Biochemical and Biophysical Research Communications **257,** 67–73 (1999)
Article ID bbrc.1999.0417, available online at http://www.idealibrary.com on IDEAL®

Characterization of *TFG* in *Mus musculus* and *Caenorhabditis elegans*

Marina Mencinger[1] and Pierre Aman
Department of Clinical Genetics, University Hospital, S-221 85, Lund, Sweden

Received February 10, 1999

***TFG* was discovered as a fusion partner of *NTRK1* in human papillary thyroid carcinoma. We assembled the mouse *TFG* cDNA from EST sequences and 5' end RACE product, identified full coding length *TFG* EST clones in pig (c17b07) and *Schistosoma mansoni* (SMNAS62), and analyzed the genomic structure of *TFG* in *Caenorhabditis elegans* (Y63D3A). The protein sequences of mouse, pig, and *S. mansoni* TFG are highly homologous to human TFG. The *C. elegans* sequence has diverged, but its predicted secondary structure is remarkably conserved. Human, mouse, and *C. elegans* TFG contain a putative trimeric N-terminal coiled-coil domain, glycosylation, myristylation, and phosphorylation sites, and SH2- and SH3-binding motifs. The SH2-binding motif is absent in *C. elegans* TFG. The expression of *TFG* does not vary among 7, 11, 15, and 19 day mouse embryonal stages. *In situ* hybridization with a *TFG* probe in 10, 5-day whole mouse embryos showed preferential staining of the limb buds, branchial arches, nasal processes, and brain, and weak staining of the primitive spinal cord and dorsal root ganglia.**

The *TFG* gene is one of the three genes identified as fusion partners of *NTRK1* in chromosomal translocations in human papillary thyroid carcinoma (1). The NTRK1 protein is a tyrosine kinase receptor for the nerve growth factor, whereas the function of the TFG protein is not known (2). As with other genes fusing to *NTRK1* in papillary thyroid carcinoma, *TFG* is ubiquitously expressed in human adult tissues and the protein can form trimer or tetramer complexes with itself (3, 4). The 400 amino acids TFG protein contains an N-terminal coiled-coil domain, phosphorylation, glycosylation, myristylation sites and an acidic C-terminal rich in amino acids S, P, Y, G, Q (5). The fusion protein between TFG and NTRK1 was localized to the cell cytoplasm and has the capability to transform NIH3T3 mouse cells (3). *NTRK1* is normally expressed only in neural tissues, the protein, however, displays a constitutive tyrosine kinase activity when fused to other proteins in papillary thyroid carcinoma (1).

Our aim was to assemble the mouse *TFG* cDNA and to identify *TFG* sequences of other species present in databases. We took advantage of the EST (Expressed Sequence Tags) databases, containing partial sequences of randomly selected cDNA clones and genomic databases using the BLAST sequence similarity searches. We compared the human TFG protein sequence to partial or full length TFG protein sequences from mouse, pig, *C. elegans* and *S. mansoni.* The expression pattern of *TFG* in mouse embryonal tissues was analyzed at different embryonal stages by Northern blot and in 10,5 days (d) whole embryos by *in situ* hybridization.

MATERIALS AND METHODS

Homology searches and sequence analysis. The human TFG amino acid sequence was used as a query in BLAST homology searches against EST databases (6). Overlapping mouse EST sequences were assembled into a full length mouse *TFG* cDNA using the GCG Wisconsin package of programs. EST clones 439282, 637422 and 367380 were obtained from official distributors. They were partially sequenced to rule out uncertain parts in the assembled sequence using vector specific primers and primers based on human or mouse *TFG* sequence (R1: gctgccataacctgagttgactgtttcc; R2: ctggtcatctgttaatccaaatgctga; R3: catagggaggttgctgcggc; F1: gtgcgattggccagctctgc; F2: catagggaggttgctgcggc F3: cagttccagggatatggccag). EST sequences from *C. elegans* (yk96f8, yk479h2, yk403f1) were used as queries against the *C. elegans* genomic database. A genomic clone (Y63D3A) from *C. elegans,* and EST clones from pig (c17b07) and *S. mansoni* (SMNAS62) were analyzed using the GCG package. Putative partial or full length TFG protein sequences from different species were aligned using ClustalX (7). Coiled-coil domains were predicted by COILS, PAIRCOIL and MULTICOIL program (8) (9) (10). A secondary structure prediction of the proteins was obtained with the Predict Protein program (11). A comparative chro-

[1] To whom correspondence should be addressed. Fax: +46-46-131061. E-mail: Marina.Mencinger@klingen.lu.se.

