Human tenascin: primary structure, pre-mRNA splicing patterns and localization of the epitopes recognized by two monoclonal antibodies

Annalisa Siri, Barbara Carnemolla, Marco Saginati, Alessandra Leprini, Giorgio Casari¹, Francisco Baralle¹ and Luciano Zardi*

Laboratory of Cell Biology, Istituto Nazionale per la Ricerca sul Cancro, Viale Benedetto XV 10, 16132 Genoa and ¹International Centre for Genetic Engineering and Biotechnology, Patriciano 99, 34012 Trieste, Italy

Received November 6, 1990; Revised and Accepted January 9, 1991

EMBL accession no. X56160

ABSTRACT

By sequencing cDNA clones which cover the complete coding region of human tenascin (TN), we have established its primary structure. This confirms that, as in the case of chicken, TN is mainly made up of three groups of sequences with a high homology to Epidermal Growth Factor (EGF) fibronectin (FN) type III repeat and fibrinogen. Furthermore, we have determined the amino-terminal sequence of the mature peptide directly on purified TN. The main differences with respect to the chicken TN molecule are that in the human there are 14 and half EGF-like repeats compared to 13 and half in the chicken and that, as previously reported, there are 15 FN-like repeats compared to 11 in the chicken. By Polymerase Chain Reaction (PCR) amplification we have also studied the different splicing patterns of the TN pre-mRNA in cultured cells. The results show the presence of at least four different isoforms containing different numbers of FN-like type III repeats. Using purified human TN as immunogen, we have obtained numerous monoclonal antibodies (Mabs) to TN. By screening a human melanoma cDNA library in the expression vector λ gt11 with these Mabs and subsequently sequencing the insert of the positive clones, we have been able to localize, within the TN molecule, the epitopes recognized by two of these Mabs: BC-4, which recognizes an epitope within the EGF-like sequence and BC-2 which recognizes an epitope within the FN like type III repeats whose expression is regulated by alternative splicing of the TN pre-mRNA. Thus, while the Mab BC-4 may be useful in studies on TN distribution (since it recognizes all different TN isoforms) BC-2 may be useful in the study of the expression of particular TN isoforms generated by the alternative splicing of the TN primary transcript.

INTRODUCTION

Tenascin (TN) is a polymorphic high molecular mass extracellular matrix glycoprotein composed of six similar subunits joined together at their NH_2 -terminal by disulfide bonds (for reviews see references 1-5).

The sequence of cDNA clones codifying for the complete chicken TN subunit has recently been reported. The deduced amino acid sequence shows that chicken TN is mainly made up of 13 and half epidermal growth factor like repeats (EGF-L), 11 units similar to fibronectin type III repeats (FN-L), and, at the C-terminal, by a sequence with 40% homology to the globular domain of the β -and γ -chain of fibrinogen (6-7). It has been shown that TN displays a restricted distribution in normal adult tissues and that its expression is ontogenically programmed. In fact, TN is transiently expressed in many developing organs and it has been proposed that TN could modulate epithelialmesenchymal and neuronal-glial interactions (8-13). However, neoexpression, or a dramatically increased expression, of TN has been documented in a variety of tumors, thus suggesting that this glycoprotein may play a role in tumor cell-stroma interdependence (2,8,14-22). Since it has also been shown that TN stimulates the growth of mammary tumor cells in vitro (8), it has been advanced that TN plays a role in supporting the growth of epithelial tumors in vivo (23).

Since transforming-growth-factor- β (TGF- β) induces a dramatic increase in TN expression in cultured fibroblasts (24), the belief has surfaced that epithelial tumor cells, which in general do not produce TN, through the release of TGF- β (and most likely other factors) force the surrounding mesenchymal cells to organize a tumor stroma which is different from the normal organ stroma and more suitable to sustain tumor growth and invasion (23).

It has recently been demonstrated that TN is the same protein as miotendinous antigen (25), JI (26), cytotactin (12), gliomaassociated-extracellular matrix (ECM) antigen (22), gp 150-225

