Complete Thrombospondin mRNA Sequence Includes Potential Regulatory Sites in the 3' Untranslated Region

Sally W. Hennessy, Beth A. Frazier, David D. Kim, Thomas L. Deckwerth, D. Mona Baumgartel, Peter Rotwein, and William A. Frazier

Departments of Biological Chemistry, Medicine and Genetics, Division of Biology and Biomedical Sciences, Washington University School of Medicine, St. Louis, Missouri 63110

Abstract. The nucleotide sequence of human thrombospondin (TS) mRNA has been determined from human fibroblast and endothelial cDNAs. The sequence of 5802 bp begins 110 bp upstream from the initiator codon and includes the entire 3' untranslated region (UTR) of the mRNA. The coding region (3510 bp) specifies a protein of 1170 amino acids with all of the known features of the TS subunit (Frazier, W. A. 1987. J. Cell Biol. 105:625-632). The long 3' UTR of 2166 nucleotides is extremely A/T-rich, particularly in the latter half. It contains 37 TATT or ATTT(A) sequences that have been suggested as mediators of the stability of mRNAs for cytokines, lymphokines, and oncogenes (Shaw, G., and R. Kamen. 1986. Cell. 46: 659-667). Another unusual feature of the 3' UTR of TS mRNA is a stretch of 42 nucleotides of which 40 are thymidines (uridine in the mRNA) including an uninterrupted sequence of 26 thymidines. This region is flanked by two sets of direct repeats suggesting that it may be an insertion element of retrotranscriptional origin. Comparison of the 3' untranslated region of TS mRNA with the GenBank data base indicates the greatest degree of similarity with an α-interferon gene which contains a number of the TATT/ATTT consensus sites. The degree of similarity between the TS and interferon sequences is the same in regions of the interferon gene corresponding to its coding and non-coding regions suggesting that most of the TS 3' UTR may be derived from an interferon gene or pseudogene. The features of the TS mRNA 3' UTR provide a potential explanation for the rapid regulation of TS message observed in cultured cells in response to PDGF and suggest that TS is a member of a group of proteins which are intimately involved in the control of cell growth and differentiation.

Thrombospondin (TS) is a modular or multidomain adhesive glycoprotein found in platelet alpha granules and involved in platelet aggregation (9, 14). The native protein is a trimer of identical subunits that consist of a globular NH₂-terminal heparin-binding domain of ~25 kD, a thin connecting strand region of ~50 kD which contains the interchain disulfide bonds, and a COOH-terminal globular domain which consists of a calcium-binding region (~30 kD) and a cell- or platelet-binding domain of ~23 kD (14, 20). In addition to heparin and similar glycosaminoglycans, TS binds collagens (15), the adhesive glycoproteins fibronectin, laminin, and fibrinogen (21), as well as plasminogen and histidine-rich glycoprotein (36). All of these interactions are thought to be mediated by the central connecting strand region of the subunit (21). The calcium-binding region of TS seems to form an extensible hinge so that in the calcium-replete form this region appears globular, while after chelating calcium, it appears as an extension of the thinner connecting strand (10, 14). The COOH-terminal cell- and platelet-binding domain of TS probably binds to a membrane glycoprotein receptor on platelets and on other cells which adhere to TS (1, 14).

In addition to its presence in platelets, TS is widely distributed. Localization of TS in human tissues using immunofluorescence methods demonstrates that the protein is present in blood vessel, skin, muscle, and glandular epithelium (43). TS is synthesized by a variety of cell types in culture where its production is inversely related to the degree of confluence of the culture (27). An interesting feature of TS regulation is its relatively rapid rate of synthesis and degradation (22), suggesting an important regulatory function for the protein. Indeed, it has recently been reported that TS mRNA (18, 23) and protein (22) are rapidly induced by PDGF in vascular smooth muscle cells and that TS mRNA appears with a time course comparable to that of protooncogenes c-myc and c-fos mRNA in these cells (17, 23, 45). Furthermore, TS mRNA levels in cells are superinduced by PDGF treatment in the presence of protein synthesis inhibitors, such as cycloheximide (23). This phenomenon also has

© The Rockefeller University Press, 0021-9525/89/02/729/8 $2.00
The Journal of Cell Biology, Volume 108, February 1989 729-736
been reported for a number of other PDGF responsive genes, including protooncogenes (45).