Abbreviations used: EST, Expressed Sequence Tag; *TFG, TRKA* (also known as NTRK1) fusion gene; NTRK1, neurotrophic tyrosine kinase receptor 1; SHPS-1, SHP substrate 1; CAMP, cAMP and cGMP-dependent protein kinase; PKC, protein kinase C; CK2, casein kinase 2; N, asparagine; mouse, *Mus musculus;* pig, *Sus scrofa; C. elegans, Caenorhabditis elegans; S. mansoni, Schistosoma mansoni;* d, days.

0006-291X/99 $30.00

mosomal map between human and mouse, which is part of the NCBI (National Center for Biotechnology Information) resources, was used to predict the chromosomal localization of the *TFG* in mouse (12).

DNA probes and hybridization to Northern membrane. The mouse EST clone 763875 was selected because it is identical to the previously published mouse gene sequence X79233, whereas the clone 475422 was selected for its high similarity to the human *FUS* gene. These purchased EST clones and the 5' end *TFG* clone were labeled with P^{32} and used separately as probes for hybridization to the Northern membrane (Clontech, 7763-1). *EWS* and *FUS* probes served as internal controls. Filter hybridization were performed as described (13).

RNA probes and whole mount hybridization. Antisense and sense mouse *TFG* RNA probes were synthesized by *in vitro* transcription reaction. The mouse *TFG* clone 367380 was linearized with EcoRI or NotI and used as a template for the generation of antisense and sense RNA probes, respectively. Probes were labeled using RNA DIG labeling kit (Roche, 1175025) according to the recommendations of the manufacturer. The T3 and T7 RNA polymerase were used for the generation of the antisense or sense probe, respectively. The reaction mix was precipitated with ethanol and sodium acetate and diluted in 20 μl of DEPC-H_2O. The RNA concentration was measured with a spectrophotometer. One μl of the probe solution was analyzed on an agarose gel as a quality control.

Embryos were dissected from mouse at day 10,5 post coitum and placed in PBS. The roof of the 4th ventricle and the telencephalon was punctured to avoid probe trapping. Embryos were fixed in 4% of paraformaldehyde in PBS at 4°C overnight with gentle rocking. After washing in PBT (Dublecos PBS and 0,1% Tween) twice at RT for 5 min, they were stored in PBT at 4°C for no longer than a week. For storage of up to 3 months, embryos were dehydrated in 25%, 50%, 75% and 100% methanol and stored at −20°C. Before use they were rehydrated in falling concentrations of methanol and washed afterwards twice in PBT for 5 min each time.

The embryos were permeabilized with 10 μg/ml of proteinase K at 37°C for 15 min. During all washes the embryos were placed on a rocking platform. Two successive washes with 2mg/ml glycine in PBT were carried out before the embryos were refixed with 0,2% glutaraldehyde and 4% paraformaldehyde in PBT at RT for 20 min. The embryos were washed twice in PBT for 5 min and transferred to round bottom 2 ml Nunclone tubes. Prehybridization was done at 70°C for at least one hour in a hybridization solution [50% deionized formamide, 5× SSC (pH=4,5), 50 μg/ml yeast tRNA, 1% SDS and 50 μg/ml heparin]. One μg of RNA probe was heated to 80°C for 2 min and resuspended in 1 ml of hybridization solution. The hybridization solution was replaced by the hybridization solution containing the riboprobe. The hybridization was carried out at 70°C overnight. Unbound probe was washed off with solutions: prewarmed solution I [50% deionized formamide, 4× SSC (pH=4,5), 2% SDS] at 70°C twice for 30 min, 1:1 prewarmed mix of solution I and II at 70°C once for 10 min solution, and with solution II [0,5M NaCl, 0,1M Tris (pH=7,5), 0,01% of Tween-20] at RT three times for 5 min. Single stranded RNA was degraded with 100 μg/ml RNAse in solution II at 37°C for 30 min. The embryos were washed with solution II and then solution III [50% deionized formamide, 2× SSC (pH=4)] each at RT for 5 min. Two washes were performed with prewarmed solution III at 65°C for 30 min. Then, three washes were performed in Maleic Acid Buffer (MAB) (pH=7,5) [100 mM Maleic Acid (Sigma M0375), 150 mM NaCl, 10 M NaOH, 0,1% Tween-20 and 2mM levamisole (Sigma L9756)] at RT for 5 min each. Non-specific antibody binding sites were blocked by preincubation of the embryos in 2% Blocking Reagent (Roche, 1096176) and 10% normal sheep serum in MAB at RT for 3 hours. Three mg of embryo powder were added to 0,5 ml MAB with 2% Blocking Reagent and the tube was incubated at 70°C for 30 min. The tube was vortexed and cooled. Five μl of normal sheep serum and 2 μl of antibodies conjugated with anti-DIG-alkaline phosphatase were added to the cooled tube, which was then placed on a rocking table at 4°C for at least 1 hour. The tube was centrifuged at 4°C for 10 min and the supernatant was diluted to 6 ml with MAB, 2% Blocking Reagent and 1% normal sheep serum. The blocking solution was removed from the embryos and replaced with the solution containing antibodies. The embryos were placed on a rocking platform at 4°C overnight. Next day the embryos were transferred to 50 ml tubes and washed with MAB buffer three times for 5 min every hour for a total period of 8 hours, and left at 4°C overnight on a rocking table. The fourth day, they were washed three times with NTMT [100mM NaCl, 100mM Tris-HCl (pH=9.5), 1M $MgCl_2$, 0.1% Tween-20] at RT for 10 min. The embryos were covered with 300 μl Purple alkaline phosphatase-substrate (Roche, 1442074) and 2mM levamisole and left in dark for 4-12 hours.