GAATTCGCTAGAGCCCTAGAGCCCCAGCAGCACCCAGCCAAACCCACCTCCACCATGGGGGGCCATGACTC 70 AGCTGTTGGCAGGTGTCTTTCTTGCTTTCCTTGCCCTCGCTACCGAAGGTGGGGTCCTCAAGAAAGTCAT 140 Q L L A G V F L A F L A L A T E G G V L K K V I CCGGCACAAGCGACAGAGTGGGGTGAACGCCACCCGCCAGAAGAGAACCAGCCACGGGGTTTAACCAC 210 R H K R Q S G V N A T L P E E N Q P V V F N H GTTTACAACATCAAGCTGCCACTGGGATCCGCTGGTGGGGGAGAAAG 280 V Y N I K L P V G S Q C S V D L E S A S G E K ACCTGGCACCGCCTTCAGAGCCCAAGGAAAGCTTTCAGAGCACAAGATGGGGAAAACCAGATTGT 350 D L A P P S E P S E S F O E H T V D G E N O I V 420 490 560 C C L Q P A T G R L D T R P F C S G R G N F S T TGAAGGATGTGGCTGGGATGTCTGCGGAACGCCCCAACTGCTCTGAGCCCGAATGTCCAGGC 630 E G C G C V C E P G N K G P N C S E P E C P G AACTGTCACCTTCGAGGCCGGTGCATTGATGGGCAGTGCATCTGTGACGACGGCCTTCACGGGCGAGGACT 700 N C H L R G R C I D G Q C I C D D G F T G E D GCAGCCAGCTGGCTTGCCCCAGCGACTGCCATGACCAGGGCAATGGAGTCTGCATCGATCTGCAT 770 840 E G Y A A D C S R E I C P V P C S E E H G T C GTAGATGGCTTGTGTGTGTGGCCACGATGGCTTTGCAGGCGATGACTGCAACAAGCCTCTGTGTCTCAACA 910 V D G L C V C H D G F A G D D C N K P L C L N ATTGCTACAACCGTGGACGATGCGTGGAGAATGAGGGGTTTCACGGGGTAGACTG 980 N C Y N R G R C V E N E C V C D E G F T G E D C Castgagetcatetgececaatgactgettegacegegeegeetgeateaatggacegeaatga 1050 S E L I C P N D C F D R G R C I N G T C Y C E GARGETTEALAGETGARGATEGEGGAAACCEACTEGECICACATEGECIGACECAGGECIGGETETE E G F T G E D C G K P T C P H A C H T Q G R C AGGAGGEGCAGTETGTATGTGATGAGGECTTTGCCGCTGTGGACTGCAGCAGAGAGTGTCCTCCTGA 1120 1190 1260 1330 G E L K C P N G C S G H G R C V N G Q C V C D AGGGCTATACTGGGGAGGACTGCAGCCAGCTACGGTGCCCCAATGACTGTCACAGTCGGGGCCGCCGCTGCTG 1400 E G Y T G E D C S Q L R C P N D C H S R G R C V CGAGGGCAAATGTGTATGTGAGCAAGGCTTCAAGGGCTATGACTGCAGTGACATGAGCTGCCCTAATGAC 1470 1540 1610 R D R Q C P R D C S N R G L C V D G Q C V C E D CGCCTTCACCGGCCCTGACTGTGCAGAACTCTCCTGTCCAAATGACTGCCATGGCAGGGTCGCTGTGTG 1680 G F T G P D C A E L S C P N D C H G Q G R C V AATGGGCAGTGCGTGTGCCATGAAGGATTTATGGGCAAAGACTGCAAGGACGAAAGATGTCCCAGTGACT 1750 H G Q C V C H E G F N G K D C K E Q R C P S D GTCATGGCCAGGGCCGTGGACGGCCAGTGCATCTGCACGGGCTTCACAGGCCTGGACGTGG 1820 1890 1960 2030 2100 E G G L E M Q F R V P G D Q T S T I I R E L E CCTGGTGTGGAGTACTTTATCCGTGTATTGCCATCCTGGAGAACAAGAAGAGCATTCCTGTCAGGGCCA 2170 P G V E Y F I R V F A I L E N K K S I P V S A GGTGGCCACGTACTTACCTGCACCTGAAGGCCTGAAATTCAAGTCCATCAAGGAGACATCTGTGGAAGT VATYLPAPEGLKFKSIKETSVEY GGAGTGGGATCCTCTAGACATTGCTTTTGAAACCTGGGAGATCATCTTCCGGAATATGAATAAAGAAGAT L D I A F E T W E I I F R N M N K E D GAGGGAGAGATCACCAAAAGCCTGAGGAGGCCAGAGACCTCTTACCGGCAAACTGGTCTAGCTCCTGGGC 2380 F K P L A E I D G I E L T Y G I K D V P G D R CCACCATCGATCTCACAGAGGACGAGAACCAGTACTCCATCGGGACCTGAGGCTGACAGTAGGAC 2660 2730 2800 PKSQQATTKKTTLTGLRPGTEYGI GGAGTTTCTGCTGTGAAGGAAGAAGAAGGAGGAGCAATCCAGCGGACCATCAACGCAGGGTTGGACA G V S A V K E D K E S N P A T I N A T E L D CGCCCAAGGACCTTCAGGTTCTGAAACTGCAGAGACCAGGCTGACCTGCTGCTGGAGACACCGTTGGC 3080 3150 K F D R Y R L N Y S L P T G Q N V G V Q L P R MCACCACTTICTATGTCCTGAGGGCCTGGACCAGGACTACATGTCCTCCTGACAGCCGAGA 3220 N T T S Y V L R G L E P G Q E Y N V L L T A E MGGCAGACACMGAGCAAGCCCGCACGTGTGAAGGCATCCACTGAACAGCCCCTCAGCTGGAAAACCT 3290 3360 I I Q V Q E A N K V E A A R N L T V P G S L GCCTGTGGACATACCGGCCTCAAGGCTGCTACGCCTTATACAGTCTCCATCTATGGGGTGATCCAGGG 3500 R A V D I P C L K A A T P Y T V S I Y G V I Q G CTATAGAACACCAGTGCTCTGCTGAGGCCTCCACAGGGGAAACTCCCAATTGGGAGAGGTCGTGGTG 3570

