BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **243,** 284–288 (1998)
ARTICLE NO. RC978003

Molecular Cloning and Expression of a cDNA Encoding a Human Thioredoxin-like Protein

Antonio Miranda-Vizuete, Jan-Åke Gustafsson, and Giannis Spyrou[1]
Department of Biosciences at Novum, Karolinska Institute, S-141 57 Huddinge, Sweden

Received December 15, 1997

This report describes the cloning of a human cDNA that encodes a new protein (Txl, Thioredoxin-like) that belongs to the expanding family of thioredoxins based on sequence comparison of the deduced amino acid sequence. This cDNA, with a total length of 1,278 bp, consists of 205 bp of 5′-untranslated sequence (including an in frame stop codon), an open reading frame of 870 bp and a 203 bp fragment of 3′-untranslated sequence. The coding sequence predicts a protein of 289 amino acids with two distinct domains: an N-terminal domain of 105 residues homologous to the rest of mammalian thioredoxins containing the conserved active site (CGPC) and a C-terminal domain of 184 residues with no homology with any other protein in the database. Northern blot analysis indicates that the txl probe hybridizes to a 1.3 Kb mRNA and is ubiquitously expressed in human tissues with the highest expression in stomach, testis and bone marrow.

Thioredoxins (Trx) are a group of small redox active proteins (M_r 12,000). Thioredoxin was initially discovered in *Escherichia coli* as an electron donor for the essential enzyme ribonucleotide reductase (1) and, since then, many functions have been assigned to thioredoxins not only associated with redox mediated processes but also structural roles (for review see introduction of ref. 2). Mammalian cells have two thioredoxins, one located in the cytoplasm (3) and one in the mitochondria (4) with an active sequence Trp-Cys-Gly-Pro-Cys which is conserved in all thioredoxins (5). The active site of thioredoxin is located in a protrusion of the protein similarly to that of *Escherichia coli* as determined by NMR and X-ray crystallography (6-9). There is an increasing number of proteins with thioredoxin-like domains in which many conserved residues in thioredoxins are present at the same position but active site residues are changed. Thus, in mammalian cells, protein-disulfide isomerase (PDI), calcium binding protein-1 (CaBP1) and phospholipase C-α (PLC-α) have two thioredoxin-like domains but the proline residue at the active site is changed to hystidine or nucleoredoxin, a recently described nuclear thioredoxin-like protein, with a proline residue instead of glycine (10-13). In prokaryotic organisms, *Corynebacterium nephridii* thioredoxin-2 has an alanine residue instead of glycine at the active site and *Bradyrhizobium japonicum* TlpA, a membrane-anchored thioredoxin-like protein has a valine residue instead of glycine (14,15). Thioredoxin reductase is a flavoenzyme that reduces oxidized thioredoxin by using the reducing power of NADPH (the thioredoxin system) (5). Homologous thioredoxin reductases have been reported to be able to activate some of the proteins with thioredoxin-like domains in the insulin assay (13,14,16,17). We describe here the cloning of a tentatively novel member of the thioredoxin-like family (Txl), with an atypical C-terminal domain not previously found in any of the members of this family. However, this protein is not a substrate for the homologous thioredoxin reductase in the insulin assay, typical for thioredoxin activity. Furthermore, Txl mRNA was detected in all the tissues investigated.

MATERIALS AND METHODS

Database search. The GenBank sequence data base at the National Center of Biotechnology Information was accessed via Internet World Wide Web (http://www2.ncbi.nlm.nih.gov/) and the Basic Local Alignment Search Tool (BLAST) (18) was used to identify EST clone-encoded proteins homologous but not identical to the region centered at the active site of human Trx1 (accession number: J04026) (3). Out of many EST hits, only ESTs displaying 50-70% identity were

[1] To whom correspondence should be addressed: Dept. of Biosciences, Center for Biotechnology, Karolinska Institutet, Novum, S-141 57 Huddinge, Sweden. Fax: 46-8-7745538; Email: Giannis.Spyrou@cbt.ki.se

Abbreviations: EST, expressed sequence tag; Trx, thioredoxin; G3PDH, glyceraldehyde-3-phosphate dehydrogenase; PCR, polymerase chain reaction

The sequence data reported in this paper have been submitted to the GenBank Data Libraries under the accession number AF003938.