Rapid amplification of cDNA ends (RACE). In order to isolate the 5' end of mouse *TFG* cDNA a 5' RACE was performed using two adapter primers (Clontech) and two nested specific reverse primers, an outer primer R2 (5'-ctggtcatctgttaatccaaatgctga-3') and an inner primer R1 (5'-gctgccataacctgagttgactgtttcc-3') for the mouse sequence. The first PCR reaction was performed using 5 μl adapter-ligated double stranded cDNA (Clontech) as a template, 0,5 μM outer specific primer, 0,5 μM of outer adapter primer, 2,5 mM $MgCl_2$, 10× PCR reaction buffer II, 0,2 mM each dNTP and 1 unit of Gold AmpliTaq polymerase. Three μl of the first PCR product were used in a second PCR reaction with the same composition as in the first, exept that the inner primers were used.

RESULTS

Mouse TFG cDNA. Several mouse EST clones encoding protein sequences with high similarity to the human TFG protein were selected using protein sequence similarity searches. EST clones 637422, 439282, and 367380 were partially sequenced in our laboratory. Fifteen overlapping sequences of mouse EST clones were aligned step-wise on each other to obtain the full length mouse *TFG* cDNA (Fig. 1). The 5' end of clone 637422 extends upstream of the ATG translation initiation codon and its 3' end sequence contains a polyadenylation-signal and a poly(A)-tail (Fig. 1). This clone, therefore, represents the full length coding mouse *TFG* cDNA sequence. A 5' RACE (rapid amplification of cDNA ends) experiment revealed several bands. The longest band of 700 bp was cloned and sequenced. The cDNA sequence of this band does not extend further than the sequence of the most 5' end mouse EST (AA185073). The non-coding 5' end of the mouse *TFG* transcript is around 200 bp longer than that published for human (Y07968), which has only 18 nucleotides. However, we found a new human EST clone AA636060 that overlaps and extends 275 nucleotides upstream of the ATG codon in the published *TFG* human sequence.

Three clones (421877, 989186 and 637422) lack 12 nucleotides between nucleotide 922 and 934 of the coding sequence, which are present in other mouse clones (1054104, 865151 and 304119), as well as in human *TFG* cDNA (Fig.2). Two of these clones (637422 and 304119) were control-sequenced in our laboratory and we confirmed the existence of both cDNA variants.