GCCGAGGTGGGCTGGGATGCCCTCAAACTCAACTGGACTGCTCCAGAAGGGGGCCTATGAGTACTTTTTCA 3640 A E V G W D A L K L N W T A P E G A Y E Y TTCAGGTGCAGGAGGCTGACACAGTAGAGGCAGCCCAGAACCTCACCGTCCCAGGAGGACTGAGGTCCAC 3710 3780 3850 T T P L S V E V L T E E V P D N G N L T V T E TRACTOGGATOCTOTOAGACTGAACTGAACCACGCCAGATGGAACCTATGACCAGTTTACTATTCAGGT 3920 VS N D A L R L N N T T P D G T Y D O F T I O V CCAGGAGGCTGACGAGGTGGAAGAGGCTCACAATCTCACGGTTCCTGGGAGCCGGGAGAGGCTGAGAAGC 3990 Q E A D Q V E E A H N L T V P G S L R S N E I CCAGGCCTCAGGGCTGCACTCCTTACACAGTCACCCTGCACGCGAGGTCAGGGCCACAGCACTCGAC 4060 P G L R A G T P Y T V T L H G E V R G H S T R CCCTTGCTGTAGAGGTCGTCACAGAGGTCTCCCCACAGCTGGCAGATTAGCCGTGTCTGAGGTCGCCTG 4130 LAVEVVTEDLPOLGDLAVSEVG **GEATGECCTCAGACTCAACTGGACCGCAGCTGACAATGCCTATGAGCACTTTGTCATTCAGGTGCAGGAG** 4200 D G L R L N W T A A D N A Y E H F V I Q V Q E GTCAACAAAGTGGAGGCAGCCCCAGAACCTCACGTTGCCTGGCAGCCTCAGGGCTGGGACATCCCGGGCC 4270 V N K V E A A O N L T L P G S L R A V D I P G TCGAGGETGCCACGCCTTATAGAGETCCATCTATGGGGGGTGATCCGGGGGCATAGAACACCAGTACTCC 4340 LEAATPYRVSIYGVIRGYRTPVLS TGCTGAGGCCTCCACAGCCCGAAAGAGCCTGAAATGGAAACTTGAATGTTTCTGACATAACTCCCGAGAGC 4410 A E A S T A K E P E I G N L N V S D I T P E S TTCAATCTCTCCTGGATGGCTACCGATGGGATCTTCGAGACCTTTACCATTGAATTATTGATTCCAATA 4480 N L S W M A T D G I F E T F T I E I I D S I $\begin{array}{c} \text{Olitotatic totatic t$ A T T E A L P L L E M L T I S D I M P T G F T TTTCTGGATGCATGCAGGATGGATGCATCGAAGCACTTTCAACAAGGTGATGGAGGTGGAGCTGGAGCTGGAGCTGGAGCTGGAGCTGAGGAGTGGAGCTGAAGCTGGAGCTGAAGCTGGAGCTGAAGCTGGAGCTGAAGCTGGAGCTGAAGCTGGAGCTGAAGCTGGAGCTGAAGCTGGAGCTGAAGCTGGAGCTGAAGCTGGAGTGGAGCTGAAGCGGAGTGGAGCTGAAGCTGGAGCTGAAGCTGGAGTGGAGCTGAAGCTGGAGTGGAGCTGAAGCGGAGTGGAGCTGAAGCGGAGTGGAGCTGAGGCTGAAGCGGAGTGGAGCTGAGGGCGATGAGGGCGATGAGGGGAGTGGAGCTGAGGCGCATAACTGGGAAT 4830 D P Q E F T L S G T Q R K L E L R G L I T G I GCTATGAGGTTATGGTCTCTGGCTTCACCCAAGGECATCAAACCAAGCCCTTGAGGGCTGAGATTGTTA G Y E V M V S G F T Q G H Q T K P L R A E I V CAGAAGCCGAACCGGAAGTTGACAACCTTCTGGTTTCAGATGCCACCCCGGAGGTTTCCGTCTGTCCTG 4970 T E A E P E V D N L L V S D A T P D G F R L S W GACAGCTGATGAAGGGGTCTTCGACAATTTTGTTCTCAAAATCAAGAATACCAAAAAGCAGTCTGAGCCA 5040 T A D E G V F D N F V L K I R D T K K Q S E P CTGGAAATAACCCTACTTGCCCCCGAACGTACCAGGGACATAACAGGTCTCAGAGAGGCTACTGAATACG 5110 LETTLLAPERTRDITGLREAT AMATTGAACTCTATGGAATAAGCAAAGGAAGGGATCCCAGACAGTGCTATAGCAACAACAGCCAT 5180 E I E L Y G I S K G R R S Q T V S A I A T T A M GCCTCCCCAMGGAACTATHTICLGGAATCATCATGAMATTCGCCTATTGCCCCCGAGGGACC G S P K B V I F S D I T B V S A T V S W R A P ACGCCCCAGTGGAGAGCTTCCCGATTACCTATGTGCCCCATTACAGGAGGTACCCCCCATGGTAACTG 5250 5320 T A Q V E S F R I T Y V P I T G G T P S N V T TGGACGGAACCAAGACTAGACCAGGCTGGTGAAACTCATACCTGGCGTGGAGTACCTTGTCAGCATCAT 5390 V D G T K T Q T R L V K L I P G V E Y L V S I I CCCCATGAAGGCCTTTGAGGAAAGTGAACCTGTCTGGGTGGCCATCT 5460 A M K G F E E S E P V S G S F T T A L D G P G L V T A N I T D S E A L A R N Q P A I A T V ACAGTTATGTCATCTCCTACACAGGCGAGAAAGGCCCAGAAATTACACGCCAGGGGGCGCGGGAGACAGAG 5600 D S Y V I S Y T G E K V P E I T R T V S G N T V GCAGTATGCTCTGACCGGACCTGGCCCGCGCATACACACTGAGAAACGGCCC 5670 EYALTDLEPATEYTLRIFAEKGP CAGAAGAGCTCAACGCCAAGTCCACAGACCTCGATCTCCCAAGAGCTTGACTGCTACTG 5740 Q K S S T I T A K F T T D L D S P R D L T A T AGGTCAGTCGGAAACTGCCCCCTTACCTGGCGACCCCCCCGGGCATCAGTCACCGGTTACCTGCTGGT VOSETALLTWRPPRASVTGYLL CTATGAATCAGTGGATGGCACAGTCAAGGAAGTCATTGTGGGTCCAGATACCACCTCCTACAGCCTGGCA GACCTGAGCCCATCCACCCACTACACAGCCAAGATCCAGGCACTCAATGGGCCCCCTGAGGAGCAATATGA 5950 D L S P S T H Y T A K I Q A L N G P L R S N H TCCAGACCATCTTCACCACAATTGGACTCCTGTACCCCTTCCCCAAGGACTGCTCCCAAGCAATGCTGAA 6020 I Q T I F T T I G L L Y P F P K D C S Q A H L N TGCAGACACGACCTCTGGCCTCTACACCATTTATCTGAATGGTGATAGGCTCAGGCGCTGGAAGTCTTC 6090 G D T T S G L Y T I Y L N G D K A Q A L E V F TGTGACATGACCTCTGATGGGGTGGATGGATGGTTCCTGAGACGCAAAAACGGACGCCGGAAACTTCT 6160 C D M T S D G G G N I V F L R R K N G R E N F ACCAAAACTGGAAGGCATATGCTGCTGGGTTTGGGGACCGCAGAGAAGAATTCTGGCTTGGGCTGGACAA 6230 O N W K A Y A A G F G D R R E E F W L G L D CCTGAACAAAATCACAGCCCAGGGGCAGTACGAGCTCCGGGTGGACCTGCGGGACCATGGGGAGACAGCC 6300 L N K I T A Q G Q Y E L R V D L R D H G E T A TITGCTGTCTATGACAAGTTCAGCGTGGGAGATGCCAAGACTCGCTACAAGCTGAAGGTGGAGGGGTACA 6370 F A V Y D K F S V G D A K T R Y K L K V E G Y GTGGGACAGCAGGTGACTCCATGGCCTACCACAATGGCAGATCCTTCTCCACCTTTGACAAGGACACAGA 6440 S G T A G D S M A Y H N G R S F S T F D K D T D TTCAGCCATCACCAACTGTGCTCTGCTACAAGGGGCTTCTGGTACAGGAACTGTCACCGTGTCAACCTG 6510 S A I T N C A L S T R G F M Y R N C H R V ATGGGGAGATATGGGGACAATAACCACAGTCAGGGCGTTAACTGGTTCCACTGGAAGGGCCACGAACACT 6650 6720 6790 6860 GTCTCCAACATGTTTCGTTTTGTTGTTGATCAGCAAAAATCTCCCAGTAGCAACAACATCGCGAATAGTTT TTTACTTCTCTTAGGTGGCTCTGGGATGGAGAGGGTAGGATGACAGGGTAGCATGTAGTTTTTAAACCA GCCGTATTTTACATGAGGCGGTATAATTAATTGTCATTATTTTTGTAGCAAAGATTAAATGTGTCATTG 6930 7000 7070 7210 AAAAAACCGGAATTC

