGCGTCAATGACAATTTCCAGGGGGTGCTGCAGAATGTG
R L R I A K G G V N D N F Q G V L Q N V 197
721 AGGTTTGTCTTTGGAACCACACCAGAAGACATCCTCAGGAACAAAGGCTGCTCCAGCTCT
R F V F G T T P E D I L R N K G C S S S 217
781 ACCAGTGTCCTCCTCACCCTTGACAACAACGTGGTGAATGGTTCCAGCCCTGCCATCCGC
T S V L L T L D N N V V N G S S P A I R 237
841 ACTAACTACATTGGCCACAAGACAAAGGACTTGCAAGCCATCTGCGGCATCTCCTGTGAT
T N Y I G H K T K D L Q A I C G I S C D 257


901 GAGCTGTCCAGCATGGTCCTGGAACTCAGGGGCCTGCGCACCATTGTGACCACGCTGCAG
E L S S M V L E L R G L R T I V T T L Q 277
961 GACAGCATCCGCAAAGTGACTGAAGAGAACAAAGAGTTGGCCAATGAGCTGAGGCGGCCT
D S I R K V T E E N K E L A N E L R R P 297
1021 CCCCTATGCTATCACAACGGAGTTCAGTACAGAAATAACGAGGAATGGACTGTTGATAGC
P L C Y H N G V Q Y R N N E E W T V D S 317
1081 TGCACTGAGTGTCACTGTCAGAACTCAGTTACCATCTGCAAAAAGGTGTCCTGCCCCATC
C T E C H C Q N S V T I C K K V S C P I 337
1141 ATGCCCTGCTCCAATGCCACAGTTCCTGATGGAGAATGCTGTCCTCGCTGTTGGCCCAGC
M P C S N A T V P D G E C C P R C W P S 357
1201 GACTCTGCGGACGATGGCTGGTCTCCATGGTCCGAGTGGACCTCCTGTTCTACGAGCTGT
D S A D D G W S P W S E W T S C S T S C 377
1261 GGCAATGGAATTCAGCAGCGCGGCCGCTCCTGCGATAGCCTCAACAACCGATGTGAGGGC
G N G I Q Q R G R S C D S L N N R C E G 397
1321 TCCTCGGTCCAGACACGGACCTGCCACATTCAGGAGTGTGACAAGAGATTTAAACAGGAT
S S V Q T R T C H I Q E C D K R F K Q D 417
1381 GGTGGCTGGAGCCACTGGTCCCCGTGGTCATCTTGTTCTGTGACATGTGGTGATGGTGTG
G G W S H W S P W S S C S V T C G D G V 437
1441 ATCACAAGGATCCGGCTCTGCAACTCTCCCAGCCCCCAGATGAACGGGAAACCCTGTGAA
I T R I R L C N S P S P Q M N G K P C E 457
1501 GGCGAAGCGCGGGAGACCAAAGCCTGCAAGAAAGACGCCTGCCCCATCAATGGAGGCTGG
G E A R E T K A C K K D A C P I N G G W 477
1561 GGTCCTTGGTCACCATGGGACATCTGTTCTGTCACCTGTGGAGGAGGGGTACAGAAACGT
G P W S P W D I C S V T C G G G V Q K R 497
1621 AGTCGTCTCTGCAACAACCCCACACCCCAGTTTGGAGGCAAGGACTGCGTTGGTGATGTA
S R L C N N P T P Q F G G K D C V G D V 517
1681 ACAGAAAACCAGATCTGCAACAAGCAGGACTGTCCAATTGATGGATGCCTGTCCAATCCC
T E N Q I C N K Q D C P I D G C L S N P 537
1741 TGCTTTGCCGGCGTGAAGTGTACTAGCTACCCTGATGGCAGCTGGAAATGTGGTGCTTGT
C F A G V K C T S Y P D G S W K C G A C 557


1801 CCCCCTGGTTACAGTGGAAATGGCATCCAGTGCACAGATGTTGATGAGTGCAAAGAAGTG
P P G Y S G N G I Q C T D V D E C K E V 577
1861 CCTGATGCCTGCTTCAACCACAATGGAGAGCACCGGTGTGAGAACACGGACCCCGGCTAC
P D A C F N H N G E H R C E N T D P G Y 597
1921 AACTGCCTGCCCTGCCCCCCACGCTTCACCGGCTCACAGCCCTTCGGCCAGGGTGTCGAA
N C L P C P P R F T G S Q P F G Q G V E 617
1981 CATGCCACGGCCAACAAACAGGTGTGCAAGCCCCGTAACCCCTGCACGGATGGGACCCAC
H A T A N K Q V C K P R N P C T D G T H 637
2041 GACTGCAACAAGAACGCCAAGTGCAACTACCTGGGCCACTATAGCGACCCCATGTACCGC
D C N K N A K C N Y L G H Y S D P M Y R 657
2101 TGCGAGTGCAAGCCTGGCTACGCTGGCAATGGCATCATCTGCGGGGAGGACACAGACCTG
C E C K P G Y A G N G I I C G E D T D L 677
2161 GATGGCTGGCCCAATGAGAACCTGGTGTGCGTGGCCAATGCGACTTACCACTGCAAAAAG
D G W P N E N L V C V A N A T Y H C K K 697
2221 GATAATTGCCCCAACCTTCCCAACTCAGGGCAGGAAGACTATGACAAGGATGGAATTGGT
D N C P N L P N S G Q E D Y D K D G I G 717
2281 GATGCCTGTGATGATGACGATGACAATGATAAAATTCCAGATGACAGGGACAACTGTCCA
D A C D D D D D D N D K I P D D R D N C P 737
2341 TTCCATTACAACCCAGCTCAGTATGACTATGACAGAGATGATGTGGGAGACCGCTGTGAC
F H Y N P A Q Y D Y D R D D V G D R C D 757
2401 AACTGTCCCTACAACCACAACCCAGATCAGGCAGACACAGACAACAATGGGGAAGGAGAC
N C P Y N H N P D Q A D T D N N G E G D 777
2461 GCCTGTGCTGCAGACATTGATGGAGACGGTATCCTCAATGAACGGGACAACTGCCAGTAC
A C A A D I D G D G I L N E R D N C Q Y 797
2521 GTCTACAATGTGGACCAGAGAGACACTGATATGGATGGGGTTGGAGATCAGTGTGACAAT
V Y N V D Q R D T D M D G V G D Q C D N 817
2581 TGCCCCTTGGAACACAATCCGGATCAGCTGGACTCTGACTCAGACCGCATTGGAGATACC
C P L E H N P D Q L D S D S D R I G D T 837
2641 TGTGACAACAATCAGGATATTGATGAAGATGGCCACCAGAACAATCTGGACAACTGTCCC
C D N N Q D I D E D G H Q N N L D N C P 857