The mouse *TFG* cDNA has 1861 nucleotides, contains an open reading frame of 397 amino acids, a polyadenylation-signal and a poly(A)-tail (Fig. 2). The

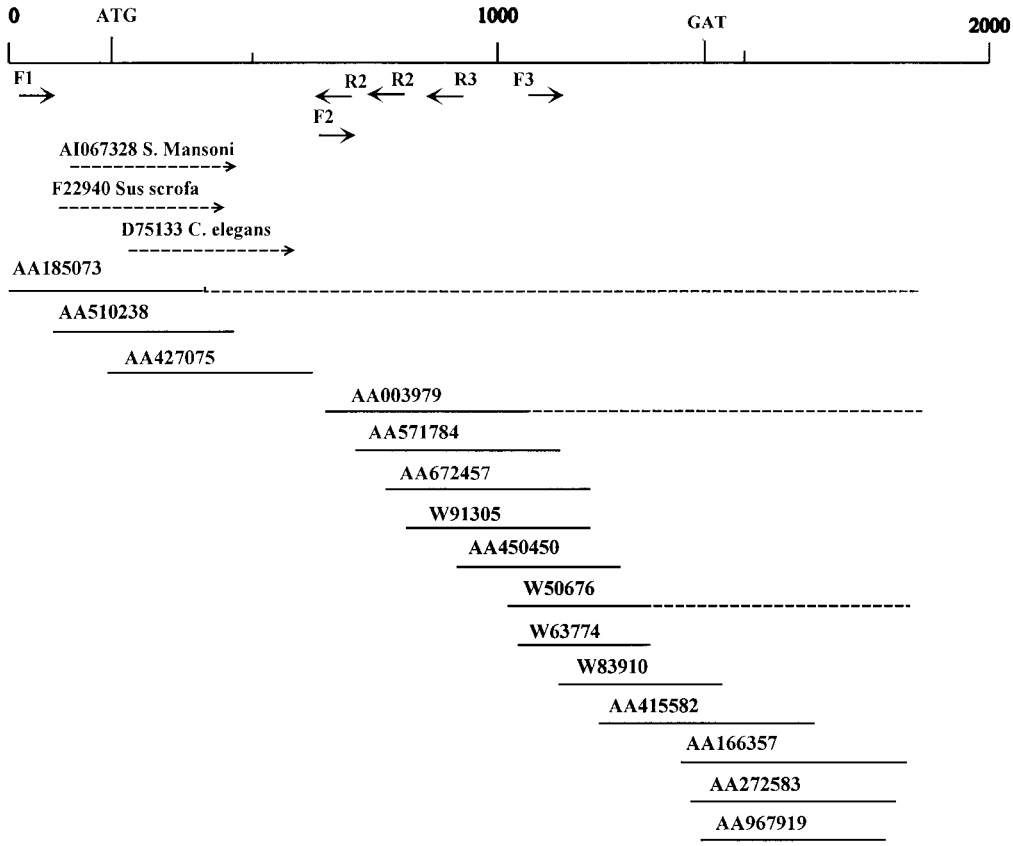

FIG. 1. An assembly of the EST clones into the full length mouse *TFG* cDNA. Sequenced parts of the mouse and other EST (pig, *S. mansoni* and *C. elegans*) clones aligned to the assembled *TFG* cDNA sequence are shown as lines or dotted arrows marked with their accession numbers, respectively. Mouse EST clones selected for partial sequencing in our laboratory are shown as solid lines continuing into dotted lines. Positions of PCR primers are indicated with arrows and names under the mouse *TFG* cDNA line. The initiation ATG and stop codon GAT are indicated above the *TFG* cDNA line.

putative mouse TFG protein shows 94% identity with human TFG (Fig. 3). N-glycosylation sites, phosphorylation sites for casein kinase 2 (CK2) and protein kinase C (PKC), as well as myristylation sites were identified in the mouse TFG protein (Fig. 2). A consensus sequence, (L/V/I)YXX(L/V/I), of a SH2-binding motif and several consensus sequences, PXXP, of a SH3-binding motifs are also present in the TFG protein sequence (Fig. 3). The SH2-binding motif is specific for the tyrosine phosphatase, SHPS-1, whereas the SH3-binding motif is common for all SH3-binding proteins (14) (15).