To provide a more detailed understanding of the TS subunit structure and to begin investigation of the structural basis for the regulation of TS expression, we undertook the cloning of TS cDNA from human fibroblasts and endothelial cells. We initially reported amino acid sequences from the amino termini of the major proteolytic fragments or domains of TS that allowed the ordering of these domains within the TS polypeptide chain (14). Using amino acid sequence data obtained on tryptic peptides of TS, we isolated and sequenced a human fibroblast cDNA clone designated 6A (Fig. 1) encoding the signal peptide, heparin binding domain, and part of the collagen binding region of TS (11). This sequence was verified by the presence of the known NH₂-terminal amino acid sequence of the mature TS subunit as well as five additional peptide sequences that we had determined. Lawler and Hynes (20) later reported the sequence of human endothelial cDNA encoding the TS subunit coding region and a portion of the 3' untranslated region (UTR), and Kobiyashi et al. (18) reported the sequence of human endothelial cDNA encoding the first one-third of the TS peptide chain. We report here the isolation and characterization of two cDNA clones, one of which encodes the entire mature TS subunit and a portion of the 3' UTR of the mRNA while the other overlapping clone extends to the poly A tail of the mRNA. The sum of this information provides a detailed picture of the structure of the TS subunit and suggests a potential explanation for the regulatory properties of TS mRNA noted above.

**Materials and Methods**

Materials and procedures used in DNA cloning and Sanger dideoxynucleotide sequencing were generally those reported previously (11). Clone S10 was isolated from a human fibroblast cDNA library in pUC 19, the construction of which has been described (11). In sequencing clone S10, we employed synthetic oligodeoxynucleotide primers in conjunction with single- and double-stranded templates. The Klbow fragment of DNA polymerase and AMV reverse transcriptase were purchased from New England Biolabs (Boston, MA) and Bethesda Research Laboratories (Gaithersburg, MD). Sequenase T7 polymerase was from United States Biochemicals Corp. (Cleveland, OH). To provide internal starting sites for sequencing, the Pvu II fragments of S10 (see Fig. 1) were used as sequencing primers. Before loading the reactions on gels, each reaction was cut with Pvu II to release the extended labeled oligonucleotides from the restriction fragment primers. Clone TXE was isolated from a Agtl I cDNA library of human umbilical vein endothelial cells prepared by Dr. Eric Sadler of this institution, and overlaps the 3' end of S10 by ~450 nucleotides. TXE was identified by hybridization with labeled S10 insert. The Xba I/Eco RI fragment of the insert was cloned into Bluescript (Stratagene, San Diego, CA), and sequencing was performed using single-stranded templates prepared by superinfection with R408 helper phage (Stratagene) and synthetic primers. Data handling and analysis of sequence data were performed using a MICROVAX II computer system and the programs of Compugene (St. Louis, MO) and the FASTN program.

RNA was extracted from human JY fibroblasts (8), both from untreated cells grown in medium containing 15% FBS and from cells maintained in low serum for 3 d and then treated with PDGF (10 ng/ml) for 4 h or with cycloheximide (10 μg/ml) for 30 min (or with both by adding cycloheximide 30 min before incubation with PDGF). The RNA was separated on a 0.8% agarose/2.2 M formaldehyde gel and was transferred to a Nitroplus 2,000 membrane (Micron Separations, Inc., Westborough, MA) and hybridized and washed as described (26). Probes were radioactively labeled with 32P using the random hexanucleotide primer method (13). A construct designated 6S was made by cloning the Sac I/Nsi I fragment of the 6A insert and the Nsi I/Xba I fragment of S10 into the Bluescript vector. This insert contains the entire TS coding region plus 120 nucleotides 5' to the ATG start codon. Clone 6SXE was constructed by ligating the insert of TXE excised with Xba I and Nci I into Bluescript clone 6S cut with the same enzymes. This clone extends 6S through the poly A tail. In vitro transcripts were made from 6S and TXE using phage T3 RNA polymerase, and this RNA was included on the RNA gels as a standard.

**Results and Discussion**

We initially reported the isolation of a cDNA clone designated 6A (see Fig. 1) encoding the NH₂-terminal heparin-binding domain of TS (11). Using this cDNA as a probe, we have isolated from the same human fibroblast library a cDNA clone containing an open reading frame that encodes the entire sequence of the TS subunit plus ~800 nucleotides of the 3' untranslated region of its mRNA (clone S10 in Fig. 1). In an effort to clone the entire 3' UTR of the mRNA, we also screened a human endothelial cell cDNA library constructed in Agtl I and recovered a clone extending from within the 3' end of S10 to a 65 nucleotide poly A sequence (clone TXE; Fig. 1). Also indicated in Fig. 1 is a schematic diagram of the TS protein subunit based on the translated amino acid sequence shown in Fig. 2 in the single letter code. The amino acid sequence and its relationship to the domain structure of TS along with implied evolutionary relationships with other proteins have been discussed elsewhere (14, 20). The TS peptide chain consists of eight discrete regions of protein structure, many of which show relatedness to previously sequenced proteins. This suggests that the TS gene has been assembled from a variety of exons “borrowed” from a number of ancestral genes such as those which gave rise to the present day genes for type I collagen, malarial cell surface proteins, the complement protein properdin, epidermal growth factor, and calcium-binding proteins (14, 20).