(27), hexabrachion (28) and neuronectin (21) all previously described.

Here we report 1) the cDNA cloning and deduced complete amino acid sequence of human TN as well as the amino-terminal sequence of the mature peptide determined directly on purified TN; 2) the splicing patterns of TN pre-mRNA in cultured cells and 3) the production of monoclonal antibodies (Mabs) to human TN as well as the localization of the epitopes recognized by these Mabs within the TN molecules using β -galactosidase-TN fusion proteins expressed in E.coli.

MATERIALS AND METHODS

Isolation of TN cDNA clones and their sequencing

Human melanoma, human fetal brain, human melanoma SK-MEL-28 cell line and human breast cDNA libraries in bacteriophage λ gt11 were obtained from Clontech Laboratories (Palo Alto, CA). The first cDNA clone was obtained by screening the human melanoma library using the Mab BC-2 to human TN (15) and an immunoenzymatic kit purchased from Promega Biotec (Madison, WI). Other cDNA clones covering the complete coding region of the human TN gene were obtained by screening the above mentioned libraries with ³²P labelled DNA fragments from successively isolated TN cDNA clones. Overlapping clones spanning the entire coding region were subcloned into M13mp19 or M13mp18 vectors (New England Bio Labs, Beverly, MA) and analyzed in both strands using the Sequenase 2.0 DNA sequencing kit (United States Biochemical Corp., Cleveland, Ohio).

RNA extraction, cDNA synthesis, Polymerase Chain Reaction (PCR) amplification and Southern blot

Total RNA was prepared from human melanoma SK-MEL-28 cells following the method described by Chirgwin (29). cDNA was synthetized from 10 μ g of total RNA primed with oligo $(dT)_{12-18}$ and reverse transcribed using a cDNA synthesis kit (Amersham International, Buckinghamshire, England), following the manufacturer's manual. PCR reactions were performed for 35 cycles (1 min 93°C, 1 min 52°C, 1 min 72°C) in a final volume of 200 µl containing 50 mM KCl, 10 mM Tris-Cl pH 8.3, 1.5 mM MgCl₂, 0.01% (w/v) gelatin, 200 mM each dNTP, 0.05% Nonidet P40, 0.05% Tween 20, 100 pmol each oligo T1 (5' ATAAGCTTCCAAGAAACACCACTT, position: 3140-3160 of the sequence shown in Fig.1) and T5-2 (5' GGGCAAGTAGGGTTATT, position: 5046-5062 of the sequence shown in Fig.1), 5 U of Taq DNA polymerase (Amplitaq, Perkin Elmer Cetus, Norwalk, CT) and 0.5 ml of the cDNA mixture as template. The PCR mixture was loaded on a preparative 1.4% agarose gel; well resolved bands were cut out, electroeluted and ethanol-precipitated. The pellets were resuspended in 7 μ l of water; 2 μ l Sequenase buffer and 1 μ l (20 pmol) of sequencing primer were added. The primer/template solution was boiled for 3 min and frozen in dry ice. From this step the direct sequencing protocol follows the Sequenase manual

indications. For Southern blot analysis 5 μ l of the PCR mixture were loaded in two different wells of 1.4% agarose gel. Following electrophoresis, the gel was immersed twice for 15 min in 300 ml of 1.5 M NaCl, 0.5 M NaOH and twice for 15 min in 1 M NH₄ acetate and transferred for 3 hours to nitrocellulose membrane following standard techniques (30). The filter was cut between the loaded lanes and hybridized separately with ³²P labelled oligonucleotides T2 (5' AGGTAGACAATAAAATCA-GTA, position: 4554–4574 (see Fig.1) and T4 (5' CTTCAT-CAGCTGTCCAGGACAGA, position: 4962–4984 (see Fig.1) according to (30). The oligonucleotides were synthesized by an Applied Biosystem 370A synthesizer (Applied Biosystem, Foster City, CA).

