Sequence of a cDNA for mouse thrombomodulin and comparison of the predicted mouse and human amino acid sequences

William A. Dittman and Philip W. Majerus*

Division of Hematology/Oncology, Departments of Internal Medicine and Biological Chemistry, Washington University School of Medicine, St Louis, MO 63110, USA
Submitted November 16, 1988

Accession no. J04060

Thrombomodulin is an endothelial cell surface anticoagulant cofactor for activation of the serum protease protein C. We have previously described a complete cDNA for human thrombomodulin (1), and a partial cDNA for mouse thrombomodulin (2). We now report the complete cDNA for mouse thrombomodulin, and compare the predicted amino acid sequence to human. The sequence contains a 5' untranslated region of 158 bases, followed by 1731 bases encoding 577 amino acids, and 1745 bases of 3' untranslated sequence terminating in a poly A tail (24 bases). The sequence was determined by sequencing both strands of two independent cDNA clones, one encoding bases 1-833, the other 446-3658.

```

1  CCGAGGAGC TAGGAGGCTG GGTCTAGAAC GTCTAGGTTG TGATGAGGCG TAGCTGCTGT CCGCGAGCG ATACGAGAAC CCGCTTGGCT GAGCTGGGTG
101  TCTGAGCAGC CCGAGGAGCTA GCTTTTCTTG GAACTTGCAC GTGCGACCCC TGACGAGCAT GCTTGGGAT TTTCTTCTGG GTGTGCTGGC TCGCTAGGAC
201  CTGGGGCTCT CCGAGCTACG CAAGCTGTGG CCGACAGGCA GTCAATGGGT GAGCATGAGG TACTTGGCC TTTTCCAGGG CCGCGGAGCC TTCTTGATGT
301  CCGAGGAGCG CTGGAGGCGG CTGAGAGAAC ATTGTATGAC AGTGCGGCTCG TCAATGGCTCT CCGATTTGAT CTGCGCTCTG CTGAGGAGCA GAGTATGTGA
401  TTTAGGCGCC TGATGCTGTT TAGAGCTTCC GAGGAGGCTGT GAGGAGGCTGT GCGGCTGGCC GCGTTTCAGT GCGTCTACTG CGATACGAC
501  ACGAGTTGCA GAGAGTGGCG GCGCGGAGG GAGCAGAGCG CTGCGACTCTG GCGGCTCTCTG TGCGTACGCG TCGTGAGAGT AACTGAGACT GAGCGGCGG
601  ACGCGGCTCT GAGAGAGAG CAGTGTGAGA CTGAGAGGCA GCGTTTCTCT TGATGATTTT ACTTACGAGC TCTGTGAGCG CTCTGAGCG TGATGAGCTG
701  CGATCTGAGT GCTGGGAGCA TCTCTAGTAC CTGAGAGGCG CCGTTTGGCG TCGTGGTCTG GAGCTTTGAA AGCTTGGCGG TAGGAGCTTC GCGCGGCTGT
801  GAGCGGCTCT GCTGTGAGCT GGTGTGAGG GCGCGGCGCG GAACTTCTGA GCGGAGCTG GCGTGGAGG CCGAGAGGCG CTGGAATTCG AGCTGAGCA
901  AGTGTGGCTG TGATGACTTG TGCAAGAGA GAGGAGATGA ACCGAGTGC CTCTGGGCTA GAGCAGTGA CCGTGGAGCG GATGAGACTT CGTGTGAGG
1001  ACGTGTGGTT CAGTGTGAGA AGCAGTCTCG CCGAGCTATT TGTGTGAGCA AGCTTGAGT GCGAGGCTCT TACTCTGTGA TGTGTGAGC AGGCTAGCAG
1101  TGTGTGAGC ACGGAGCGCG GGTGTGAGG CTGATGACT GTAGCGAGCG GCGCAATCTA GCGGCGAGC TCGTGTGATA CAGCAGAGCG GCGCTTGGAT
1201  GCTTCTGTGA TGATGGCTAT GAGTGTGAGG ATGAGAGTGT CGTGAGCTTT CTGATGAGTT GTTGTGGACT TACTTGGAGG TTTGAGTGGC AGCGCTAGCA