Fig. 2 shows the nucleotide sequence derived from clones 6A, S10, and TXE. The initiation codon is embedded in a perfect match of the Kozak consensus sequence (11) (boxed in Fig. 2). The open reading frame is 3,510 nucleotides long and is preceded by 126 nt of 5' untranslated sequence. It is followed by another 2,101 nucleotides of 3' untranslated sequence and a poly A tract of 65 nucleotides. This yields a total of 5,802 nucleotides, of which 12–15 at the 5' end of the sequence represent a homopolymeric C tail added during the construction of the library. This value of 5.8 kb is in good agreement with estimates of 5.5 (23) to 6.0 (18, 20, 29) kb for the mRNA size obtained from Northern blots (Fig. 3). Hence, it is likely that there are very few 5' untranslated nucleotides preceding the ~120 nucleotides which are represented in clone 6A.

Recently TS has been found to be a member of a class of growth-related proteins whose mRNAs respond rapidly to induction by growth factors, specifically PDGF (17, 18, 23, 45). TS mRNA levels in rat aortic smooth muscle cells are inversely related to the degree of confluency of the culture, being highest in sparse cell populations (23). Treatment of quiescent cells with either serum or PDGF results in the rapid induction of TS mRNA with a time course similar to that of the induction of c-myc mRNA in the same cells (17, 18, 23). Further, inhibition of protein synthesis with cycloheximide during induction causes a larger increase in the amount of TS mRNA than found with PDGF alone (23). This behavior is characteristic of mRNAs for several proteins which respond to PDGF induction (17, 45), as well as several cytokines and lymphokines such as interleukins, interferons,
tumor necrosis factor, colony stimulating factors and oncogenes such as myc, fos, and myb (17, 32, 35, 45). In addition to rapid induction, this group of mRNAs also undergoes rapid degradation, a property thought to be mediated by A/T-rich sequences within the 3' UTR of each mRNA. These sequences, identified as ATTT(A) (35), TATT (32), and TTA-TTTAT (5) by different workers, occur multiple times in the 3' UTR of TS mRNA, being grouped within the latter half of the 3' UTR (Figs. 1 and 2). The oligomers TATT and ATTT(A) occur 22 and 27 times, respectively, within the 3' UTR, while the longer octamer sequence TTA-TTTAT appears only once as it does in many of the cytokine and lymphokine cDNAs (5).

The TS sequence that we have characterized is isolated from human fibroblast and endothelial cell cDNA libraries, while most of the data regarding TS induction and turnover has been obtained in smooth muscle cells. Yet it seems likely that the structure of TS mRNA is the same in all cells thus far examined, since the size of the major mRNA identified on Northern blots is identical in rat smooth muscle cells and human fibroblasts (references 23, 29 and Fig. 3). To ascertain that the human fibroblast mRNA contained the same 3' UTR as the human endothelial cells, a Northern blot of JY fibroblast RNAs obtained from cultures grown to different cell densities and with or without PDGF treatment was probed with the insert derived from clone S10, which contains the full length coding region for TS. This established that human fibroblasts do indeed contain a predominant mRNA species of 6.0 kb in length containing the TS coding sequences (Fig. 3). This blot was then stripped at high temperature and reprobed with the insert of clone TXE which had been derived from the human endothelial library and contains only the extreme 3' UTR region. As seen in Fig. 3A, the same 6 kb band is identified by both probes in all cases. Lane S contains the in vitro transcript from the insert of 6S and provides a control for the stripping procedure since the 6S transcript (i.e., the coding region) does not overlap with TXE (i.e., the 3' UTR) used to reprobe the blot. In several experiments, a smaller mRNA band of ~4.0–4.5 kb was visualized in the human fibroblasts after hybridization with the 6S coding region probe (Fig. 3, A and B). Similar results were seen with RNA extracted from HEL cells, HT1080 cells, and rat smooth muscle cells (not shown), suggesting that the smaller mRNA species may represent a specific processing or degradation product. This appears to be the case for JY cells since the smaller RNA band is not detected with a probe prepared from the insert of TXE which represents the most 3' portion of the mRNA (Fig. 3, A and B). Only the larger RNA hybridizes with these 3' sequences. Given the potential regulatory function of the latter half of the 3' UTR of TS mRNA, the finding that cells contain a form of the mRNA missing this region is of interest. At present the mode of generation of the smaller form of TS mRNA is not known.