Purification of TN and characterization of Monoclonal Antibodies (Mabs) BC-2 and BC-4 specific to TN

TN was purified from the conditioned medium of SK-MEL-28 human melanoma cell line by chromatography on hydroxyapatite (DNA-grade, Bio-Rad Laboratories, Richmond, CA) followed by immunoaffinity on Mab BC-4 conjugated to cyanogen bromide-activated-sepharose-4B (Pharmacia Fine Chemicals, Uppsala, Sweden). The complete procedure will be reported elsewhere. The purified TN was reduced and alkylated as described by Canfield and Anfisen (31). Protein sequencing was performed by the Center for Analysis and Synthesis of Macromolecules at Stony Brook (Stony Brook, NY).

cDNA libraries, in the expression vector $\lambda gt11$, were screened using the Mabs BC-2 and BC-4 and an immunoenzymatic kit purchased from Promega Biotec (Madison, WI). Bacteriophage DNA was prepared from positive clones, and the EcoRI inserts were excised, subcloned into a M13mp19 vector (New England Bio Labs, Beverly, MA) and sequenced.

For immunoblotting analysis the β -galactosidase-TN fusion protein were obtained as previously reported (32). All the cloning and subcloning procedures were carried out according to Sambrook et al. (30). SDS-PAGE and immunoblotting were performed as previously described (33-34).

RESULTS

cDNA sequence and pre-mRNA alternative splicing analysis of human TN

The cDNA and the deduced amino acid sequences of human TN are shown in Fig.1. The primary structure of human TN, arranged in groups of homologous repeats as suggested for chicken TN (7), is shown in Fig.2. The first 22 amino acids correspond to a typical signal peptide (35-36). Furthermore, the amino-terminal sequence of the mature peptide was determined directly on purified TN and it begins from residue 23 (Fig.2). The domain structure of a human TN subunit is schematically represented in Fig.3, showing the 14 and half EGF-L, the 15 FN-L and the fibrinogen like sequences and the potential N-glycosylation sites. The longer human TN reading frame is 2199 amino acids. Northern blot analysis of total RNA from the

Figure 1. cDNA and deduced amino acid sequences of human TN. The consensus polyadenylation site AATAAA (position 7258) is underlined. With respect to the sequence reported by Gulcher et al. (37) we found the following main differences: 1) our residue 538 is Q instead of R reported (residue 109 of ref.37). 2) our residue 679 is R instead of Q (residue 250 of ref.37) 3) our residue 1065 is R instead of H (residue 636 of ref.37). 4) from residue 1599 to the residue 1607 we have the following sequence: SGFTQGHQT instead of LWLHPRASN (residue 1170 to 1178 of ref.37). 5) our residue 1676 is I instead of L (residue 1247 of ref. 37). 6) We also found differences at the level of the bases, that do not, however, change the amino acids. At bases 1667 (A instead of G), 2100 (G instead of A), 3339 (G instead of A), 5080 (A instead of T), 5253 (G instead of A), 6214 and 6215 (TG instead of CT). These sequence data are available from EMBL/Genbank/DDBJ/NBRF-PIR under accession number X56160.



Figure 2. Human TN sequence arranged in groups of homologous repeats as suggested for chicken TN by Spring et al. (7). Potential N-linked glycosylation sites are underlined. All cysteine residues are highlighted as well as the conserved tryptophan, leucin and threonine residues used to align the FN-L type III repeats. Numbers on the right indicate the residue number; numbers on the left indicate the FN-L repeat number. (*) indicates the FN-L repeats which undergo alternative splicing. The amino-terminal residues identified by protein sequencing are underlined in bolder type.

human melanoma cell line SK-MEL-28, using the insert of the clone HT-12 as probe, showed a broad band of about 6,000-7,000 bases (data not shown) as expected from our cloning data indicating the presence of several TN mRNAs in this cell line. In fact, shorter reading frames generated by alternative splicing of the TN primary transcript have also been observed. We have found a cDNA clone (HT-5) in which the FN-L repeats 6,7,8,9 and 11 were lacking and another one (HT-33) in which only the FN-L repeat 11 was lacking (Fig.4). These observations agree and further the results by Gulcher et al. (37), who reported the sequences of two human TN cDNA clones, one containing all 15 FN-L repeats and one in which the FN-L repeats 6,7,8,9,10,11 and 12 were absent because of splicing of the TN primary transcript.

In order to confirm these various splicing patterns of the human TN primary transcript, we carried out PCR amplification experiments on cDNA derived from human melanoma cell line SK-MEL-28 total RNA. PCR amplification experiments were performed using as primers the oligonucleotides T1, hybridizing a coding region within the 5th repeat, and T5-2, hybridizing

a coding region within the 12th repeat. After electrophoresis of the PCR mixture, four bands were visualized of about 1900, 1650, 550 and 280 bp respectively (Fig.4). These bands were electroeluted and completely sequenced. Furthermore, Southern blots of the amplified fragments were hybridized with an oligonucleotide (T2) that specifically recognizes the 10th repeat, and separately with an oligonucleotide specific for the 12th repeat (T4). All four fragments hybridized to T4 as was expected due to the presence of the 12th repeat while T2 failed to recognize the smallest 284 bp fragment, confirming that it lacks the 10th FN-L repeat.