TFG ESTs in Schistosoma mansoni, Sus scrofa and Caenorhabditis elegans, and a TFG genomic clone in Caenorhabditis elegans. *S. mansoni* (SMNAS62), *S. scrofa* (pig) (c17b07) and three overlapping *C. elegans* clones (yk96f8, yk479h2, yk403f1) from EST databases showed high sequence homology to the human TFG protein. The translated pig and *S. mansoni* EST sequences had nearly 96% identity over 73 amino acids and 67% identitiy over 90 amino acids to the human TFG protein, respectively (Fig. 3). Sequences of the three *C. elegans* ESTs were found present on a genomic clone (Y63D3A) deposited in a C. elegans genomic database. Three exons, comprising the whole *TFG* gene, were predicted in the sequence of Y63D3A (Table 1). Exon 1, 2 and the upstream part of exon 3 were identified by aligning 5' end sequences of selected ESTs to the genomic sequence of Y63D3A. The exon 3 was translated into an open reading frame with continuing similarity to the human TFG until the stop codon. A polyadenylation signal was found 418 nucleotides downstream of this stop codon. The splice signal sequences followed the "gt-ag" rule (Table 1). The predicted exons of *C. elegans* genomic *TFG* sequence were translated into the 487 amino acid protein, including the initiation codon ATG and the stop codon TAA, with 26% identity and around 50% similarity over full length of the human TFG protein (Fig. 3). A trimeric coiled-coil domain (spanning amino acid 92-116 or 92-125, predicted by COILS and PAIRCOILS, respectively), N-glycosylation sites, myristylation sites, phosphorylation sites for cAMP and cGTP dependent kinase (CAMP), PKC and CK2, as well as several SH3-