2701 TATGTGCCCAATGCCAACCAGGCTGACCATGACAAAGATGGCAAGGGAGATGCCTGTGAC
Y V P N A N Q A D H D K D G K G D A C D 877
2761 CACGATGATGACAACGATGGCATTCCTGATGACAAGGACAACTGCAGACTCGTGCCCAAT
H D D D N D G I P D D K D N C R L V P N 897
2821 CCCGACCAGAAGGACTCTGACGGCGATGGTCGAGGTGATGCCTGCAAAGATGATTTTGAC
P D Q K D S D G D G R G D A C K D D F D 917
2881 CATGACAGTGTGCCAGACATCGATGACATCTGTCCTGAGAATGTTGACATCAGTGAGACC
H D S V P D I D D I C P E N V D I S E T 937
2941 GATTTCCGCCGATTCCAGATGATTCCTCTGGACCCCAAAGGGACATCCCAAAATGACCCT
D F R R F Q M I P L D P K G T S Q N D P 957
3001 AACTGGGTTGTACGCCATCAGGGTAAAGAACTCGTCCAGACTGTCAACTGTGATCCTGGA
N W V V R H Q G K E L V Q T V N C D P G 977
3061 CTCGCTGTAGGTTATGATGAGTTTAATGCTGTGGACTTCAGTGGCACCTTCTTCATCAAC
L A V G Y D E F N A V D F S G T F F I N 997
3121 ACCGAAAGGGACGATGACTATGCTGGATTTGTCTTTGGCTACCAGTCCAGCAGCCGCTTT
T E R D D D Y A G F V F G Y Q S S S R F 1017
3181 TATGTTGTGATGTGGAAGCAAGTCACCCAGTCCTACTGGGACACCAACCCCACGAGGGCT
Y V V M W K Q V T Q S Y W D T N P T R A 1037
3241 CAGGGATACTCGGGCCTTTCTGTGAAAGTTGTAAACTCCACCACAGGGCCTGGCGAGCAC
Q G Y S G L S V K V V N S T T G P G E H 1057
3301 CTGCGGAACGCCCTGTGGCACACAGGAAACACCCCTGGCCAGGTGCGCACCCTGTGGCAT
L R N A L W H T G N T P G Q V R T L W H 1077
3361 GACCCTCGTCACATAGGCTGGAAAGATTTCACCGCCTACAGATGGCGTCTCAGCCACAGG
D P R H I G W K D F T A Y R W R L S H R 1097
3421 CCAAAGACGGGTTTCATTAGAGTGGTGATGTATGAAGGGAAGAAAATCATGGCTGACTCA
P K T G F I R V V M Y E G K K I M A D S 1117
3481 GGACCCATCTATGATAAAACCTATGCTGGTGGTAGACTAGGGTTGTTTGTCTTCTCTCAA
G P I Y D K T Y A G G R L G L F V F S Q 1137
3541 GAAATGGTGTTCTTCTCTGACCTGAAATACGAATGTAGAGATCCCTAATCATCAAATTGT
E M V F F S D L K Y E C R D P 1152

```
3601  TGATTGAAAGACTGATCATAAACCAATGCTGGTATTGCACCTTCTGGAACTATGGGCTTG

3661  AGAAAACCCCCAGGATCACTTCTCCTTGGCTTCCTTCTTTTCTGTGCTTGCATCAGTGTG

3721  GACTCCTAGAACGTGCGACCTGCCTCAAGAAAATGCAGTTTTCAAAAACAGACTCAGCAT

3781  TCAGCCTCCAATGAATAAGACATCTTCCAAGCATATAAACAATTGCTTTGGTTTCCTTTT

3841  GAAAAAGCATCTACTTGCTTCAGTTGGGAAGGTGCCCATTCCACTCTGCCTTTGTCACAG

3901  AGCAGGGTGCTATTGTGAGGCCATCTCTGAGCAGTGGACTCAAAAGCATTTTCAGGCATG

3961  TCAGAGAAGGGAGGACTCACTAGAATTAGCAAACAAAACCACCCTGACATCCTCCTTCAG

4021  GAACACGGGGAGCAGAGGCCAAAGCACTAAGGGGAGGGCGCATACCCGAGACGATTGTAT

4081  GAAGAAAATATGGAGGAACTGTTACATGTTCGGTACTAAGTCATTTTCAGGGGATTGAAA

4141  GACTATTGCTGGATTTCATGATGCTGACTGGCGTTAGCTGATTAACCCATGTAAATAGGC

4201  ACTTAAATAGAAGCAGGAAAGGGAGACAAAGACTGGCTTCTGGACTTCCTCCCTGATCCC

4261  CACCCTTACTCATCACCTGCAGTGGCCAGAATTAGGGAATCAGAATCGAAACCAGTGTAA

4321  GGCAGTGCTGGCTGCCATTGCCTGGTCACATTGAAATTGGTGGCTTCATTCTAGATGTAG

4381  CTTGTGCAGATGTAGCAGGAAAATAGGAAAACCTACCATCTCAGTGAGCACCAG
```

*Figure 2.* Nucleotide sequence of cDNA clones and corresponding amino acid sequence. The initiation and termination codons are boxed. The $NH_2$ terminal (*asterisks*) and cysteine residues (*solid arrowheads*) are indicated and the regions of known amino acid sequence are underlined. The potential sites for addition of N-linked carbohydrate are marked by open triangles. The amino acid sequence is given in the single letter code.

termined on both strands. The results established a continuous sequence of 4,434 bases from the 5′ end of M10 to the PvuII site just beyond the 3′ end of M2 (Fig. 2). The region between the 3′ end of M2 and that of M6 was not completely sequenced because this region was determined to be noncoding (see below).

The cDNA sequence has the following features. A consensus sequence (ACCATGG) for initiation codons begins at position 73 (Fig. 2) (Kozak, 1986). The initiation codon is followed by an open reading frame of 3,508 bases and a TAA termination codon at position 3,586. The coding region is followed by 848 bases of 3′ untranslated sequence. An additional 140 bases have been determined by sequencing the 3′ end of M6 (Fig. 1). Both segments of 3′ untranslated sequence have multiple termination codons in all three frames, but do not include a polyadenylation signal sequence. These data are consistent with the Northern blotting data which indicate that the thrombospondin message is 6.1 kb, presumably including more 3′ untranslated sequences that are not included in our clones.

### *Amino Acid Sequence*

The cDNA sequence predicts that the mature thrombospondin chains (i.e., minus the signal sequence, see below) are composed of 1,152 amino acids and have a molecular weight of 127,524 without carbohydrate (Fig. 2). The amino acid composition of the predicted polypeptide (K,55; H,26; R,60; D,131; N,81; E,51; Q,50; T,61; S,74; G,101; A,50; C,69; M,12; V,70; P,70; I,45; L,60; F,34; Y,31; W,21) agrees well with previously published compositions for thrombospondin (Lawler et al., 1978).