Localization within the TN molecule of the epitopes recognized by the Mabs BC-2 and BC-4 by using TN- β -galactosidase fusion proteins expressed in E.coli

In order to localize the epitopes recognized by the Mabs to human TN within the TN molecule, we have screened by an immunoenzymatic procedure, a human melanoma cDNA library in the expression vector $\lambda gt11$. We obtained two clones, $\lambda pTN1$

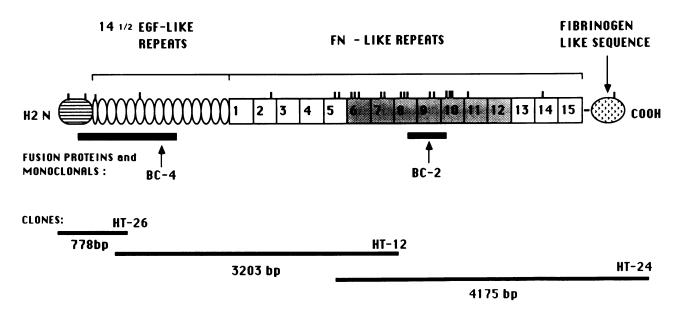


Figure 3. Model of the domain structure of a human TN subunit. The EGF-L, FN-L repeats as well as the fibrinogen like sequence are indicated. The FN-L repeats 6 to 12, whose expression is regulated by the alternative splicing of the pre-mRNA, are shaded. The potential N-linked glycosylation sites are indicated by small dashes. Also indicated are the overlapping cDNA clones covering the complete coding region of human TN which have been sequenced to deduce the TN primary structure. The figure also shows the β -galactosidase-TN fusion proteins expressed in E. coli used to localize within the TN molecule the epitopes recognized by the Mabs BC-2 and BC-4.

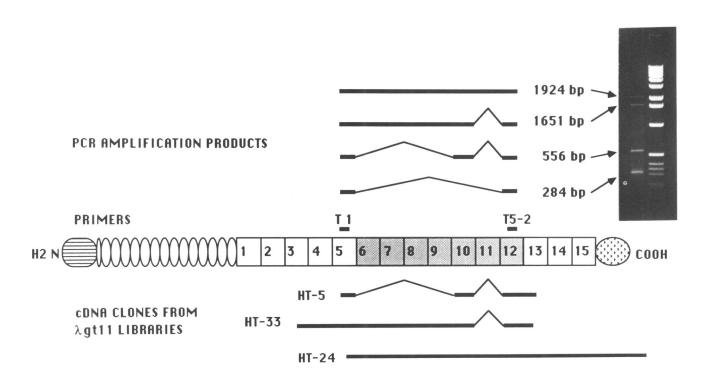


Figure 4. Study of different mRNA generated by alternative splicing of the TN primary transcript. The figure shows a model of the domain structure of a human TN subunit, the FN-L repeats 6 to 12 which undergo alternative splicing are shaded. In the lower part of the figure cDNA clones containing different amounts of FN-L repeats are shown. In the cDNA clone HT-5, obtained from a human melanoma cDNA library, only repeats 10 and 12 of the FN-L repeats undergoing alternative splicing are present, while in the clone HT-24 from a fetal human brain library all the repeats 6,7,8,9,10,11 and 12 are present, and in the clone HT-33 from a human melanoma cDNA library, only repeat 11 is lacking. The upper part of the figure shows the PCR amplification products obtained using the oligonucleotides T1 and T5-2 indicated in the figure (see also Materials and Methods) and total RNA from the human melanoma cell line SK-MEL-28. The 1% agarose gel of these PCR amplification products is shown in the inset. We obtained the following DNA fragment: a 1924 bp fragment containing repeats 5,6,7,8,9,10,11 and 12; a 1651 bp fragment containing the repeats 5,6,7,8,9,10 and 12; a 556 bp fragment containing the repeats 5,10 and 12 and a 284 bp fragment containing the repeats 5 and 12. Molecular weight markers (1kb, GIBCO BRL, Grand Island, NY, USA) are also shown. For details see Results. Λ indicates contiguity.

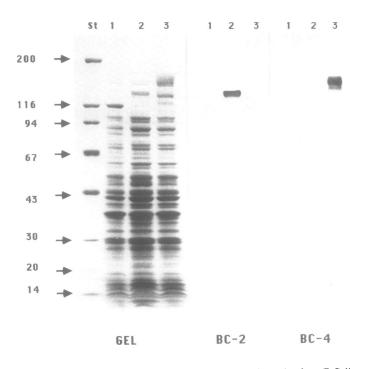


Figure 5. On the left, a 4-18% SDS-PAGE gradient of proteins from E.Coli infected by the expression vector $\lambda gt11$, by clones $\lambda pTN1$ and $\lambda pTN27$ (lanes 1,2 and 3 respectively). On the right, immunoblots of gels identical to those shown on the left using the Mabs BC-2 and BC-4, respectively. The values on the left indicate the molecular masses, in kD, of the standards (St).

and λ pTN27, producing β -galactosidase-TN fusion proteins which react strongly with the Mabs BC-2 and BC-4 respectively (Fig.5). The cDNA inserts of these two clones were purified, inserted into a M13mp19 vector and completely sequenced. The sequencing data demonstrated that the TN part of the fusion protein produced by the clone λ pTN1 is made up of a sequence from the residue 1316 to the residue 1450 while the fusion protein produced by the clone λ pTN27 contains the TN sequence from the residue 50 to the residue 453 (Fig.2 and 3).Thus, while the epitope recognized by the Mab BC-4 is present in all the TN isoforms the Mab BC-2 recognizes an epitope whose expression is regulated by alternative splicing of the TN pre-mRNA (Fig.3).