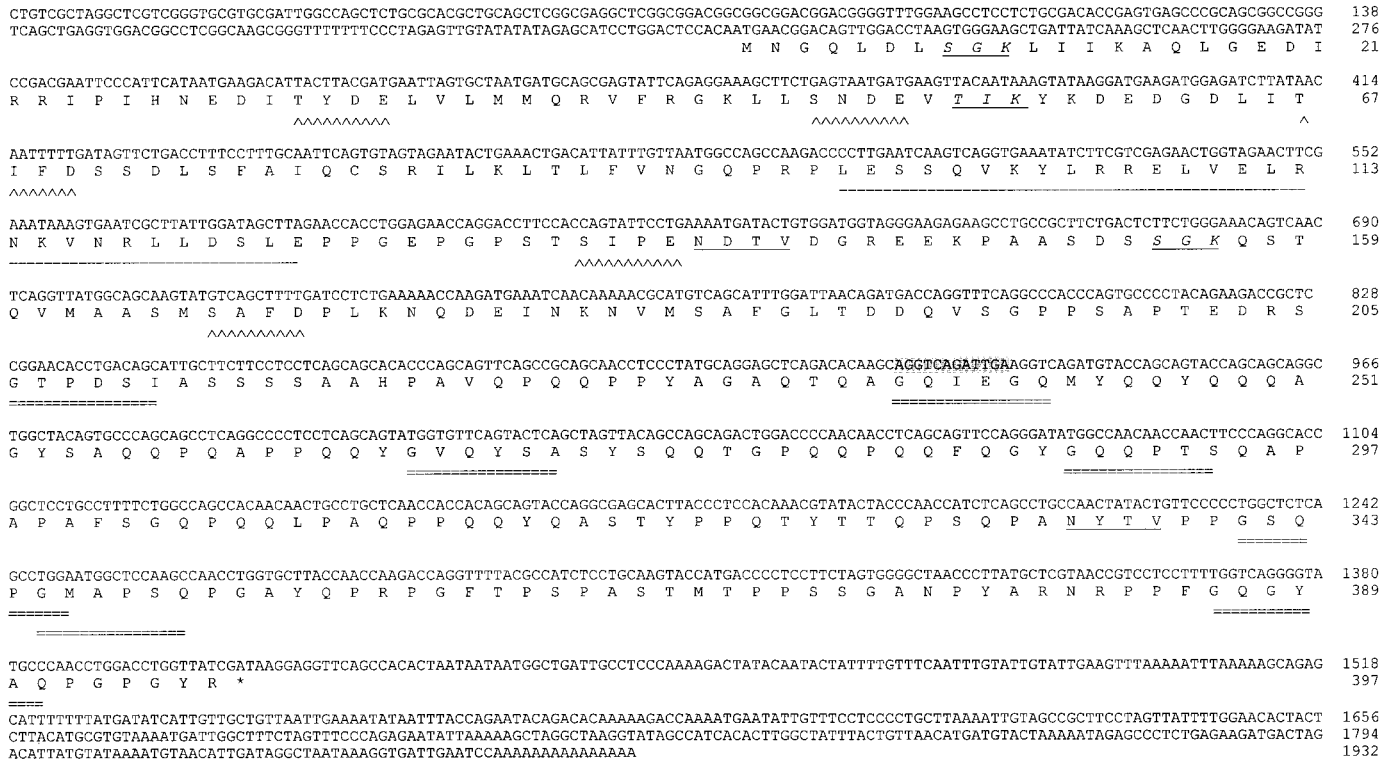

FIG 2. Nucleotide sequence and putative amino acid sequence of the mouse *TFG* cDNA (Accession No. U94662). Twelve nucleotides that are not present in some clones are shadowed. Asterisk indicates the stop codon. Regions matching different putative protein sites are marked: N-glycolysation sites are underlined, the phosphorylation sites for PKC are underlined and written in italic, and myristylation sites are indicated by double-dashed underline. The putative coiled-coil is indicated by dashed underline and the region rich in S,P,Y,G,Q amino acids is delimited with angle brackets.

binding motifs were found in the *C. elegans* TFG (Fig. 3).

Expression in mouse embryo and putative chromosomal localization. No significant difference in intensity of the signal was observed among the lanes on the Northern membrane with RNA from 7, 11, 15 and 19 days mouse embryos using *TFG*, *EWS* or *FUS* probes (Fig. 4a). A distinct staining among organs was observed by using the *TFG* antisense probe for ISH on 10,5 d whole embryos. The limb buds, branchial arches, nasal processes and brain stained moderately, whereas the primitive spinal cord and the dorsal root ganglia stained less. Other organs appeared negative. The sense *TFG* probe, used as a negative control, did not produce any signal (Fig. 4b).

The chromosomal location 3q12 of the human *TFG* was found synthenic with the chromosomal arm 16 q, between 22 and 26 cM from centromere, according to the mouse human homology map (12). No phenotype or disease loci was reported associated with this region on mouse chromosome.

DISCUSSION

In this work we describe the full length *TFG* cDNA in mouse. The cDNA was assembled based on the criteria, that at least three EST clones had identical sequences over an overlapping region. Regions covered by less than three identical EST sequences were control-sequenced in our laboratory using clones 439282, 637422, 367380 and the PCR amplified 5' end of *TFG* as templates. The size of the longest 5' RACE product is identical to the distance between the specific reverse primer and the 5' end of the assembled cDNA, suggesting that the longest EST clone reaches the 5' end of the transcript. PCR bands smaller than 700 bases are probably the result of uncompleted cDNAs during the cDNA synthesis, since only one transcript is seen on the Northern membrane (Fig. 4a).

The gap of 12 coding nucleotides in three of the mouse EST clones could be explained by an alternative splicing, a polymorphism or existence of two TFG genes in the mouse genome. The first half of these 12 nucleotides is repeated once after the 12th nucleotide, which also suggest that the elongation may be the result of a duplication event (Fig. 2). The significance of this variant cDNA is not known at present. No human clones with the gap were found, suggesting that it may be specific for the mouse *TFG*.

We have also identified EST clones encoding pig, *S. mansoni* and *C. elegans* TFG protein (Fig. 3). Pig and *S. mansoni* EST clones include the putative translation

FIG. 3. Multiple sequence alignment of the putative TFG protein in mouse, pig, *S. mansoni*, and *C. elegans* to the TFG in human. The double slash indicates the end of a protein sequence where no downstream sequence was available in database. The questionmark indicates an unrecognized amino acid. The asterisk indicates stop codon. Identical residues are shadowed and similar residues are bolded. Beta-strands are indicated by single-dashed underline and alfa-helices by double underline. The coiled-coil domain is delimited with arrows. The regions in *C. elegans* TFG matching different putative protein sites are marked: N-glycosylation sites are underlined, phosphorylation sites for PKC are underlined and written in italic, the phosphorylation site for CAMP is indicated with "+", and myristylation sites are indicated by double-dashed underline. SH2-binding motif is indicated by asterisks and SH3-binding motifs are indicated with circles in human and *C. elegans* TFG. SH2-binding motif and/or SH3-binding motifs conserved in the TFG sequences from several species are indicated only in human TFG.

initiation ATG codon, therefore they probably represent the full length *TFG* coding cDNA clones in these species, whereas the analyzed *C. elegans* ESTs lack some nucleotides at the coding 5' end (Fig. 1). The homology between human TFG protein and translated "single pass" 5' end EST sequences of pig and *S. mansoni* is high enough to assume that these EST clones represent the *TFG* transcript in these species (Fig. 3).