Three regions of known amino acid sequence of human

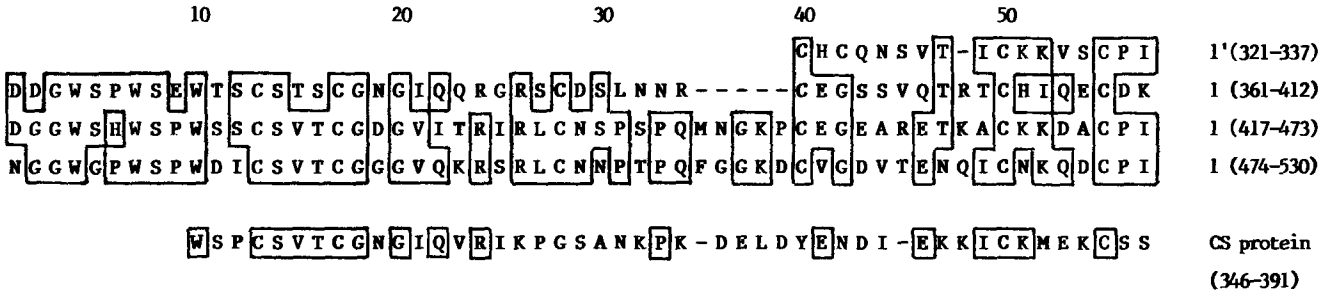



*Figure 3.* Alignment of type 1 homologies. The four type 1 homologies (three complete and one partial) are aligned with the position of the first and last amino acids in the complete thrombospondin sequence indicated to the right of each line. Regions where two or more of the four aligned residues are identical are enclosed in boxes. Dashes indicate gaps introduced to maximize alignment. The amino acid sequence of a homologous region from the circumsporozoite (CS) protein of *Plasmodium falciparum* is given at the bottom with homologous residues enclosed in boxes (Dame et al., 1984). Note that the CS protein sequence lacks the central pair of cysteine residues present in the thrombospondin repeats.

platelet thrombospondin can be identified. The amino acid sequence between N(1) and G(25) is identical with the reported sequence for the $NH_2$-terminal of the thrombospondin chains, with one exception (Coligan and Slayter, 1984; Dixit et al., 1984; Raugi et al., 1984). Arginine (23) has previously been reported to be tryptophan (Coligan and Slayter, 1984). The amino acid sequence between I(241) and I(251) is identical with the reported sequence for the $NH_2$ terminal of a 70,000-D chymotryptic fragment of thrombospondin (Galvin et al., 1985). In addition, the amino acid sequence between D(1031) and Y(1046) is identical with the reported sequence for the $NH_2$ terminal of an 18,000-D chymotryptic fragment of thrombospondin (Galvin et al., 1985). These data prove that the M1–M7 and M9–M11 clones are authentic clones of human thrombospondin.

The predicted amino acid sequence of thrombospondin has the following features. The $NH_2$ terminal N(1) is preceeded by an 18 amino acid signal sequence of uncharged residues (−18 to −1). The majority of the cysteine residues are located in the center one-third of the molecule. Three types of repeating sequence, designated homology types 1–3, occur within the amino acid sequence (see below). There are six potential sites for N-linked glycosylation, although two of them include proline and are less likely to be used (Fig. 2) (Hubbard and Ivatt, 1981).

The amino acid sequences of the type 1 homologies are shown in Fig. 3. Three complete type 1 homologies of 57 amino acids occur between D(361) and I(530). Alignment of these three segments reveals that 30% of the residues are identical in all three and that the positions of all six cysteine residues are conserved (Fig. 3). In addition, the region between C(321) and I(337) was found to be homologous with the last 17 amino acids of the type 1 repeating sequences (Fig. 3).

Three adjacent type 2 homologies follow immediately after the type 1 homologies (Fig. 4). Six cysteine residues occur in each of the three repeats along with other conserved residues (Fig. 4). The type 2 repeats are not as well conserved as the type 1 homologies. The second type 2 repeat shows the least homology with the other two (Fig. 4). A sequence of 13 amino acids (positions 42–54) is present in this repeat only. Ignoring this insertion and with other gaps as in Fig. 4, the repeats show 30–35% pairwise identity (20% threefold identity) over 46 residues.

The amino acid sequences of the type 3 homology region are shown in Fig. 5. This region includes eight repeating sequences that show a well conserved pattern 38 amino acids long. The eight type 3 homologies form a continuous sequence of 260 amino acids. A consensus sequence for the type 3 repeats shows that aspartate (D) occupies 11 of the 38 positions (Fig. 5). In addition, the spacing of the D residues and one of the glycine (G) residues in the first half of the consensus sequence (positions 6–17) is recapitulated in the second half of the consensus sequence (positions 21–32). These sequences are homologous with the calcium-binding sites of parvalbumin and calmodulin (see Discussion).

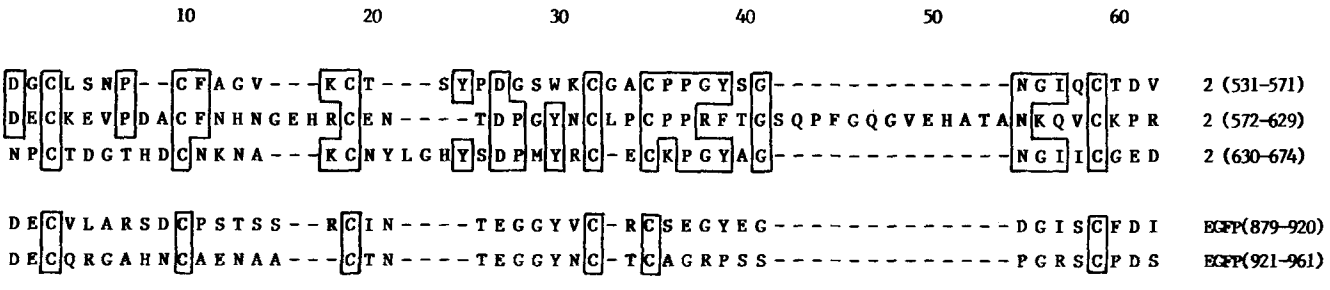



*Figure 4.* Alignment of type 2 homologies. The three type 2 homologies are aligned with the positions of the first and last amino acids in the complete amino acid sequence of thrombospondin indicated at the right. The sequence is continuous from D(531) to D(674). The amino acid sequence of the homologous region of mouse epidermal growth factor precursor (EGFP) is given at the bottom with the characteristic cysteine residues enclosed in boxes (Gray et al., 1983; Scott et al., 1983).