0006-291X/98 $25.00

selected. Being guided by homology, 53 ESTs were aligned along with human Trx1 and a contig sequence was deduced by merging the overlapping ESTs.

Cloning of human Txl cDNA. Two primers (hTxl-F1, 5′-TCCCCA-TCCCAGCCCGCCCCCTCGGCCGC and hTxl-R1, 5′-AAGCTTGGC-AAAAGTGGTGACTTTTATTTAC) were designed based on the DNA sequence of the EST contig obtained. These primers were used to amplify the full-length cDNA from human testis cDNA (Clontech) with the Expand PCR Long Template System (Boehringer Manheim) as described in the Clontech Manual. The PCR product was cloned into the pGEM-Teasy vector (pGEM-T/hTxl) and sequenced.

Protein expression and purification. The sequence of the open reading frame of hTxl was amplified by PCR from the pGEM-T/hTxl construct by using two mutagenic primers, hTxl-NdeI (5′-CTGCCG-GGCTCTCTGTGCATATGGTGGGGGTGAAG-3′) and hTxl-BamHI (5′-GGTTGTCCAGGATCCTTTGTACCTTAGTGGCTTTCTCC-3′) that introduce a *Nde*I and a *Bam*HI site at the N-terminus and C-terminus of the protein, respectively. The amplified DNA was cloned into the *Nde*I-*Bam*HI sites of the pET-15b expression vector (AMS Biotechnology) and *Escherichia coli* strain BL21(DE3) was transformed with the recombinant plasmid (pET/hTxl). A single positive colony was inoculated in 1 liter of LB medium plus ampicillin and grown at 37 °C until A_{600} = 0.5. Then, fusion protein was induced by addition of 0.5 mM IPTG (isopropyl-1-thio-β-D-galactopyranoside) and growth was continued for another 3.5 h. The cells were harvested by centrifugation at 10,000 × *g* for 10 min and the pellet was resuspended in 50 ml of 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 1 mM phenylmethylsulfonyl fluoride. Lysozyme was added to a final concentration of 0.5 mg/ml with stirring for 30 min on ice. Subsequently, $MgCl_2$ (10 mM), $MnCl_2$ (1 mM), DNase I (10 μg/ml) and RNase (10 μg/ml) were added and the incubation was continued for another 45 min on ice. The cells were sonicated and the supernatant was cleared by centrifugation at 15,000 × *g* for 30 min and loaded onto a Talon Resin Column (Clontech). The His-Txl protein was eluted with 50 mM imidazole, dialyzed against 20 mM Tris-HCl pH 8.0, concentrated with Centricon concentrators (Amicon Inc.) and the size and purity of His-Txl was determined by SDS-PAGE. When indicated the His-tag was removed by thrombin incubation (1U/mg protein). Protein concentration was determined from the absorbance at 280 nm using a molar extinction coefficient of 9,800 M^{-1} cm^{-1}.

The truncated form of human Txl (hΔTxl) lacking the last 184 amino acids at the C-terminal part of the protein was amplified using the primer hΔTxl-BamHI 5′-GCTTCCAGGATCCTTTTCTTAGTG-CTGCTTGATTTTTTC-3′ as reverse primer and hTxl-NdeI as a forward primer. The cloning, overexpression and purification of this truncated form was identical to that described above for the full-length one. Protein concentration was determined from the absorbance at 280 nm using a molar extinction coefficient of 2,920 M^{-1} cm^{-1} for human ΔTxl.