1301  CCGAGGAGCG TACGAGTCTG TCTGCGCTCG AGGCTTGGCA CCGAGGCGCG ATRGAGCGCA CAGTGTGATA ATGTTCTTGA ATGAACTTCA ATGAACTTCA
1401  GAGTGTGAGC CTAGCTCTCT TACTGTGTTT GAGTGGGCTG AAGGCTTCAT CTGAGAGGAG GGTTGGTAT CCGAGGAGCT TGATGAGTGC AGTCAAGGCG
1501  AATGCTTCTC CAGTGTGAT GGAAGCTTCT CTGCTTCTGA TGAGTGTATC TCGCGGCTCTG ACAGAGGCTG TCGTGTGAGG ATTAGCAGAG ACCTGTGAGC
1601  CATGCTGTTT ACGGAGAGCA CAGAGGAGCA GAGGCGCTCT GCGGAGGCTC CAGTCAAGCG TAGCGAGCG TCTGTGAGG GTGCGGCTCT TCGAGGCGCA
1701  TGTGACTCTG GGTGCTCTAT TGGCTATTTC ATTGGGAGCG TGTGCGCTCT GGTGGGCTCTG TGTGGGCTCTG TGTGGGCTCTG TGTGGGCTCTG TGTGGGCTCTG
1801  CTGTGAGCA GCTGAGTAC AAGTGGGAT ATTGGGAGCG GAGGAGTGTG CTGAGAGCTG TCGAGAGCTA TCGAGAGCTG CAGAGAGTCT GAGGAGTGTG
1901  CTGTGAGCA CAGGAGTGGG TTTGTCTCTT CCGGCTCTCT ACTCTCTCT TCTCTCTCT TCTCTCTCT TCTCTCTCT TCTCTCTCT TCTCTCTCT
2001  CAGCTGCTCT GGTATATATA CAGAGAGCA GCGGCTCTCT TCGAGAGCG CAGGAGTGTG GAGGAGTGTG AGGAGAGCTA CCGAGTGTCT TCGAGAGCA
2101  TACTGTGACA CTGGGCGGAG GTGGGAGATA CAGGAGGAT CCGAGAGCT CCGAGTGTCT CAGTGTCTCT CCGAGTGTCT CCGAGTGTCT CCGAGTGTCT
2201  CAGAGCTTGA CCGAGGAGCG TGAGAGTATC CCGAGTATT ATTCTTTTGA AAGTATTTAG TATCTCTCT CCGAGTGTCT TCTCTCTCT GAGAGTCTCT
2301  ACGCGGAGC AGCTTCTCTA GTCGCTGAT TCGGCGCTCT CAGTGTCTCT CCGAGTGTCT CCGAGTGTCT CCGAGTGTCT CCGAGTGTCT CCGAGTGTCT
2401  TATGTGAGC AAGAGAGGAA CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT
2501  TACAGAGCG TAAAGATTT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT CAGTGTGAT
2601  AACTCTTGTG TAAAGAGG TAAAGAGG TAAAGAGG TAAAGAGG TAAAGAGG TAAAGAGG TAAAGAGG TAAAGAGG TAAAGAGG
2701  GCGTCTGAGC TGTATATAT TTTTGTGAT TCGAGAGCT TCGAGAGCT TCGAGAGCT TCGAGAGCT TCGAGAGCT TCGAGAGCT TCGAGAGCT
2801  TACTGTGAGC TGTATATAT TTTTGTGAT TCGAGAGCT TCGAGAGCT TCGAGAGCT TCGAGAGCT TCGAGAGCT TCGAGAGCT TCGAGAGCT
2901  TGTGCTGCTG CCGAGCTGAG TCTGAGTCT GCGTGGGCTCT TGTAGTCTCG AAGTGTGAGC AAGTGTGAGC AAGTGTGAGC AAGTGTGAGC AAGTGTGAGC
3001  TCTGGGAGTG CCGAGTGTG TCTGAGTCT TGTAGTCTCG AAGTGTGAGC AAGTGTGAGC AAGTGTGAGC AAGTGTGAGC AAGTGTGAGC AAGTGTGAGC
3101  GAGTGTGAGC TGTGCTGCTG AATGCTGAG AATGCTGAG AATGCTGAG AATGCTGAG AATGCTGAG AATGCTGAG AATGCTGAG AATGCTGAG
3201  AAGTGTGAGC GCGAGTGTG CAGTGTGAT CAGGAGACT GGTCTTGA GGTCTTGA GGTCTTGA GGTCTTGA GGTCTTGA GGTCTTGA
3301  GCGATGTATA GCGATTTAT TATGAGGAA CTGAGGAT TCTCTGAT AGCTTGTCT TCTGTGAGC AATGAGTGT AAGTGTGAGC CCGTGTGAGC
3401  GTGCGAGCA GAGTGTGAG GGTGTTGTT TATGAGGAG TAGGATGTG TTAGTGTG TTAGTGTG TTAGTGTG TTAGTGTG TTAGTGTG
3501  AAGAGGTTT TCTGTGAGG TTTGTGAGG TTAGTGTG TTAGTGTG TTAGTGTG TTAGTGTG TTAGTGTG TTAGTGTG TTAGTGTG
3601  CAGATGTGA AATGAGTGTG AATGAGTGTG AATGAGTGTG AATGAGTGTG AATGAGTGTG AATGAGTGTG AATGAGTGTG AATGAGTGTG AATGAGTGTG