Another unusual feature of the 3' UTR of TS mRNA is the occurrence of a string of 42 nucleotides of which 40 are T residues (U in the RNA) including a homopolymer run of 26 T residues. This poly T run ends 163 nucleotides from the beginning of the poly A tail. The region bracketed by the poly T and poly A tracts could not be sequenced until part of the poly A tail was deleted resulting in disruption of the base pairing between these two homopolymeric regions in the single-stranded template DNA. TS is the only 12th example of an mRNA containing a poly U tract of 10 nucleotides or longer in the 3' UTR. Other mRNAs with poly U tracts in their 3' UTRs include those for the human insulin receptor (42), chick procollagen type II (34), bovine acetylcholine receptor β-chain (39), mouse urokinase plasminogen activator (2), human β-actin (40), chick ovalbumin Y gene (16), human tissue factor (37), and human cap-binding protein (33). In none of these cases is there a known function for the long poly U tract in the 3' UTR of the TS mRNA is also flanked by two sets of direct repeats. It was recently reported that the mRNA for laminin B2 chain contains a similar sequence element in the 3' UTR of the message (12). Interestingly, this laminin mRNA contains 22 of the TATT/ATT(A) sequence elements and one octamer sequence in its 3' UTR and these are clustered in its latter half. One explanation for these long poly U tracts in UTRs of mRNAs is that the gene has had a retroprotein inserted in the antisense orientation such that its poly A tail is transcribed in the sense strand as a poly T tract, or a poly U tract in the mRNA. This may explain the occurrence of the poly T tract in the TS gene since it is followed after 18 nucleotides by a TTTACT sequence whose complement reads AGTTAA, a 5/6 match with the polyadenylation signal consensus AATAAA.

Whatever the origin of the long poly U tract in the mRNA, it is quite clear that it can hybridize with the poly A tract forming a loop of some 163 nucleotides. Perhaps this duplex region plays some role in stabilization of the TS mRNA.
| Sequence | Start | End |
|----------|-------|-----|
| GAACACAAATTCCGGATCACGTGGACTCTGACTCGACACCAGCTATTGGGATACTCTGACACCAACAAATCAGGATATTTGAGAATGGGCCACGAAAGAATCTGGACACCTACGTCCCTATGTCGCC | 1 | 2760 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 2761 | 2880 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 2881 | 3000 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3001 | 3120 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3121 | 3240 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3241 | 3360 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3361 | 3480 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3481 | 3600 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3601 | 3720 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3721 | 3840 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3841 | 3960 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 3961 | 4080 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 4081 | 4200 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 4201 | 4320 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 4321 | 4440 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 4441 | 4560 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 4561 | 4680 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 4681 | 4800 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 4801 | 4920 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 4921 | 5040 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 5041 | 5160 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 5161 | 5280 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 5281 | 5400 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 5401 | 5520 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 5521 | 5640 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 5641 | 5760 |
| EHNPDQLDSDSDRIGDTODNNQDIDEDGHQNNLDDNPYYP | 5761 | 5801 |

---

| Chromosome | Length |
|------------|--------|
| 1          | 2760   |
| 2          | 2880   |
| 3          | 3000   |
| 4          | 3120   |
| 5          | 3240   |
| 6          | 3360   |
| 7          | 3480   |
| 8          | 3600   |
| 9          | 3720   |
| 10         | 3840   |
| 11         | 3960   |
| 12         | 4080   |
| 13         | 4200   |
| 14         | 4320   |
| 15         | 4440   |
| 16         | 4560   |
| 17         | 4680   |
| 18         | 4800   |
| 19         | 4920   |
| 20         | 5040   |
| 21         | 5160   |
| 22         | 5280   |
| 23         | 5400   |
| 24         | 5520   |
| 25         | 5640   |
| 26         | 5760   |
| 27         | 5800   |

---

| Gene | Description |
|------|-------------|
| 1    | Gene 1      |
| 2    | Gene 2      |
| 3    | Gene 3      |
| 4    | Gene 4      |
| 5    | Gene 5      |
| 6    | Gene 6      |
| 7    | Gene 7      |
| 8    | Gene 8      |
| 9    | Gene 9      |
| 10   | Gene 10     |
| 11   | Gene 11     |
| 12   | Gene 12     |
| 13   | Gene 13     |
| 14   | Gene 14     |
| 15   | Gene 15     |
| 16   | Gene 16     |
| 17   | Gene 17     |
| 18   | Gene 18     |
| 19   | Gene 19     |
| 20   | Gene 20     |
| 21   | Gene 21     |
| 22   | Gene 22     |
| 23   | Gene 23     |
| 24   | Gene 24     |
| 25   | Gene 25     |
| 26   | Gene 26     |
| 27   | Gene 27     |