Insulin disulfide reduction assay. The insulin disulfide reduction assay was performed essentially as described elsewhere (19) with a slight modification to activate human Trx1 and Txl by reduction. Aliquots of both proteins were incubated 37°C for 10 min with 1 μl of: 50 mM HEPES pH 7.6, 100 μg/ml BSA and 2 mM DTT. Then, 20 μl of a reaction mixture composed of 200 μl HEPES (1 M), pH 7.6, 40 μl EDTA (0.2 M), 40 μl NADPH (40 mg/ml) and 500 μl of insulin (10 mg/ml) were added plus distilled water to a volume of 55 μl in a 96-wells plate. The reaction started with the addition of 5 μl of calf thymus thioredoxin reductase (50A_{412} unit) (I.M.C.O., Sweden) and incubation was continued for 20 min at 37°C. The reaction was stopped by the addition of 0.5 ml of 6 M guanidine-HCl, 1 mM DTNB, and the absorbance at 412 nm was measured. Human Trx1 was always used as a positive control.

Northern blot analysis. Human multiple tissue Northern blots and human RNA Master Blot with poly(A)$^+$ RNA from different human tissues were purchased from Clontech. Human Txl open reading

```
-205                                                    TCCCCCATCCCAG
-192 CCCGCCCCCTCGGCCGCTCCGCGTCCCTGGCGTGACTCGGCACTGAGAGTCCCGG
-157 GAGAAGACTCGGCGGTCGCACCTCTTCTGTCCAGGCCTCGGCCTTCC TGA GCA
                                                          ***

 -84 TCT CTC CTT CCT CTC CCA GAT CGT CTT CTC CTT CAG TTT CAA

 -42 AGC CAG TGG CGT CGC GGC CAC CCT GCC GGG CTC TCT GTG AGG

   1 ATG GTG GGG GTG AAG CCC GTC GGG AGC GAC CCG GAT TTC CAG
 (1) M  V  G  V  K  P  V  G  S  D  P  D  F  Q

  43 CCA GAG CTG AGC GGC GCG GGC TCC AGA CTC GCC GTG GTC AAG
(15) P  E  L  S  G  A  G  S  R  L  A  V  V  K

  85 TTC ACC ATG AGA GGG TGT GGG CCA TGT TTG AGG ATT GCC CCA
(29) F  T  M  R  G  C  G  P  C  L  R  I  A  P

 127 GCA TTC AGT TCT ATG AGT AAT AAA TAT CCA CAG GCT GTT TTC
(43) A  F  S  S  M  S  N  K  Y  P  Q  A  V  F

 169 TTG GAA GTC GAT GTA CAT CAG TGT CAG GGA ACA GCT GCC ACC
(57) L  E  V  D  V  H  Q  C  Q  G  T  A  A  T