DISCUSSION

We here report the complete primary structure of human TN deduced from cDNA clones, a study of the splicing patterns of the TN primary transcripts in cultured human melanoma cells and the characterization of two Mabs specific to human TN using β -galactosidase-TN fusion proteins expressed in E. coli.

By comparing the human and chicken TN sequences, we found two main differences: 14 1/2 instead of 13 1/2 EGF-L repeats and 15 instead of 11 FN-L repeats.

The sequence of the first 22 amino acids of human TN corresponds to that of a typical signal peptide (35-36). The sequence of the following 10 amino acids (residues 23-33), is strongly basic since it contains 5 positively charged residues. In chicken a homologous sequence has been considered a putative propeptide since the mature chicken peptide started at residue 34 (6). However, direct amino-terminal sequencing of purified human TN shows that the mature peptide starts at residue 23 thus including this positively charged sequence that in chicken has been reported not to be present in the mature peptide. This

discrepancy may be explained by the fact that the chicken TN used to determine the amino-terminal sequence was purified from tissues (insoluble TN) (6) while we have purified TN from the medium of cultured cells (soluble TN). The removal of this very basic sequence might be a crucial step in the process of TN assembly in the extracellular matrix. We are currently conducting experiments to confirm this possibility. As in the case of chicken TN a secondary structure prediction of the sequence from residue 119-147 shows a high probability to form α -helix. Thus, the primary structure of human TN is compatible with the model of the TN hexamer suggested by Spring et al. (7).

About 2/3 of the primary structure, deduced from the sequence of cDNA clones, have already been reported by Gulcher et al. (37), more specifically from base 1342 to base 6217 of our sequence reported in Fig. 1. With respect to this sequence we found only few differences (see Legend of Fig. 1) with the exception of residues 1599-1607 where we have the following sequence: SGFTQGHQT while in (37) LWLHPRASN (residue 1170 to 1178 of their sequence) has been reported. This difference is due to the fact that we found CT at bases 4848-4849 instead of CCT thus, this single base insertion introduces a frameshift. Furthermore, we found CCA at bases 4875-77 instead of CA reestablishing the frame of the TN sequence previously reported (37). Since the double C reported by Gulcher et al. (37) at bases 4848-4849 should generate a cleaving site for the restriction enzyme Ava II we have tested an appropriate cDNA clone with this enzyme. The absence of this cleaving site confirms our sequence.

In chicken TN there are three FN-L repeats that undergo alternative splicing which generates three different TN subunit isoforms (6-7). It has been reported that in human TN seven FN-L repeats undergo alternative splicing since two different cDNA clones, one with all these seven FN-L repeats and the other one in which all seven were missing, have been isolated (37). From the cDNA libraries tested we have found many cDNA clones covering the human TN area which were obviously derived from alternative splicing of pre-mRNA. However, the smallest clone we found contains two repeats of the seven which undergo alternative splicing (repeats 10 and 12). We also found a clone in which only repeat 11 was missing and clones in which all the seven repeats were present. We did not find any clones in which all the seven repeats were missing. We have also studied, by PCR amplification on cDNA obtained from total RNA from the melanoma cell line SK-MEL-28, the possible different splicing patterns of the TN primary transcript. As reported in 'Results' we observed 4 different bands, one containing all the seven repeats, one only repeats 10 and 12, one containing only repeat 12 and one in which only repeat 11 was missing. However, the above reported results do not give any quantitative estimation on the relative abundance of the different isoforms of TN since the efficiency of amplification of the various cDNA species may be different. We are currently studying the relative abundance of TN isoforms in different tissues by RNAase protection analysis.

By homology comparison between the seven FN-like repeats involved in alternative splicing in human and three FN-L repeats involved in alternative splicing in chicken, it has been reported that human repeats 6,7,8,9 have an extremely high homology with repeat 6 in chicken, thus supporting a recent reduplication of this sequence (38). In fact, our results have shown that human repeats 6 to 9 (which should correspond to the chicken repeat 6) have up till now been observed to always be included or excised as a block. Furthermore, there is a high homology of chicken FN-Like repeat 7 to human FN-L repeat 10, and a high homology of the chicken FN-L repeat 8 to human FN-L repeat 12. However, there are no chicken repeats with high homology to the human FN-L repeat 11. Considering that the cDNA clones containing repeat 11 are the rarest, in our study at least, a chicken FN-L repeat corresponding to the human FN-L repeat 11 might have escaped previous analysis (6-7). Furthermore, repeat 11 has been, up till now, found only in cDNA clones containing all other repeats undergoing alternative splicing and only in a fetal brain cDNA library. These data indicate that the splicing patterns of human TN are quite complex and require further investigation.

At present possible biological functions of different TN isoforms may only be a matter of speculation. However, it is worth noting the high concentration of potential N-glycosylation sites in the repeats which may undergo alternative splicing. Furthermore, Prieto et al. (39) reported spatio-temporal changes during chicken development of two forms of TN mRNA. We are presently approaching the study of the biological functions of different TN isoforms by 1) preparing Mabs able to recognize different TN isoforms in order to perform a detailed analysis of the distribution of various TN in different tissues and in different physiological conditions and 2) establishing procedures for the production and/or purification of different TN isoforms.

ACKNOWLEDGEMENTS

This study has been partially supported by AIRC funds. We thank Miss Antonella Gessaga for skillful secretarial assistance and Mr.Thomas Wiley for manuscript revision. We furthermore thank Miss Monica Peruzzi and Mr.Germano Querzé for their technical assistance. We are indebted to Professor Leonardo Santi for his support and encouragement.