---

| Protein | Description |
|---------|-------------|
| 1       | Protein 1   |
| 2       | Protein 2   |
| 3       | Protein 3   |
| 4       | Protein 4   |
| 5       | Protein 5   |
| 6       | Protein 6   |
| 7       | Protein 7   |
| 8       | Protein 8   |
| 9       | Protein 9   |
| 10      | Protein 10  |
| 11      | Protein 11  |
| 12      | Protein 12  |
| 13      | Protein 13  |
| 14      | Protein 14  |
| 15      | Protein 15  |
| 16      | Protein 16  |
| 17      | Protein 17  |
| 18      | Protein 18  |
| 19      | Protein 19  |
| 20      | Protein 20  |
| 21      | Protein 21  |
| 22      | Protein 22  |
| 23      | Protein 23  |
| 24      | Protein 24  |
| 25      | Protein 25  |
| 26      | Protein 26  |
| 27      | Protein 27  |

---

| mRNA | Description |
|------|-------------|
| 1    | mRNA 1      |
| 2    | mRNA 2      |
| 3    | mRNA 3      |
| 4    | mRNA 4      |
| 5    | mRNA 5      |
| 6    | mRNA 6      |
| 7    | mRNA 7      |
| 8    | mRNA 8      |
| 9    | mRNA 9      |
| 10   | mRNA 10     |
| 11   | mRNA 11     |
| 12   | mRNA 12     |
| 13   | mRNA 13     |
| 14   | mRNA 14     |
| 15   | mRNA 15     |
| 16   | mRNA 16     |
| 17   | mRNA 17     |
| 18   | mRNA 18     |
| 19   | mRNA 19     |
| 20   | mRNA 20     |
| 21   | mRNA 21     |
| 22   | mRNA 22     |
| 23   | mRNA 23     |
| 24   | mRNA 24     |
| 25   | mRNA 25     |
| 26   | mRNA 26     |
| 27   | mRNA 27     |

---

| Peptide | Description |
|---------|-------------|
| 1       | Peptide 1   |
| 2       | Peptide 2   |
| 3       | Peptide 3   |
| 4       | Peptide 4   |
| 5       | Peptide 5   |
| 6       | Peptide 6   |
| 7       | Peptide 7   |
| 8       | Peptide 8   |
| 9       | Peptide 9   |
| 10      | Peptide 10  |
| 11      | Peptide 11  |
| 12      | Peptide 12  |
| 13      | Peptide 13  |
| 14      | Peptide 14  |
| 15      | Peptide 15  |
| 16      | Peptide 16  |
| 17      | Peptide 17  |
| 18      | Peptide 18  |
| 19      | Peptide 19  |
| 20      | Peptide 20  |
| 21      | Peptide 21  |
| 22      | Peptide 22  |
| 23      | Peptide 23  |
| 24      | Peptide 24  |
| 25      | Peptide 25  |
| 26      | Peptide 26  |
| 27      | Peptide 27  |
against 3' exonucleolytic attack as has been proposed for bacterial and histone mRNAs (4, 30). The loop in the TS mRNA would contain seven of the TATT/ATTT sequences. If these are in fact sites for single-stranded endonuclease attack, the subsequent exonuclease digestion of the nicked mRNA would be limited by the duplex region just upstream. Thus, the combination of a large stem-loop structure containing endonuclease recognition sites could be a way of limiting or coordinating two types of control mechanisms operating on mRNA stability. It is becoming increasingly clear that the 3' UTRs of messages can play an important role in regulation of the expression of the proteins which the messages encode. For example, iron regulation of expression of the transferrin receptor is mediated by a specific stem-loop structure in the 3' UTR of that mRNA (6).

A search of the GenBank nucleic acid database (3) (release

Figure 3. Northern blots of JY human fibroblast RNA preparations probed with labeled insert from clone 6S (left) and the same blots stripped and reprobed with the labeled insert of clone TXE (right). (A) Lane 1 contains RNA from cells grown normally. Lanes 2-5 contain RNA from cells maintained in low serum before treatment as follows: lane 2, none (quiescent control); lane 3, plus cycloheximide; lane 4, plus PDGF; lane 5, plus both PDGF and cycloheximide (see Materials and Methods for details). Lanes 1 and 3-5 contain 10 μg RNA, and lane 2 contains 4 μg. Lane S is the in vitro transcript from S10 included as a stripping control. It is not visualized after probing with TXE sequences. (B) Lanes 1-3 represent three different preparations of RNA from JY cells grown normally. Lanes S1 and S2 contain the 6S and TXE in vitro transcripts, respectively. All lanes contain 10 μg RNA.
52.0) with the program FASTN using either the entire 3' UTR of the TS cDNA sequence or the portion contained within clone TXE (Fig. 1) revealed that the most related sequences in the primate data base were found in the human leukocyte α-interferons which displayed a level of 40–45% identity over the entire length of the TS 3' UTR. The interferons are a multigene family, and the α-interferon region having the most similarity to the TS 3' UTR sequence was region L (41). The interferon gene family includes a number of pseudogenes (41). The rather evenly distributed relatedness between the 3' UTR of TS mRNA and the α-interferon genes suggests that the TS 3' UTR may have resulted from the incorporation of a complete interferon-like pseudogene into the TS gene. This mode of evolution is certainly consistent with the modular assembly of the TS coding region which appears to be an excellent example of exon shuffling (14, 20).