 211 AAC AAT ATA TCA GCA ACA CCT ACA TTT TTG TTT TTT CGA AAC
(71) N  N  I  S  A  T  P  T  F  L  F  F  R  N

 253 AAA GTG AGA ATT GAT CAA TAT CAA GGA GCA GAT GCT GTG GGA
(85) K  V  R  I  D  Q  Y  Q  G  A  D  A  V  G

 295 TTA GAA GAA AAA ATC AAG CAG CAC TTA GAA AAT GAC CCT GGA
(99) L  E  E  K  I  K  Q  H  L  E  N  D  P  G

 337 AGC AAT GAG GAC ACA GAT ATT CCA AAA GGC TAT ATG GAT TTA
(113) S  N  E  D  T  D  I  P  K  G  Y  M  D  L

 379 ATG CCT TTT ATT AAC AAA GCT GGT TGT GAA TGT CTT AAT GAA
(127) M  P  F  I  N  K  A  G  C  E  C  L  N  E

 421 AGT GAT GAG CAT GGA TTT GAC AAC TGT TTA CGA AAA GAC ACA
(141) S  D  E  H  G  F  D  N  C  L  R  K  D  T

 463 ACC TTC TTG GAA TCT GAC TGT GAT GAA CAG CTG CTT ATT ACT
(155) T  F  L  E  S  D  C  D  E  Q  L  L  I  T

 505 GTG GCA TTC AAT CAA CCT GTT AAG CTT TAT TCC ATG AAA TTT
(169) V  A  F  N  Q  P  V  K  L  Y  S  M  K  F

 547 CAA GGG CCA GAT AAT GGT CAG GGC CCT AAA TAT GTA AAA ATT
(183) Q  G  P  D  N  G  Q  G  P  K  Y  V  K  I

 589 TTT ATC AAC CTA CCC CGA TCT ATG GAT TTT GAA GAG GCA GAA
(197) F  I  N  L  P  R  S  M  D  F  E  E  A  E

 631 AGA AGT GAA CCA ACT CAA GCT CTG GAA CTG ACA GAG GAT GAT
(211) R  S  E  P  T  Q  A  L  E  L  T  E  D  D

 673 ATT AAA GAA GAT GGC ATT GTT CCA CTT CGT TAT GTT AAG TTT
(225) I  K  E  D  G  I  V  P  L  R  Y  V  K  F

 715 CAG AAT GTT AAC AGT GTA ACT ATA TTT GTT CAG TCG AAT CAA
(239) Q  N  V  N  S  V  T  I  F  V  Q  S  N  Q

 757 GGT GAA GAG GAA ACA ACA AGA ATT TCA TAT TTT ACT TTT ATT
(253) G  E  E  E  T  T  R  I  S  Y  F  T  F  I

 799 GGT ACT CCA GTC CAG GCA ACA AAT ATG AAT GAC TTC AAA CGA
(267) G  T  P  V  Q  A  T  N  M  N  D  F  K  R

 841 GTA GTT GGC AAA AAA GGA GAA AGC CAC TAA GGTACAAAAGACACT
(281) V  V  G  K  K  G  E  S  H  ***

 886 GGACAACCATATTGCAATCAGATCTACAGCTCCTGGATAATTGCTTGATTCTCGC
 941 CAGAGACTGTCAATGTTTCATTCATTGCCATTACCAATAAATCATTGCTTTTGTT
 996 CAGATGGTATCACTAGTGTTTCCATTGTAATCTTGACACATGCAATTGTAAATAA
1051 AAGTCACCACTTTTGCCAAGCTTAAAAAAAAAAAAAAAAAAA
```