REFERENCES

- 1. Erickson, H.P. and Lightner, V.A. (1988) Advanc. Cell Biol., 2, 55-90.
- 2. Natali, P.G. and Zardi, L. (1989) Int. J. Cancer, Suppl. 4, 66-68.
- 3. Erickson, H.P. and Bourdon, M.A. (1989) Annu. Rev. Cell Biol., 5, 71-92.
- 4. Chiquet-Ehrismann, R. (1990) FASEB J., 4, 2598-2604.
- 5. Chiquet, M. (1989) Dev. Neurosci., 11, 266-275.
- Jones, F.S., Hoffman S., Cunnigham, B.A. and Edelman, G.M. (1989) Proc. Natl. Acad. Sci. USA., 86, 1905–1909.
- Spring, J., Beck, K. and Chiquet-Ehrismann, R., (1989) *Cell*, **59**, 325-334.
 Chiquet-Ehrismann, R., Mackie, E.J., Pearson, C.A. and Sakakura, T. (1986)
- Cell, 47, 131–139.
- Aufderheide, E., Chiquet-Ehrismann, R. and Ekblom, P. (1987) J. Cell Biol., 105, 599-608.
- 10. Aufderheide, E. and Ekblom, P. (1988) J. Cell Biol., 107, 2341-2349.
- Inaguma, Y., Kusakabe, M., Mackie, E.J., Pearson, C.A., Chiquet-Ehrismann, R. and Sakakura, T. (1988) Dev. Biol., 128, 245-255.
- Grumet, M., Hoffman, S., Crossin, K.L. and Edelman, G.M. (1985) Proc. Natl. Acad. Sci. USA, 82, 8075-8079.
- Crossin, K.L., Hoffman, S., Grumet, M., Thiery, J.P. and Edelman, G.M. (1986) J. Cell Biol., 102, 1917–1930.
- Mackie, E.J., Chiquet-Ehrismann, R., Pearson, C.A., Inaguma, Y., Taya, K., (1987) Proc. Natl. Acad. Sci. USA, 84, 4621-4625.
- Natali, P.G., Nicotra, M.R., Bartolazzi, A., Coscia, N., Di Filippo, F., Bigotti, A. and Zardi, L., (1990) Int. J. Cancer, 46, 586-590.
- Natali, P.G., Nicotra, M.R., Bigotti, A., Botti, C., Castellani, P., Risso, A.M. and Zardi L., Int. J. Cancer (in press).
- 17. Nicolò, G., Salvi, S., Olivieri, G., Borsi, L., Castellani, P. and Zardi L., *Cell Diff. and Develop.* (in press).
- 18. Stamp, G.W.H., (1989) J. Pathol., 159, 225-229.
- Vollmer, G. Siegal, G.P., Chiquet-Ehrismann, R., Lightner, V.A., Arnholdt, H. and Knuppen, R. (1990) Lab. Invest., 62 (6), 725-730.
- Anbazhagan, R., Sakakura, T., Gusterson , B.A. (1990) Virchows Arch., 59 (1), 59-63.

- Garin-Chesa, P., Old, L.J. and Retting, W.J. (1990) Proc. Natl. Acad. Sci, USA, 87, 7235-7239.
- Bourdon, M.A., Wikstrand, C.J., Furthmayer, H., Matthews, T.J. and Bigner, D.D. (1983) *Cancer Res.*, 43, 2796–2805.
- Chiquet-Ehrismann, R., Kalla, P. and Pearson, C.A. (1989) Cancer Res. 49, 4322-4325.
- Pearson, C.A., Pearson, D., Shibahara, S., Hofsteenge, J. and Chiquet-Ehrismann, R. (1988) EMBO J., 7, 2677-2681.
- 25. Chiquet, M. and Fambrough, D.M.(1984) J. Cell Biol., 98, 1937-1946
- Kruse, J., Keilhauer, G., Faissner, A., Timpl, R. and Schachner, M. (1985) *Nature*, **316**, 146-148.
- Gulcher, J.R., Marton, L.S. and Stefansson, K.(1986) Proc. Natl. Acad. Sci. USA, 83, 2118-2122.
- 28. Erickson, H.P. and Iglesias, J.L. (1984), Nature, 311, 267-269.
- Chirgwin, J.M., Przybyla, A.E., Mac Donald, R.J. and Rutter, W.J., (1979) Biochemistry, 18, 5294-5299.
- Sambrook, J., Fritsch, E.F. and Maniatis, T., (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor University Press, II Edition.
- 31. Canfield, R.E. and Anfinsen, C.B. (1963) J. Biol. Chem., 238, 2684-2690.
- 32. Carnemolla, B., Balza, E., Siri, A., Zardi, L., Nicotra, M.R., Bigotti, A. and Natali, P.G., (1989) J. Cell Biol., 108, 1139-1148.
- 33. Laemmli, U.K. (1970) Nature, 227, 680-685.
- Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. USA. 76, 4350-4354.
- 35. Von Heijne, G. (1986) Nucleic Acid Res., 14, 4683-4690.
- 36. Kreil, G. (1981) Ann. Rev. Biochem., 50, 317-348.
- Gulcher, J.R., Nies, D.E., Marton, L.S. and Stefansson, K. (1989) Proc. Natl. Acad. Sci. USA., 86, 1588-1592.
- Gulcher, J.R., Alexakcs, M.J., Le Beau, M.M., Lemons, R.S. and Stefansson, K., (1990), Genomics, 6, 616-622.
- Prieto, A.L., Jones, F.S., Cunningham, B.A., Crossin, K.L. and Edelman, G.M. (1990), J. Cell Biol., 111, 685-698.