When 1,756 nucleotides of α-interferon gene region L were compared with 100 randomized sequences generated from 1,748 nucleotides of the TS 3' UTR, the mean aligned score was 325 while the aligned score for the bona fide TS 3' UTR sequence was 536, or 2.42 standard deviations higher. The most A/T-rich region of the TS 3' UTR aligned with the A/T-rich part of the α-interferon 3' UTR. This is precisely the region of the interferon mRNA which is thought to contain sequences that specify a rapid turnover for this and other mRNAs. A high degree of similarity of the A/T-rich region of the TS 3' UTR was also found with the 3' UTRs of the human β- and γ-interferons, interleukin-2, and B-cell leukemia (lymphoma) protooncogene mRNAs. When only the 280 nucleotides representing the most A/T-rich segment of the TS 3' UTR were used to probe the GenBank database, the α-interferons were again identified as the best match. Thus, the 3' UTR of TS mRNA contains multiple copies of A/T-rich sequences which have been suggested by others to specify rapid turnover. It is thought that these sites provide targets or recognition sequences for extremely labile endonucleases, thus explaining the effects of inhibitors of protein synthesis to stabilize these mRNAs (5, 32, 35).

Interestingly, one of the other mRNAs that displayed the highest level of relatedness to the TS 3' UTR was the human vitronectin receptor α-subunit mRNA (38). Closer inspection of the alignment of these two RNA sequences revealed that, like TS mRNA, vitronectin receptor mRNA contains a large 3' UTR which is extremely rich in the putative target sequences TATT and ATTT(A), having 51 of these sites in the portion of the 3' UTR that has been sequenced. Thus we suggest that the vitronectin receptor mRNA may be regulated in a manner similar to the TS and cytokine mRNAs.

The fact that TS mRNA responds rapidly to PDGF induction and has a very rapid turnover rate suggests that TS is critically involved in the growth response of cells that produce it. It has been reported that TS appears in the extracellular matrix of only proliferating cells and that providing a sink for TS in the form of a monoclonal antibody slows the cells' growth rate (24). The well-known inhibitory effect of heparin on the growth of smooth muscle cells (7) could also be explained by an inhibition of TS association with these cells. TS binds to a variety of cell types by association of its NH2-terminal heparin-binding domain with cell surface heparin-like glycosaminoglycans (28). Taken together, these observations suggest that TS may be an essential component of an autocrine stimulation pathway triggered by growth factors such as PDGF. Thus, it is not inconceivable that TS could act as an oncogene under the proper circumstances. Abnormally high or prolonged production of TS triggered by mutation could lead to a proliferative response that is no longer regulated. In this regard, it is interesting that deletion of 67 bp of the A/T-rich region of the 3' UTR of the c-fos gene converts it into a transforming gene (25). It was suggested that this effect was due to a stabilization of the fos mRNA and a concomitant increase in the cellular content of the protein (31). The reverse effect on mRNA stability has been observed by grafting the homologous 3' UTR region from the granulocyte–monocyte colony stimulating factor (GM-CSF) mRNA onto the normally quite stable β-globin mRNA (35). The possibility that cells also contain a significant amount of TS mRNA apparently lacking the A/T-rich sequences implicated in the regulation of RNA stability may mean that, in some cases, cells normally produce a highly unstable, and hence rapidly regulated form, and a more stable form of a message at the same time. Experiments to define the potential role of the 3' UTR in determining the stability of TS mRNA are in progress.

We thank Karen Pollock and Sheela Mohan for helpful advice and assistance; Drs. Jeffrey Gordon and Marilyn Kozak for helpful discussions; and Dr. Gregory Grant and Mark Frazier of the Washington University Protein Chemistry Facility for automated amino acid sequence analysis and oligodeoxyribonucleotide synthesis. We also thank Dr. Mary Edmonds of the University of Pittsburgh for information that she has compiled on poly U tracts in UTRs of mRNAs.

This work was supported by National Institutes of Health grants NS 13269 and HL 1417 and a grant from the Monsanto Co. (St. Louis, MO) to W. A. Frazier. P. Rotwein acknowledges support from National Institutes of Health grants DK 37449 and HD 20805.