FIG. 1. Nucleotide and deduced amino acid sequences of human Txl. Amino acid numbers are shown in parenthesis, active site is boxed and the two polyadenylation signals are underlined. The stop codons are marked with asterisks.

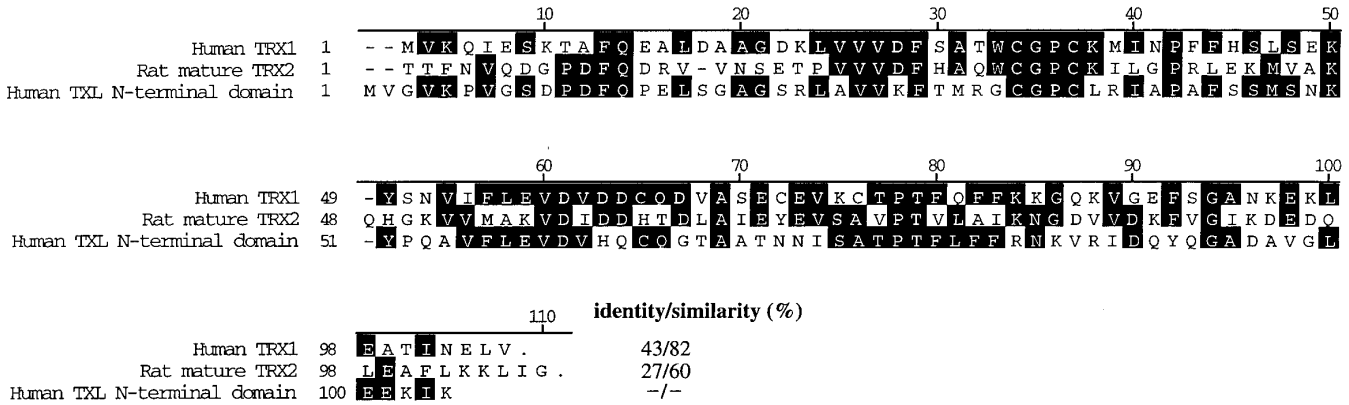

FIG. 2. Alignment of the predicted amino acid sequence of human Txl with that of human Trx1 and mature portion of rat Trx2. Black boxes indicate conserved amino acid residues. The sequence of human Txl was used as a reference for the identity/similarity values.

frame probe was labelled with [α-^{32}P]dCTP (Rediprime random primer labelling kit, Amersham) and hybridized at 65°C overnight in ExpressHyb Solution following the protocol provided by Clontech. The blots were also hybridized with human G3PDH probe as control. The blots were scanned and quantified with the Gel Pro Analyzer Program (Media Cybernetics) and the values were corrected for the G3PDH values.

RESULTS AND DISCUSSION

Many different EST entries were found to be homologous but not identical to human Trx1 sequence when dbEST databases were BLAST-searched against the region centered at the active site of human Trx1. These ESTs most likely represented partial cDNA segments of a novel human thioredoxin-like protein and were so abundant in the database that it was possible to construct a contig, a hypothetical cDNA, simply by splicing the 53 overlapping ESTs obtained.

To actually clone the cDNA we designed two primers at the 5′ and 3′ positions of the contig and used them to amplify human testis cDNA by PCR. The resulting PCR product was cloned into the pGEM-Teasy vector and sequenced, confirming the sequence obtained from the overlapping ESTs. The complete sequence of the cDNA (1,278 bp) consists of 205 bp of 5′-untranslated sequence (including a stop codon in frame), an open reading frame of 870 bp and a 203 bp fragment of 3′-untranslated, including two polyadenylation signals (AATAAA) before the poly(A)$^+$ tail (Fig. 1). The open reading frame codes for a protein of 289 amino acids with an estimated molecular mass of 32.3 kDa. This protein has two distinct domains: an N-terminal domain highly homologous to human Trx1 (3) and mature portion of rat mitochondrial Trx2 (4) comprising the first 105 amino acids, and a C-terminal domain of 184 amino acids with no homology with any other protein in the database. The N-terminal part of the protein contains the active site amino acid sequence (GCGPC) very similar to that found in all thioredoxins (WCGPC) except that the tryptophan residue is exchanged for a glycine residue (Fig. 2). This is the first report where

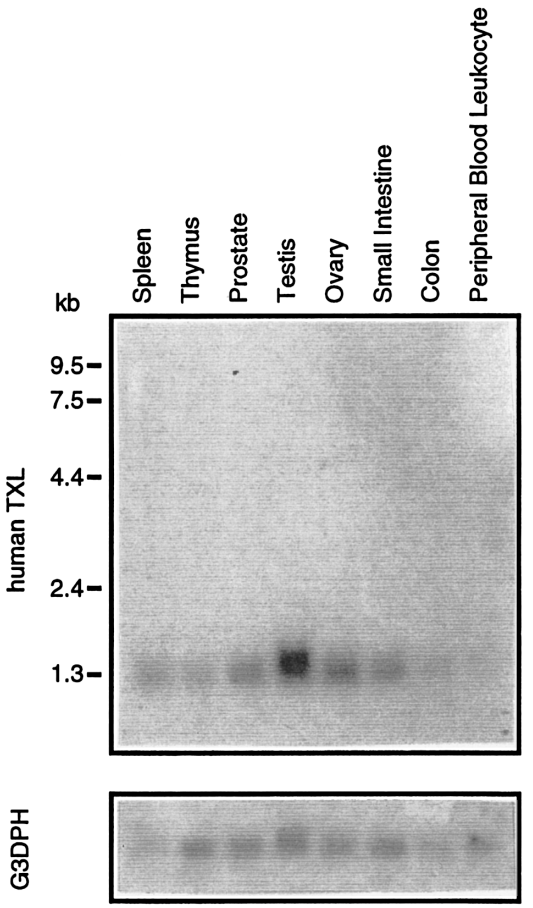

FIG. 3. Multiple tissue Northern Blot analysis of human Txl mRNA expression. G3PDH was used as a control to determine the relative amount of mRNA from each tissue.