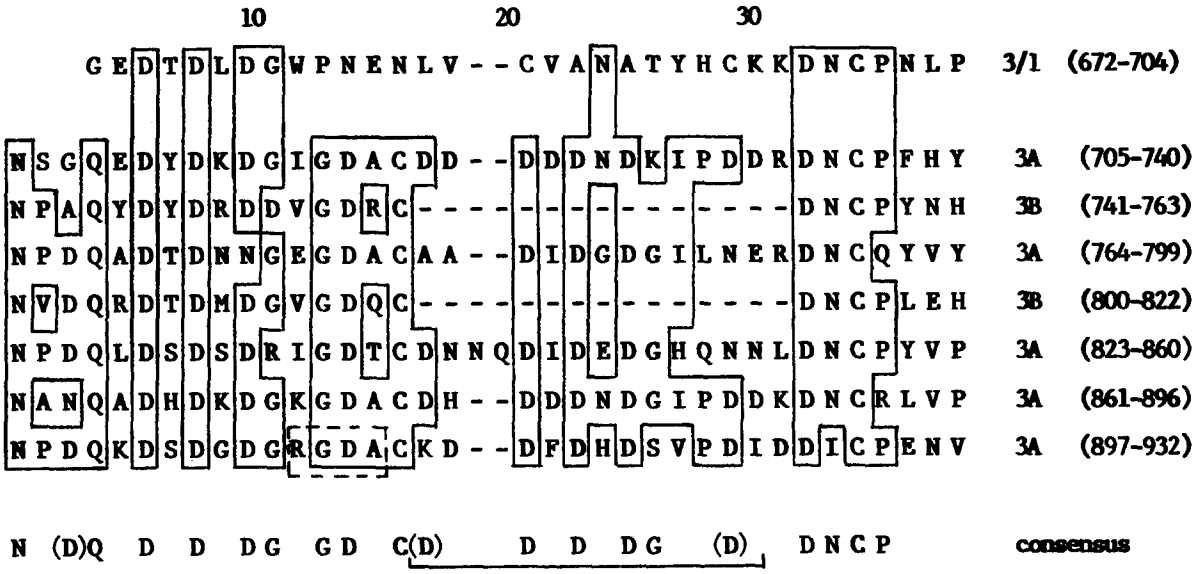



*Figure 5.* Alignment of type 3 homologies. The eight type 3 homologies are aligned with the positions of the first and last amino acids in the complete amino acid sequence of thrombospondin indicated at the right. The sequence given is continuous from G(672) to V(932). This sequence follows directly after the type 2 homologies shown in Fig. 4. A consensus sequence of the most strongly conserved amino acids is indicated at the bottom. Note that the pattern of aspartate (D) residues, which probably constitutes a calcium-binding site, is repeated twice in the type 3A repeats. The second half of each of the type 3A repeats, which is absent in the two type 3B repeats is bracketed. The first repeat (3/1) appears to be a hybrid of type 3 and type 1 homologies (see text). The RGDA sequence, a potential cell-binding site, is enclosed in the dashed box. Dashes indicate where the sequences have been gapped to maximize alignment. Regions where four or more of the aligned residues are identical are enclosed in boxes.

The type 3 repeating sequences can be subdivided into three subtypes. The first subtype, designated 3A, includes five of the eight repeating sequences (Fig. 5). The type 3A repeats conform to the consensus sequence and include both putative calcium-binding sites. The second subtype, designated 3B, occurs twice (Fig. 5). Positions 17–31, constituting one of the two calcium binding sites, are absent from the type 3B homologies. The third subtype, designated 3/1, shows the least homology with types 3A and 3B (Fig. 5). Positions 6, 8, and 10 are occupied by D and position 11 is occupied by G as in the consensus sequence. However, the latter half (positions 21–38) of this repeat is dissimilar to the consensus sequence (Fig. 5). The latter half of the 3/1 repeat is instead homologous with the last 17 residues of the type 1 repeats (Fig. 3). This repeat, therefore, appears to be a hybrid of type 3 and type 1 homologies.

## Discussion

The data presented here indicate that each of the three chains of thrombospondin has a molecular weight of 127,524 without carbohydrate. A mass of 10,000 D has been reported for the carbohydrate groups, bringing the total mass of each chain to ~138,000 D (Vischer et al., 1985). This value agrees well with molecular weights of 133,000 determined by sedimentation equilibrium and 145,000 determined by SDS PAGE on phosphate-buffered gels (Margossian et al., 1981; Lawler et al., 1982). These data indicate that the value of 185,000 obtained by SDS PAGE by the method of Laemmli (1970) is an overestimate (Lawler et al., 1982; and see Fig. 6).

The amino acid sequence obtained in this study for endothelial cell thrombospondin is consistent with earlier structural studies of platelet thrombospondin. The amino acid sequence predicts that there is a peptide of 26,173 D, with one glycosylation site, between the $NH_2$ terminus of the intact molecule and the $NH_2$ terminus of the 70,000-D chymotryptic fragment (Fig. 6) (Galvin et al., 1985). The glycosylation site is eleven residues from the COOH-terminus of this fragment. Chymotryptic digestion at low enzyme-to-substrate ratios results in the production of a 30,000-D fragment which binds the Lens culinaris lectin (Lawler et al., 1986*a*). At higher enzyme-to-substrate ratios the 30,000-D fragment is converted to a 25,000-D fragment which no longer binds the lectin. Since the $NH_2$-terminal sequence of the 25,000-D fragment is identical to that of the intact molecule, it can be concluded that the carbohydrate is near the COOH terminal of the 30,000-D fragment (Galvin et al., 1985; Lawler et al., 1986*a*) exactly as predicted by the sequence.

The 25,000-D fragment also contains the epitope for MA-II (Lawler et al., 1985), consistent with the fact that this region is encoded by the clones M9 and M11, which were selected by MA-II (Figs. 1 and 6). The 25,000-D fragment also contains the binding site for heparin (Lawler and Slayter, 1981; Dixit et al., 1984). Two possible heparin-binding sites can be identified in this region. Clusters of positively charged amino acids exist between R(23) and K(32) and between R(77) and R(83). The latter segment is predicted to be in an α-helical conformation by the algorithm of Chou and Fasman (1978), as is the heparin-binding region of platelet factor 4 and β-thromboglobulin (Lawler, 1981).