Received for publication 17 March 1988 and in revised form 11 October 1988.

Note Added in Proof: Based on recently published TS genomic DNA sequence and a determination of the transcription starting site (Donoviel, D. B., P. Framson, C. E Eldridge, M. Cooke, S. Kobayashi, and P. Bornstein. 1988. J. Biol. Chem. 263:18590–18593), our cDNA clones reported here lack only 66 nucleotides at the 5' end. Thus the total length of the TS mRNA is 5,855 nucleotides.

References

1. Asch, A. S., J. Barnwell, R. L. Silverstein, and R. L. Nachman. 1987. Isolation of the thrombospondin membrane receptor. J. Clin. Invest. 79:1054–1061.
2. Belin, D., J.-D. Vassalli, C. Combépine, F. Goddeau, Y. Nagamine, E. Reich, H. P. Kocher, and R. M. Duvoisin. 1985. Cloning, nucleotide sequencing and expression of cDNAs encoding mouse urokinase-type plasminogen activator. Eur. J. Biochem. 148:225–232.
3. Bilovsky, H. S., C. Burks, J. W. Fickett, W. B. Goad, F. I. Lewitter, W. P. Rindone, C. D. Swindell, and C.-S. Tung. 1986. The GenBank genetic sequence database. Nucleic Acids Res. 14:1–4.
4. Brawerman, G. 1987. Determinants of messenger RNA stability. Cell 48:5–6.
5. Caput, D., B. Beutler, K. Hartog, R. Thayer, S. Brown-Shimer, and A. Cerami. 1986. Identification of a common nucleotide sequence in the 3'-untranslated region of mRNA molecules specifying inflammatory mediators. Proc. Natl. Acad. Sci. USA. 83:1670–1674.
6. Casey, J. L., M. W. Hentze, D. M. Koeller, S. W. Caughman, T. A. Rouzi, R. D. Klausner, and J. B. Hartford. 1988. Iron-responsive elements: regulatory RNA sequences that control mRNA levels and translation. Science (Wash. DC). 240:924–928.
7. Castellot, J. J., Jr., M. L. Addonizio, R. Rosenberg, and M. J. Karnofsky. 1981. Cultured endothelial cells produce a heparitinase inhibitor of smooth muscle cell growth. J. Cell Biol. 90:372–379.
8. Dillman, W. H., A. Barrieux, W. E. Neeley, and P. Contreras. 1983. Influence of thyroid hormone on the in vitro translational activity of specific mRNAs in the rat heart. J. Biol. Chem. 258:7738–7745.

Hennessy et al. Thrombospondin mRNA Sequence
9. Dixit, V. M., D. M. Havercost, 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.

10. 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: characterization and mapping of their epitopes. J. Biol. Chem. 261:1962-1968.

11. Dixit, V. M., S. W. Hennessy, G. A. Grant, P. Rotwein, and W. A. Frazier. 1986. Characterization of a cDNA encoding the heparin and collagen binding domains of human thrombospondin. Proc. Natl. Acad. Sci. USA 83:5449-5453.

12. Durkin, M. E., B. B. Bartos, S.-H. Liu, S. L. Phillips, and A. E. Chung. 1988. Primary structure of the mouse laminin B2 chain and comparison with laminin B1. Biochemistry. 27:5198-5204.

13. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6-13.

14. Frazier, W. A. 1987. Thrombospondin: A modular adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. Annual Rev. Biochem. 56:1205-1230.

15. Galvin, N. J., P. M. Vance, V. M. Dixit, B. Fink, and W. A. Frazier. 1987. Interaction of human thrombospondin with Types I-V collagen: direct binding and electron microscopy. J. Cell Biol. 104:1413-1422.

16. Heilig, R., R. Murauskas, C. Kleeper, and J. L. Mandel. 1982. The ovalbumin gene family: Complete sequence and structure of the Y gene. Nucleic Acids Res. 10:4363-4382.

17. Kindy, M. S., and G. E. Sonenshein. 1985. Regulation of oncogene expression in cultured aortic smooth muscle cells. Post-transcriptional control of c-myc mRNA. J. Biol. Chem. 261:12865-12868.

18. Kobayashi, S., F. Eden-McCutchan, P. Framson, and P. Bornstein. 1986. Partial amino acid sequence of human thrombospondin as determined by analysis of cDNA clones: homology to malarial circumsporozoite proteins. Biochemistry. 25:8418-8425.

19. Kulkosky, J. W., W. M. Wood, and M. Edmonds. 1985. Location of oligo(arginic acid) sequences within messenger ribonuclease acid molecules of HeLa cells. Biochemistry. 24:3678-3686.