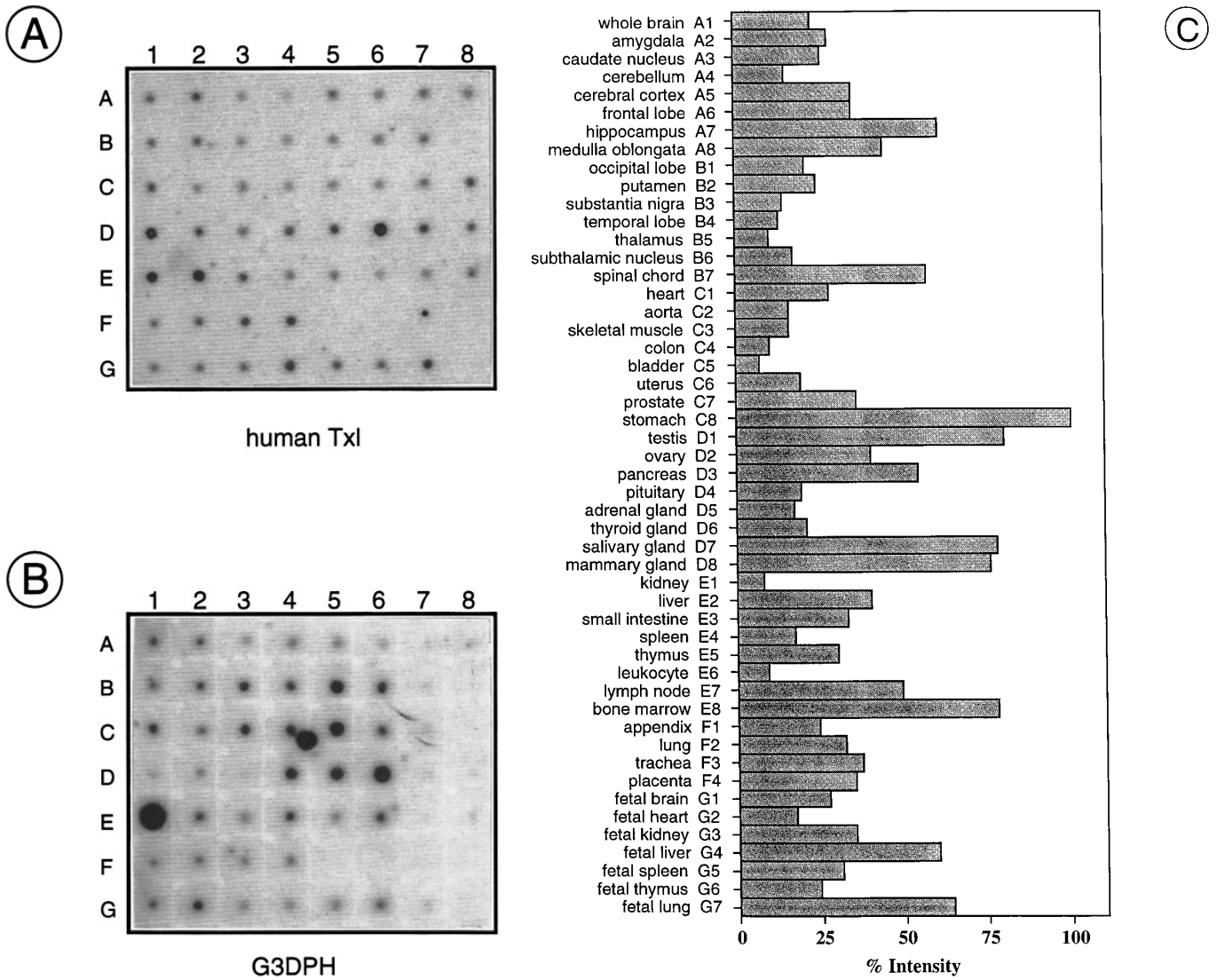

FIG. 4. Human RNA Master Blot analysis of A) human Txl mRNA; B) human G3DPH mRNA. C) The blots were scanned, quantitated and the intensity divided by its corresponding G3DPH value. The graphic shows the intensity percentage referred to the highest value (stomach).