The 70,000-D chymotryptic fragment (see Fig. 6) contains the sites where the chains are cross-linked by disulfide bonds and binds Lens culinaris lectin (Lawler et al., 1985; Galvin et al., 1985; Lawler et al., 1986). Consistent with these observations, the sequence that follows the $NH_2$ terminal of the 70,000-D fragment is rich in cysteine and includes three

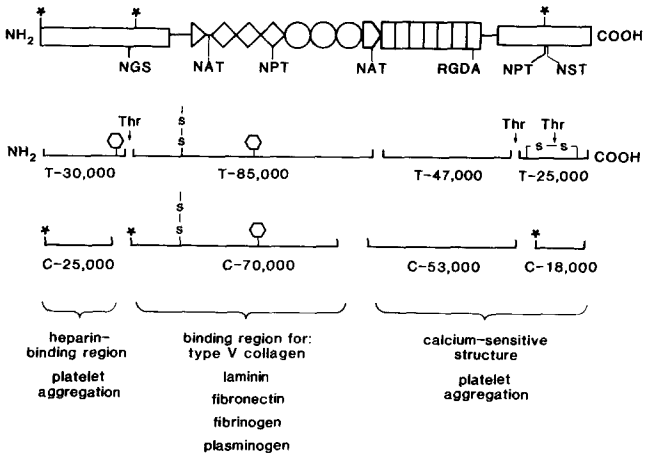



*Figure 6.* Correlation of the primary sequence data with the properties of the major proteolytic fragments of thrombospondin. The position of the types 1 (*open diamond*), 2 (*open circle*), and 3 (*open rectangle*) repeats are indicated at the top. The open boxes at the $NH_2$ terminal and COOH terminal are regions where strong repeating sequences could not be identified. The locations of known amino acid sequence are indicated (*asterisk*). NGS, NAT, NPT, and NST mark sequences of potential asparagine linked carbohydrate acceptor sites (NxT and NxS). RGDA (arg-gly-asp-ala) marks a potential cell-binding site. The masses of the fragments produced by trypsin (T) and chymotrypsin (C) are indicated underneath each fragment. The sites for thrombin (thr) cleavage are indicated by arrows. Those fragments which contain carbohydrate moieties (*open hexagon*) which bind Lens culinaris lectin are indicated. The position and number of Lens culinaris lectin binding sites within the 85,000-D tryptic fragment (T-85,000) has not been determined. The molecular weights given in this figure are based on the migration in the Laemmli gel system which overestimates the size of thrombospondin and several of its fragments. However, we use them here for ease of comparison with earlier data.

potential sites for N-linked glycosylation (Fig. 6). A computer search of the National Biomedical Research Foundation and Newat (Doolittle, 1981) data bases revealed that this area (residues 285–354) also shows homology with the $NH_2$-terminal prosegment of human (36.6% identity in a 71 amino acid overlap) and bovine (37.5% identity in a 72 amino acid overlap) type I procollagen, respectively (Horlein et al., 1979; Chu et al., 1984). Interestingly, short sequences in this area also show homology with human fibronectin (50% identity in a 16 amino acid overlap, thrombospondin residues 298–313) and human von Willebrand factor (53.8% identity in a 13 amino acid overlap, thrombospondin residues 348–360) (Garcia-Pardo et al., 1983; Kornblihtt et al., 1985; Ginsburg et al., 1985). It is possible that these short sequences are involved in a common aspect of matrix protein function in much the same way the RGD sequence is involved in cell adhesion (see below).

The 85,000-D tryptic fragment, which overlaps the 70,000-D chymotryptic fragment (Fig. 6), has been reported to contain a cross-linking site for factor $XIII_a$ (Bale et al., 1985). The sequence Gly-Gln-Gln (GQQ) has been shown to be the transglutaminase reactive site in the γ-chain of fibrinogen (Chen and Doolittle, 1971), and Ala-Gln-Gln (AQQ) is the major reactive site in fibronectin (McDonough et al., 1981). Thrombospondin contains the sequence I(381)-Q(382)-Q(383) in the region, which should be included in the 85,000-D tryptic fragment.

The precise COOH terminal of the 70,000-D chymotryptic fragment is not known. The molecular weight would predict that all of the type 1 and all or most of the type 2 repeating sequences are located in this fragment (Fig. 6). A computer search (Lipman and Pearson, 1985) of known protein sequences revealed that the type 1 repeating sequences are homologous with the COOH terminal of circumsporozoite proteins from *Plasmodium falciparum* (38.3% identify in a 47 amino acid overlap) and *Plasmodium knowlesi* (31.4% identify in a 51 amino acid overlap) (Ozaki et al., 1983; Dame et al., 1984). The significance of this finding is not clear. Thrombospondin has been reported to be involved in the adhesion of *Plasmodium falciparum*–infected red blood cells to endothelial cells (Roberts et al., 1985). However, this phenomenon involves the trophozoite or schizont stage of the parasite, whereas the sporozoite stage infects hepatocytes. The cysteine-rich segment of the circumsporozoite protein is conserved between isolates and species of the parasite and has been proposed to be involved in interactions with cell surfaces (Dame et al., 1984). The region of thrombospondin that contains the three type 1 repeats, which are homologous with this segment of the circumsporozoite protein, is known to bind to several matrix proteins (Fig. 6) (Mumby et al., 1984; Lawler et al., 1986*b*). The precise function of this motif in both *Plasmodium* proteins and thrombospondin clearly deserves further investigation.

The type 2 repeats are somewhat homologous with mouse epidermal growth factor precursor (24% homology in a 200 amino acid overlap by the program of Lipman and Pearson, 1985; see also Gray et al., 1983; Scott et al., 1983) (Fig. 4). Transforming growth factors, vaccinia virus growth factor, coagulation factors IX and X, tissue plasminogen activator, urokinase, the low density lipoprotein receptor, and the proteins encoded by notch and lin-12 genes also have a region of epidermal growth factor (EGF)–like homology, based primarily on the positions of the six cysteine residues (for a review see Bender, 1985). As shown in Fig. 4, the central region of thrombospondin, which is homologous with epidermal growth factor precursor, can be arranged as three type 2 repeats, which show some homology with epidermal growth factor–like repeats, especially in the arrangement of cysteine residues. The significance of the homologies among these diverse proteins is uncertain. One possibility is that the epidermal growth factor–like module is involved in protein–protein interactions in each case. The 70,000-D chymotryptic fragment of thrombospondin binds to a variety of other proteins (Fig. 6) (Mumby et al., 1984; Lawler et al., 1986). Some of these affinities are shared with other proteins in the family of proteins which contain some homologies with epidermal growth factor. For example, tissue plasminogen activator and urokinase both interact with plasminogen, as does thrombospondin, and tissue plasminogen activator and thrombospondin both bind fibronectin and fibrinogen (Leung and Nachman, 1982; Lahav et al., 1984; Silverstein et al., 1984). Several lines of investigation are suggested by these homologies and parallels.