Despite the relatively low amino acid sequence similarity between the *C. elegans* and the human TFG, their predicted secondary structure seem to be highly conserved (Fig. 3). Thus, coiled-coil was predicted in the N-terminal of *C. elegans* TFG protein, even though this region show almost no sequence similarity to the predicted coiled-coil domain in human TFG (Fig. 3). Furthermore, coiled-coil domains of human and *C. el-*

TABLE 1

Exon/Intron Boundaries of Putative *C. elegans TFG*

Exon	3' Splice acceptor	Exon size	5' Splice donor	Intron size
1	ttttttccag/ATGGTGCATT	138	AAGGATGAAG/gtaggaattgc	51
2	tcaattttag/AAGGCGATTT	148	GCAGATTAAG/gtaaggagga	648
3	tttctttcag/CTTGATGTTG	859		

Note. Empty field means that it was not possible to obtain information from the sequence in the database.

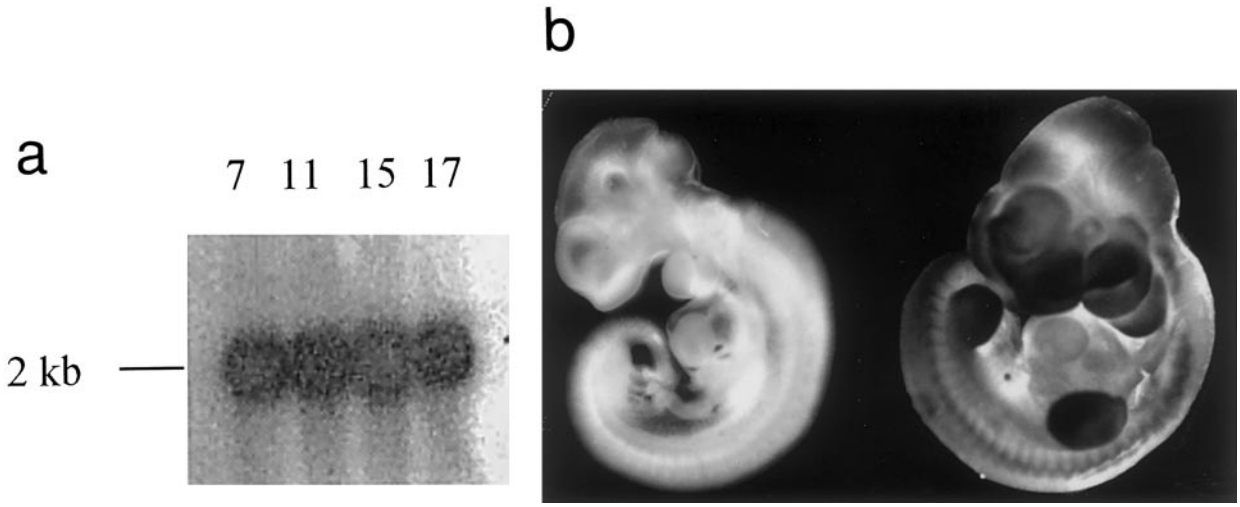

FIG 4. (a) Northern blot analysis of different mouse embryonal stages using mouse *TFG* probe. (b) ISH to 10, 5-d whole embryo with antisense (left) and sense (right) mouse *TFG* probe.

egans TFG seem to form trimers in preference to dimers as predicted by MULTICOIL program, which is in agreement with experimental data for the human TFG provided by Greco et al. (3). The human TFG protein has no close sequence homology with any other human protein, indicating that only one *TFG*-like gene exists in the human genome. Taking together, these data suggest that the described gene in *C. elegans* is indeed the *TFG*.

Most of the predicted phosphorylation sites for PKC and CK2 are conserved between human and mouse TFG, except two sites for CK2: one site (S 134) in mouse TFG, which is absent in human TFG, and one site (S 148) in human TFG, which is absent in mouse TFG (Fig. 2), (5). N-glycosylation sites are identical between human and mouse TFG. Six putative myristylation sites are identical or similar between the two species, except one (G 341) in mouse TFG, which is absent in human. Glycosylation, phosphorylation and myristylation sites were also identified in *C. elegans* TFG, but the distribution and number of these sites differ from those in human TFG (Fig. 3), (5). Human and *C. elegans* TFG differ also in one phosphorylation site for CAMP, which was only found in *C. elegans* TFG. Similar patterns in human, mouse and *C. elegans* TFG may suggest that all proteins are anchored to the plasma membrane by myristic acids and their activity is regulated by common kinases.