a thioredoxin-like protein has the residue corresponding to tryptophan 31 in human Trx1 substituted by a glycine residue. The N-terminal domain of human Txl shows 43% identity with human Trx1 and 36% identity with the mature form of rat Trx2 (without the mitochondrial targeting sequence). Some of the structural amino acids (Phe-12, Pro-40, Asp-59) or residues involved in protein-protein interaction (Ala-93 and Glu-57) (20) are also conserved in Txl (Fig. 2).

The main difference between Txl and the two previously described mammalian thioredoxins (3,4) is its atypical molecular weight (32,2 kDa) due to the existence of 184 extra amino acids at the C-terminus of the protein. This domain does not match any consensus sequence for translocation to any compartment of the cell and does not contain any cleavage or maturation sequence that could lead to a mature protein with characteristics similar to the rest of the mammalian thioredoxins. In addition, the full-length Txl protein has five additional cysteine residues apart from the two cysteines at the active site.

To test if Txl is a substrate for thioredoxin reductase we expressed the full-length protein in *Escherichia coli*. When assayed with NADPH and calf thymus thioredoxin reductase in the insulin assay, Txl (up to 30 μM final concentration) showed no activity while human Trx1, used as a positive control, was able to reduce insulin disulfides at a final concentration 100-fold

lower than Txl (data not shown). To investigate whether the C-terminal domain could affect the interaction of Txl with thioredoxin reductase we expressed a truncated form of the protein (105 amino acids) lacking the C-terminal domain. When assayed under the same conditions as the full-length protein, again no activity was obtained (data not shown). A possible explanation for this fact could be the substitution of the tryptophan residue at the active site for a glycine. At this position, the tryptophan residue has been described to stabilize the three-dimensional structure of the active site thus facilitating the interaction with proteins like ribonucleotide reductase or thioredoxin reductase (20). Indeed, a substitution of the tryptophane residue at the active site of the *Escherichia coli* thioredoxin to alanine leads to a decrease of the enzymatic activity to only 4% compared to wild type (21). Since alanine and glycine residues only differ in a methyl group, is reasonable to assume that the Gly substitution of the Trp residue could result in reduction of the enzymatic activity. As mentioned in the Introduction, some proteins with thioredoxin-like domains can be reduced by their homologous thioredoxin reductases. However, none of these proteins have a substitution at the tryptophan residue, only in any of the two internal residues of the CXXC motif, further supporting this hypothesis. Mutagenesis of the glycine residue to tryptophan would be required to clarify this point.

We used a human multiple tissue Northern blot to determine the size of human Txl mRNA. The open reading frame of the Txl, used as a probe, hybridized to an mRNA of 1.3 kb (Fig. 3), in good agreement with the size of the cDNA (1,278 bp). Highest expression was seen in testis and lowest in colon and peripheral blood leukocytes. The same probe was hybridized with an "RNA Master Blot" to determine the tissue distribution of human Txl mRNA in a much wider array of tissues. Txl is ubiquitously expressed with highest values in stomach, testis, bone marrow and mammary and salivary glands (Fig. 4). The lowest expression was in colon, bladder, kidney and leukocytes. The results obtained with the "RNA Master Blot" are in agreement with the results from the Northern blot analysis (high expression in testis and low in colon and leukocytes). The expression pattern of Txl is different from that of rat Trx1 and rat Trx2 mRNAs that are also widely expressed in many tissues (4). When a phylogenetic analysis of thioredoxins is performed, the N-terminal part of Txl is placed much closer to Trx1 than Trx2 (data not shown). The fact that human Txl is expressed in all the tissues assayed, together with its unability to be reduced by human thioredoxin reductase raises the question of its cellular function. In addition, the fact that Txl is highly expressed in very specialized tissues like testis or bone marrow suggests that it has an important role in the metabolism of these tissues.