The type 3 homologies are typified by the consensus sequence shown in Fig. 5. The primary structure of this region suggests that it consists of a series of calcium-binding sites immediately adjacent to each other. The amino acid se-

quences of the type 3 repeats are similar to the sequences of the calcium-binding sites of calmodulin, parvalbumin, and the fibrinogen β- and γ-subunits (Kretsinger, 1980; Henschen et al., 1983; Dang et al., 1985). The aspartic acid residues at positions 6, 8, 10, 14, and 17 contribute oxygens for calcium binding. In the type 3A repeating sequence a second set of identically spaced aspartic acid residues is present at positions 21, 23, 25, 29, and 32. The glycine residues in positions 11 and 26 are also homologous with the calcium-binding sites of calmodulin and parvalbumin (Kretsinger, 1980). The deletion in the type 3B repeats preserves one copy of the aspartate motif, with D(32) replacing D(17) (Fig. 5). In contrast to calmodulin and parvalbumin, thrombospondin has aspartic acid residues in positions 17 and 32 instead of glutamic acid. The calcium-binding site of the β-chain of fibrinogen also has aspartic acid in this position (Henschen et al., 1983). While the immediate coordinates of the calcium-binding sites are likely to be similar to those in calmodulin, the overall molecular architecture would seem to be quite different. Since the sites are so close together in thrombospondin, the α-helical segments which form the EF-hand motif are not present (Kretsinger, 1980). The presence of the two cysteine residues in each type 3 homology suggests that the structure in thrombospondin is stabilized by disulfide bonds rather than by secondary structure. These differences may account for the fact that, although some high affinity interactions have been reported, for the most part, thrombospondin binds to calcium with a much lower affinity ($\sim 10^{-4}$) than does calmodulin ($\sim 10^{-6}$) (Klee et al., 1980; Lawler and Simons, 1983; Dixit et al., 1986). Both the disulfide-bonded structure and lower affinity for calcium are concordant with the fact that thrombospondin functions in an extracellular environment. The presence of two cysteine residues in each of the type 3A and 3B repeats is consistent with recent reports that thrombospondin contains an intrachain disulfide bond that is protected from reducing agents by calcium (Turk and Detwiler, 1986). These authors also describe a reactive thiol group that is protected by calcium. The hybrid 3/1 repeat contains an uneven number of cysteine residues, one of which may be the relevant sulfhydryl group.

The location of the calcium-binding sites, based on the sequence data presented here, is consistent with the biochemical and electron microscopic data (see Fig. 6 and the introduction). The production of the 47,000- and 53,000-D fragments by trypsin and chymotrypsin, respectively, is dependent upon the removal of calcium from the molecule (Lawler et al., 1985; Lawler et al., 1986 *a*, *b*). Dixit and coworkers (1986) have described two monoclonal antibodies that have higher affinity for thrombospondin after EDTA treatment. The binding of one of these antibodies, designated A6.1, to thrombospondin has a calcium ion concentration dependence similar to the structural parameters (50–120 μM). The binding of the other antibody, designated D4.6, has a sharp transition at a calcium ion concentration of 100 nM. Epitope mapping indicates that monoclonal antibodies A6.1 and D4.6 bind to the region of polypeptide that produces the 47,000-D tryptic fragment (Dixit et al., 1986). The structure of this region is profoundly affected by the removal of calcium (Lawler et al., 1985; Dixit et al., 1986). In the presence of calcium this region is part of globular region C. When calcium is removed from the molecule this region adopts an extended conformation.

The RGD sequence, which is present in the last type 3 homology of thrombospondin, mediates the interaction of several extracellular matrix proteins with cell surfaces (for a review see Ruoslahti and Pierschbacher, 1986). The glycoprotein IIb/IIIa complex of platelets has been identified as a receptor that recognizes the RGD sequence of fibronectin, fibrinogen, von Willebrand factor, and vitronectin (Gardner and Hynes, 1985; Pytela et al., 1986). Parallel studies with thrombospondin indicated that this receptor may also bind thrombospondin weakly (Pytela et al., 1986). Thrombospondin has been reported to associate with the surface of thrombin-treated normal platelets in the presence of EDTA, which inactivates IIb/IIIa, and with the surface of thrombin-treated thrombasthenic platelets, which lack IIb/IIIa (Hourdille et al., 1985). By contrast, monoclonal antibodies to the glycoprotein IIb/IIIa complex inhibit the binding of thrombospondin to platelets (Plow et al., 1985). Wolff et al. (1986) have recently described the results of studies designed to measure the binding of thrombospondin to platelets. They detected two classes of binding sites; one class on resting platelets and another that is expressed on thrombin-treated platelets in the presence of calcium. One of these classes may represent a specific receptor for thrombospondin, which recognizes the RGDA sequence. Since this sequence is in a calcium-sensitive domain of thrombospondin, it is possible that its function could be regulated by divalent cation effects.

After the last type 3A repeating sequence there is a region of 25,000 D in which a well-conserved repeating sequence could not be identified. A region of known amino acid sequence is found in this area (Galvin et al., 1985). This sequence is the $NH_2$-terminal sequence of an 18,000-D chymotryptic fragment. The amino acid sequence indicates that the peptide portion of this fragment has a molecular weight of 13,978 and that there are two potential sites for N-linked glycosylation, which probably accounts for the observed molecular weight (Fig. 6). The 18,000-D fragment is known to be disulfide bonded to the rest of the molecule (Fig. 6) (Galvin et al., 1985), suggesting that C(1149) is linked to C(974).

In conclusion, the sequence data presented here provide structural information that is consonant with all of the prior structural and functional data on thrombospondin and offer new insights and lines of experimental investigation into the function of the molecule.

We wish to thank Robert Handin for providing the λgtll library of cDNA derived from human endothelial cells and bacterial strains. Robert Weinstein and Karen Wenc kindly provided the cultures of human endothelial cells. We wish to thank Steven Henikoff for his protocol for exonuclease III digestion. We are grateful to Jean Schwarzbauer, Douglas DeSimone, Mary Ann Stepp, Ramila Patel, and Deborah Fonda for constant help and support. Expert technical assistance was provided by Paula Ferro and Joseph Connolly. The art work was done by Joan Joos, and the manuscript was edited by Sami Lawler and typed by Noha Maalouf.

This work was supported by National Institutes of Health Research Award HL-28749 (to J. Lawler) and CA-26712 (to R. O. Hynes), and by a grant from the Whitaker Health Sciences Fund.

Received for publication 28 July 1986, and in revised form 31 July 1986.

## References


Bale, M. D., L. G. Westrick, and D. F. Mosher. 1985. Incorporation of thrombospondin into fibrin clots. *J. Biol. Chem.* 260:7502–7508.

Bender, W. 1985. Homeotic gene products as growth factors. *Cell.* 43:559–560.

Chou, P. Y., and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol.* 47:45–148.

Chen, R., and R. F. Doolittle. 1971. γ-γ Cross-linking sites in human and bovine fibrin. *Biochemistry.* 78:4486–4491.