```

Predicted Amino Acid Sequence of Thrombomodulin

```

MM 1  MLGIFTLVAPLASLGLSALAKLTQTSQVEHCFALGQFATLDAASQNRQLQHLMIVRSSVAADVLSLLSQSMDLGE...MLGLPQGGDDFV 100
HM  VLV A LG FP EP G D IP N I DG R NCGVGRRRL P G K
101  HLGLPAGFQVWVSTSYSSWARRNDQAPLQGLGVSTANLAEQWAEQCEETELGFLCFEYPTASCRLLVNTVTRPAHLSSTHTPHG 200
R N LIDNG A A ENY S I QQ WKD H P T A E PQA AV I G RA
201  SGNDFITLVSSAAVEPLGLEILHAPPTSGSHAWATGAWNCVSVENGGCEYLRNSNEPRCLQRNDLQNGRSQRPVWQSNLELHCFVSN 300
R A A QMT AVQ R P D HA AIRQA Q AGA TASAT D P
301  AEPVGSYSCHETGYQLAADRQCEEDMDCKGGRNCPGLVNTGQFQCYDYELVMDGQVELLDYFQSNRQEQQVSPVLYRCICARFAPRPD 400
PDQ ILE R Q ILE S R Q H PND PV RA Y LNQ SLVE I H
401  EPHKCFNENSTQCPADDPNSPTVCCPFGILDEGSVCTDIDEGSGQECFTSECRNPGYECICGPDYALAQISKDDCPVPRVETKEEGSGSEPP 500
R Q Q A TQAS Y D FTI EN GCSG H L TF S RH GT SKG DGS...DS
501  VSPSTPSPTSPSPARVHVSGLIGISIASISVALLALLCHLRKQGNARALEYKASSAEVVLQHRDRLQKF 580
P TLT AVGL L C NM APS E P RL

```

Mouse is in the top line. Differences found in human are shown underneath.

Supported by NIH Grants HLBI14147, HLBI 16634, T32 HLBI 07088, and the Monsanto-Washington University Biomedical Research Grant.

*To whom reprint requests should be addressed

References: (1) Wen, D., Dittman, W.A., Ye, R.D., Deaven, L.L., Majerus, P.W., and Sadler, J.E. (1987) *Biochemistry* **26**:4350-4357; (2) Dittman, W.A., Kumada, T., Sadler, J.E., and Majerus, P.W. (1988) *J. Biol. Chem.* **263**:15815-15822.