ACKNOWLEDGMENTS

This work was supported by grants from the Swedish Medical Research Council (Project 13X-10370) and Spanish Ministerio de Educación y Cultura (A.M.-V.).

REFERENCES

1. Laurent, T. C., Moore, E. C., and Reichard, P. (1964) *J. Biol. Chem.* **239,** 3436–3444.
2. Miranda-Vizuete, A., Damdimopoulos, A. E., Gustafsson, J.-Å., and Spyrou, G. (1997) *J. Biol. Chem.* **272,** 30841–30847.
3. Wollman, E., d'Auriol, L., Rimsky, L., Shaw, A., Jacquot, J., Wingfield, P., Graber, P., Dessarps, F., Robin, P., Galibert, F., Bertoglio, J., and Fradelizi, P. (1988) *J. Biol. Chem.* **263,** 15506–15512.
4. Spyrou, G., Enmark, E., Miranda-Vizuete, A., and Gustafsson, J.-Å. (1997) *J. Biol. Chem.* **272,** 2936–2941.
5. Holmgren, A. (1985) *Annu. Rev. Biochem.* **54,** 237–271.
6. Jeng, M.-F., Campbell, A. P., Begley, T., Holmgren, A., Case, D. A., Wright, P. E., and Dyson, H. J. (1994) *Structure* **2,** 853–868.
7. Qin, J., Clore, G. M., and Gronenborn, A. M. (1994) *Structure* **2,** 503–522.
8. Katti, S. K., LeMaster, D. M., and Eklund, H. (1990) *J. Mol. Biol.* **212,** 167–184.
9. Weichsel, A., Gasdaska, J. R., Powis, G., and Montfort, W. R. (1996) *Structure* **4,** 735–751.
10. Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A., and Rutter, W. J. (1985) *Nature* **317,** 267–270.
11. Fullekrug, J., Sonnichsen, B., Wunsch, U., Arseven, K., Nguyen Van, P., Soling, H. D., and Mieskes, G. (1994) *J. Cell Biol.* **107,** 2719–2727.
12. Bennett, C. F., Balcarek, J. M., Varrichio, A., and Crooke, S. T. (1988) *Nature* **334,** 268–270.
13. Kurooka, H., Kato, K., Minoguchi, S., Takahashi, Y., Ikeda, J.-E., Habu, S., Osawa, N., Buchberg, A. M., Moriwaki, K., Shisa, H., and Honjo, T. (1997) *Genomics* **39,** 331–339.
14. Lim, C.-J., Fuchs, J. A., McFarlan, S. C., and Hogenkamp, H. P. C. (1987) *J. Biol. Chem.* **262,** 12114–12119.
15. Loferer, H., Bott, M., and Hennecke, H. (1993) *EMBO J.* **12,** 3373–3383.
16. Lundstrom, J., and Holmgren, A. (1990) *J. Biol. Chem.* **265,** 9114–9120.
17. Lundstrom-Ljung, J., Birnbach, U., Rupp, K., Soling, H. D., and Holmgren, A. (1995) *FEBS Lett.* **357,** 305–308.
18. Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990) *J. Mol. Biol.* **215,** 403–410.
19. Holmgren, A., and Björnstedt, M. (1995) *Methods Enzymol.* **252,** 199–208.
20. Eklund, H., Gleason, F. K., and Holmgren, A. (1991) *Proteins: Structure, Function & Genetics* **11,** 13–28.
21. Krause, G., and Holmgren, A. (1991) *J. Biol. Chem.* **266,** 4056–4066.