Chu, M. L., W. DeWet, M. Bernard, J-F. Ding, M. Morabito, J. Myers, C. Williams, and F. Ramirez. 1984. Human proα1(I) collagen gene structure reveals evolutionary conservation of a pattern of introns and exons. *Nature (Lond.).* 310:337–340.

Coligan, J. E., and H. S. Slayter. 1984. Structure of thrombospondin. *J. Biol. Chem.* 259:3944–3948.

Dame, J. B., J. L. Williams, T. F. McCutchan, J. L. Weber, R. A. Wirtz, W. T. Hockmeyer, W. L. Maloy, J. D. Haynes, I. Schneider, D. Roberts, G. S. Sanders, E. P. Reddy, C. L. Diggs, and L. H. Miller. 1984. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite Plasmodium falciparum. *Science (Wash. DC).* 225: 593–599.

Dang, C. V., R. F. Ebert, and W. R. Bell. 1985. Localization of a fibrinogen calcium binding site between γ-subunit positions 311 and 336 by terbium fluorescence. *J. Biol. Chem.* 260:9713–9719.

Dixit, V. M., G. A. Grant, S. A. Santoro, and W. A. Frazier. 1984. Isolation and characterization of a heparin-binding domain from the amino terminus of platelet thrombospondin. *J. Biol. Chem.* 259:10100–10105.

Dixit, V. M., D. M. Haverstick, K. M. O'Rourke, S. W. Hennessy, G. A. Grant, S. A. Santoro, and W. A. Frazier. 1985. A monoclonal antibody against human thrombospondin inhibits platelet aggregation. *Proc. Natl. Acad. Sci. USA.* 82:3472–3476.

Dixit, V. M., N. J. Galvin, K. M. O'Rourke, and W. A. Frazier. 1986. Monoclonal antibodies that recognize calcium-dependent structures of human thrombospondin. *J. Biol. Chem.* 261:1962–1968.

Doolittle, R. F. 1981. Similar amino acid sequences: chance or common ancestry. *Science (Wash. DC).* 214:149–159.

Galvin, N. J., V. M. Dixit, K. M. O'Rourke, S. A. Santoro, G. A. Grant, and W. A. Frazier. 1985. Mapping of epitopes for monoclonal antibodies against human platelet thrombospondin with electron microscopy and high sensitivity amino acid sequencing. *J. Cell Biol.* 101:1434–1441.

Garcia-Pardo, A., E. Pearlstein, and B. Frangione. 1983. Primary structure of human plasma fibronectin. The 29,000-dalton $NH_2$-terminal domain. *J. Biol. Chem.* 258:12670:12674.

Gardner, J. M., and R. O. Hynes. 1985. Interaction of fibronectin with its receptor on platelets. *Cell.* 42:439–448.

Gartner, T. K., D. A. Walz, M. Aiken, L. Starr-Spires, and M. L. Ogilvie. 1984. Antibodies against a 23 kd heparin-binding fragment of thrombospondin inhibit platelet aggregation. *Biochem. Biophys. Res. Commun.* 124:290–295.

Ginsburg, D., R. I. Handin, D. T. Bonthron, T. A. Donlon, G. A. P. Burns, S. A. Latt, and S. H. Orkin. 1985. Human von Willebrand factor (vWF): isolation of complementary DNA (cDNA) clones and chromosomal localization. *Science (Wash. DC).* 228:1401–1406.

Gray, A., T. J. Dull, and A. Ullrich. 1983. Nucleotide sequence of epidermal growth factor cDNA predicts a 128,000-molecular weight protein precursor. *Nature (Lond.).* 303:722–725.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene.* 28:351–359.

Henschen, A., F. Lottspeich, M. Kehl, and C. Southan. 1983. Covalent structure of fibrinogen. *Ann. NY Acad. Sci.* 408:28–43.

Horlein, D., P. P. Fietzek, E. Wachter, C. M. Lapiere,and K. Kuhn. 1979. Amino acid sequence of the aminoterminal segment of dermatosparactic calf-skin procollagen type 1. *Eur. J. Biochem.* 99:31–38.

Hourdille, P., M. Hasitz, F. Belloc, and A. T. Nurden. 1985. Immunocytochemical study of the binding of fibrinogen and thrombospondin to ADP- and thrombin-stimulated human platelets. *Blood.* 65:912–920.

Hubbard, S. C., and R. J. Ivatt. 1981. Synthesis and processing of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* 50:555–583.

Hynes, R. 1985. Molecular biology of fibronectin. *Annu. Rev. Cell Biol.* 1:67–90.

Hynes, R. O., J. E. Schwarzbauer, and J. W. Tamkun. 1986. Isolation and analysis of cDNA and genomic clones of fibronectin and its receptor. *Methods Enzymol.* In press.

Jaffe, E. A., J. T. Ruggiero, L. K. Leung, M. J. Doyle, P. J. McKeown-Longo, and D. F. Mosher. 1983. Cultured human fibroblasts synthesize and secrete thrombospondin and incorporate it into extracellular matrix. *Proc. Natl. Acad. Sci. USA.* 80:998–1002.

Klee, C. B., T. H. Crouch, and P. G. Richman. 1980. Calmodulin. *Annu. Rev. Biochem.* 489:489–515.

Kornblihtt, A. R., K. Umezawa, K. Vibe-Pedersen, and F. E. Baralle. 1985. Primary structure of human fibronectin: differential splicing may generate at least ten polypeptide from a single gene. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:1755–1759.

Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell.* 44: 283–292.

Kretsinger, R. H. 1980. Crystallographic studies of calmodulin and homologs. *Ann. NY Acad. Sci.* 356:14–19.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (Lond.).* 227:680–685.

Lahav, J., M. A. Schwartz, and R. O. Hynes. 1982. Analysis of platelet adhesion with a radioactive chemical crosslinking reagent. Interaction of thrombospondin with fibronectin and collagen. *Cell.* 31:253–262.

Lahav, J., J. Lawler, and M. A. Gimbrone. 1984. Thrombospondin interactions with fibronectin and fibrinogen. Mutual inhibition in binding. *Eur. J. Biochem.* 145:151–156.

Lawler, J. 1981. Prediction of the secondary structure of platelet factor 4 and β-thromboglobulin from their amino acid sequences. *Thromb. Res.* 21:121–127.

Lawler, J. 1986. Review: the structural and functional properties of thrombospondin. *Blood.* 67:1197–1209.

Lawler, J. W., and H. S. Slayter. 1981. The release of heparin-binding peptides from platelet thrombospondin by proteolytic action of thrombin, plasmin and trypsin. *Thromb. Res.* 22:267–279.

Lawler, J., and E. R. Simons. 1983. Cooperative binding of calcium to thrombospondin. *J. Biol. Chem.* 258:12098–12101.