20. Lawler, J., and D. Hynes. 1986. The structure of human thrombospondin, an adhesive glycoprotein with multiple calcium-binding sites and homologies with several different proteins. J. Cell Biol. 103:1635-1648.

21. Lawler, J., J. F. Comolly, P. Ferro, and L. H. Derick. 1986. Thrombin and chymotrypsin interactions with thrombospondin. Ann. NY Acad. Sci. 485:273-287.

22. 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.

23. Majack, R. A., J. Mildbrandt, and V. M. Dixit. 1987. Induction of thrombospondin messenger RNA levels occurs as an immediate primary response to platelet-derived growth factor. J. Biol. Chem. 262:8821-8825.

24. Majack, R. A., L. V. Goodman, and V. M. Dixit. 1986. Platelet-derived growth factor beta increases mRNA for platelet proteins both in the presence and in the absence of changes in mRNA stability. Proc. Natl. Acad. Sci. USA 85:1105-1108.

25. Raghow, R. 1987. Regulation of messenger RNA turnover in eukaryotes. Trends Biochem. Sci. 12:358-360.

26. Ramhdorf, H. J. A., Schönhärl, P. Angel, M. Lütthig, U. Rüther, and P. Herrlich. 1987. Posttranscriptional regulation of c-fos mRNA expression. Nucleic Acids Res. 15:1643-1659.

27. Reeves, R. T., S. Elton, M. S. Nissen, D. Lehn, and K. R. Johnson. 1987. Posttranscriptional gene regulation and specific binding of the nonhistone protein HMGI-1 by the 3′ untranslated region of bovine interleukin 2 cDNA. Proc. Natl. Acad. Sci. USA. 84:6531-6535.

28. Rychlik, W., L. L. Domier, P. R. Gardner, G. M. Hellmann, and R. E. Rhoads. 1987. Amino acid sequence of the mRNA cap-binding protein from human tissues. Proc. Natl. Acad. Sci. USA. 84:945-949.

29. Sandell, L. J., H. L. Prestetze, D. Kravis, and W. B. Upholt. 1984. Structure and sequence of the chicken type II procollagen gene. J. Biol. Chem. 259:7826-7834.

30. Shaw, G., and R. Kamen. 1986. A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA mediates selective mRNA degradation. Cell. 46:659-667.

31. Silverstein, R. L., L. L. Keung, P. C. Harpel, and R. L. Nachman. 1985. Platelet thrombospondin forms a trimolecular complex with plasminogen and histidine-rich glycoprotein. J. Clin. Invest. 75:2063-2073.

32. Spicer, E. K., R. Horton, L. Bloem, R. Bach, K. R. Williams, A. Guha, J. Kraus, T.-C. Lin, Y. Nemerson, and W. H. Konigsberg. 1987. Isolation of cDNA clones coding for human tissue factor: Primary structure of the protein and cDNA. Proc. Natl. Acad. Sci. USA. 84:15148-15152.

33. Suzuki, S., W. S. Argraves, R. Pyrela, H. Arai, T. Krussul, M. D. Pierschbacher, and R. Ruoslab. 1986. CDNA and amino acid sequences of the cell adhesion protein receptor recognizing vitronectin reveal a transmembrane domain and homologies with other adhesion protein receptors. Proc. Natl. Acad. Sci. USA. 83:8614-8618.

34. Tanabe, T., M. Noda, Y. Furutani, T. Takai, H. Takahashi, K. Tanaka, T. Hirose, S. Inayama, and S. Numa. 1984. Primary structure of the subunit precursor of calf muscle acetylcholine receptor deduced from cDNA sequence. Eur. J. Biochem. 144:11-17.

35. Tomokuna, K., H. Taniguchi, K. Yoda, M. Shimizu, and S. Sakiyama. 1986. Nucleotide sequence of a full-length cDNA for mouse cytokes skeletal beta-actin mRNA. Nucleic Acids Res. 14:2829.

36. Ullrich, A., J. R. Bell, E. Y. Chen, R. Herrera, L. M. Petruzzelli, T. J. Dull, A. Gray, L. Causens, Y-C. Liu, M. Tsukikawa, A. Mason, P. H. Seeburg, C. Grunfeld, O. M. Rosen, and J. Ramachandran. 1985. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature (Lond.). 313:756-761.

37. Wight, T. N., G. J. Raugi, S. M. Murphy, and P. Bornstein. 1985. Light microscopy immunolocalization of thrombospondin in human tissues. J. Histocom. Cytochem. 33:295-302.

38. Wood, W. M., J. C. Wallace, and M. Edmonds. 1985. Sequence content of oligo(arginic acid)-containing messenger ribonuclease acid stimulate expression of the same genes in 3T3 cells. Cell. 45:793-800.