Lawler, J., H. S. Slayter, and J. E. Coligan. 1978. Isolation and characterization of a high molecular weight glycoprotein from human blood platelet. *J. Biol. Chem.* 253:8609–8616.

Lawler, J., F. C. Chao, and C. M. Cohen. 1982. Evidence for calcium-sensitive structure in platelet thrombospondin: isolation and partial characterization of thrombospondin in the presence of calcium. *J. Biol. Chem.* 257: 12257–12265.

Lawler, J., L. H. Derick, J. E. Connolly, J. H. Chen, and F. C. Chao. 1985. The structure of human platelet thrombospondin. *J. Biol. Chem.* 260:3762–3772.

Lawler, J., A. M. Cohen, F. C. Chao, and D. J. Moriarty. 1986*a*. Thrombospondin in essential thrombocythemia. *Blood.* 67:555–558.

Lawler, J., J. E. Connolly, P. Ferro, and L. H. Derick. 1986*b*. Thrombin and chymotrypsin interactions with thrombospondin. *Ann. NY Acad. Sci.* In press.

Leung, L. L. K., and R. L. Nachman. 1982. Complex formation of platelet thrombospondin with fibronogen. *J. Clin. Invest.* 70:542–549.

Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science (Wash. DC).* 227:1435–1441.

Majack, R. A., S. C. Cook, and P. Bornstein. 1985. Platelet-derived growth factor and heparin-like glycosaminoglycans regulate thrombospondin synthesis and deposition in the matrix by smooth muscle cells. *J. Cell Biol.* 101:1059–1070.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 149–163.

Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sin, and A. Efstratiadis. 1978. The isolation of structural genes from libraries of eukaryotic DNA. *Cell.* 15:687–701.

Margossian, S. S., J. Lawler, and H. S. Slayter. 1981. Physical characterization of platelet thrombospondin. *J. Biol. Chem.* 256:7495–7500.

McDonough, R. P., J. McDonough, T. E. Petersen, H. C. Thogersen, K. Skorstengaard, L. Sottrup-Jensen, S. Magnusson, A. Dell, and H. R. Morris. 1981. Amino acid sequence of the factor $XIII_a$ acceptor site in bovine plasma fibronectin. *FEBS (Fed. Eur. Biochem. Soc.) Lett.* 127:174–178.

McKeown-Longo, P. J., R. Hanning, and D. F. Mosher. 1984. Binding and degradation of platelet thrombospondin by cultured fibroblasts. *J. Cell Biol.* 98:22–28.

Mumby, S. M., G. J. Raugi, and P. Bornstein. 1984. Interactions of thrombospondin with extracellular matrix proteins: selective binding to type V collagen. *J. Cell Biol.* 98:646–652.

Murphy-Ullrich, J. E., and D. F. Mosher. 1985. Localization of thrombospondin in clots formed in situ. *Blood.* 66:1098–1104.

Nussenzweig, V., and R. S. Nussenzweig. 1985. Circumsporozoite proteins of malaria parasites. *Cell.* 42:401–403.

Ozaki, L. S., P. Svec, R. S. Nussenzweig, V. Nussenzweig, and G. N. Godson. 1983. Structure of the Plasmodium knowlesi gene coding for the circumsporozoite protein. *Cell.* 34:815–822.

Pirrotta, V., M. Ptashne, P. Chadwick, and R. Steinberg. 1971. The isolation of repressors. *In* Procedures in Nucleic Acid Research. Vol. 2. G. L. Cantoni and D. R. Davies, editors. Harper and Row, New York. 703–715.

Plow, E. F., R. P. McEver, B. S. Coller, V. L. Woods, Jr., G. A. Marguerie, and M. H. Ginsberg. 1985. Related binding mechanisms for fibrinogen, fibronectin, von Willebrand factor, and thrombospondin on thrombin-stimulated human platelets. *Blood.* 66:724–727.

Pytela, R., M. D. Pierschbacher, M. H. Ginsberg, E. F. Plow, and E. Ruoslahti. 1986. Platelet membrane glycoprotein IIb/IIIa: member of a family of arg-gly-asp-specific adhesion receptors. *Science (Wash. DC).* 231:1559–1562.

Raugi, G. J., S. M. Mumby, D. Abbott-Brown, and P. Bornstein. 1982. Thrombospondin: synthesis and secretion by cells in culture. *J. Cell Biol.* 95:351–354.

Raugi, G. J., S. M. Mumby, C. A. Ready, and P. Bornstein. 1984. Location

and partial characterization of the heparin-binding fragment of platelet thrombospondin. *Thromb. Res.* 36:165–175.

Roberts, D. D., J. A. Sherwood, S. L. Spitalnik, L. J. Panton, R. J. Howard, V. M. Dixit, W. A. Frazier, L. H. Miller, and V. Ginsburg. 1985. Thrombospondin binds falciparum malaria parasitized erythrocytes and may mediate cytoadherence. *Nature (Lond.).* 318:64–66.

Ruoslahti, E., and M. D. Pierschbacher. 1986. Arg-Gly-Asp: a versatile cell recognition signal. *Cell.* 44:517–518.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463–5467.

Schwarzbauer, J. E., J. W. Tamkun, I. R. Lemischka, and R. D. Hynes. 1983. Three different fibronectin mRNAs arise by alternative splicing within the coding region. *Cell.* 35:421–431.

Scott, J., M. Urdea, M. Quiroga, R. Sanchez-Pescador, N. Fong, M. Selby, W. J. Rutter, and G. I. Bell. 1983. Structure of mouse submaxillary messenger RNA encoding epidermal growth factor and seven related proteins. *Science (Wash. DC).* 221:236–240.

Silverstein, R. L., L. L. K. Leung, P.C. Harpel, and R. L. Nachman. 1984. Complex formation of platelet thrombospondin with plasminogen. Modulation of activation by tissue activator. *J. Clin. Invest.* 74:1625–1633.

Turk, J. L., and T. C. Detwiler. 1986. Thiol-disulfide exchange by thrombospondin: evidence for a thiol and a disulfide bond protected by calcium. *Arch. Biochem. Biophys.* 245:446–454.

Vischer, P., H. Beeck, and B. Voss. 1985. Synthesis, intracellular processing and secretion of thrombospondin in human endothelial cells. *Eur. J. Biochem.* 153:435–443.

Vogelstein, B., and D. Gillespie. 1979. Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. USA.* 76:615–619.

Wolff, R., E. F. Plow, and M. H. Ginsberg. 1986. Interaction of thrombospondin with resting and stimulated human platelets. *J. Biol. Chem.* 261(15): 6840–6846.

Young, R. A., and R. W. Davis. 1985. Immuno-screening λgt11 recombinant DNA expression libraries. Genetic Engineering Principles and Methods. Vol. 7. J. Setlow and A. Hollaender, editors. Plenum Publishing Corp., NY. 29–41.