The SH2-binding motif specific for the second SH2 domain of the tyrosine-SHPS-1 was found at the N-terminus of human, mouse, pig and *S. mansoni*, but not of *C. elegans* TFG (Fig. 3). SHPS-1 was suggested to play a regulatory role in many cellular signaling processes (16). It is interesting to note that a deletion of the TFG N-terminus spanning amino acid 2-90, which contains the putative SH2-binding domain, abolishes the transforming ability of the TFG-NTRK1 fusion mutant (3). This deletion, however, does not affect tyrosine phosphorylation and protein-complex formation of the TFG-NTRK1 mutant (3). It may be that interaction of the TFG with the SHPS-1 is important in the transformation process. The phosphorylation on the tyrosine in the identified SH2-binding motif and its ability to bind to SHPS-1 need to be confirmed experimentaly. Several SH3-binding motifs for SRC, a human gene homologous in sequence to the v-src gene of the Rous sarcoma virus, and other nonreceptor tyrosine kinases have been shown to interact with the motif PXXP, found in the human, mouse and *C. elegans* TFG (Fig. 3) (15). The actual interaction needs an experimental proof.

The putative N-termini of *S. mansoni* and the human TFG are highly conserved suggesting an essential role for this part of the protein. The putative *C. elegans* TFG shows lower degree of conservation with the full length of human TFG (Fig. 3). According to the phylogenetical hypotheses, platyhelminthes (*S. mansoni*) and nematodes (*C. elegans*) are equally distant to vertebrata (human) and therefore their sequence homology with the human TFG would be expexted to be nearly the same (17). One can speculate that the higher degree of homology between the *S. mansoni* and the human TFG is due to the fact that *S. mansoni* is a human parasite and has adopted its metabolism and structure to a man host.

The expression of the *TFG* gene does not vary among different mouse embryonal stages as assessed by Northern blot analysis. However, variations in staining among different tissues in 10, 5 d embryos were seen by *in situ* hybridization with a *TFG* probe. *TFG*, for example, is not expresssed in heart at 10,5 d (Fig. 3b) and 12,5 d (data not shown) of mouse embryonal stage, while it is ex-

pressed in an adult human heart as shown by Northern blot analysis (3). It seems that the expression of the *TFG* is differentially regulated among certain tissues during early mouse development.

In conclusion, the TFG gene seems to be highly conserved among eutheria (human, mouse, pig) and in a species *S. mansoni*. Different mouse developmental stages need to be investigated in order to provide more detailed spatio-temporal expression data of *TFG*. The identification of SH2- and SH3-binding motifs in TFG is interesting and needs to be explored further. The identification of *TFG* cDNA or gene in evolutionary lower species provides an important tool for the study of the function of *TFG*.

ACKNOWLEDGMENTS

The authors thank Kenneth Campbell and Håkan Toresson for tutelage in ISH and for commenting on the manuscript, Juan-Antonio Contreras for advice with the secondary structure prediction and the multiple sequence alignment, Anna Härlid for comments on phylogeny, and Ioannis Panagopoulos for supporting the initiation of this study. This study was supported by grants from the Swedish Cancer Society and the Medical Faculty of Lund University, Sweden.

REFERENCES

1. Pierotti, M. A., Bongarzone, I., Grazia Borello, M., Greco, A., Pilotti, S., and Sozzi, G. (1996) *Genes Chromosom. Cancer* **16,** 1–14.
2. Barbacid, M. (1994) *J. Neurobiol.* **25**(11), 1386–1403.
3. Greco, A., Mariani, C., Miranda, C., Lupas, A., Pagliardini, S., Pomati, M., and Pierotti, M. A. (1995) *Mol. Cell. Biol.* **15,** 6118–6127.
4. Greco, A., Fusetti, L., Miranda, C., Villa, R., Zanotti, S., Pagliardini, S., and Pierotti, M. A. (1998) *Oncogene* **16**(6), 809–816.
5. Mencinger, M., Panagopoulos, I., Andreasson, P., Lassen, C., Mitelman, F., and Aman, P. (1997) *Genomics* **41**(3), 327–331.
6. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
7. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **25**(24), 4876–4882.
8. Lupas, A., Van Dyke, M., and Stock, J. (1991) *Science* **252,** 1162–1164.
9. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M., and Kim, P. S. (1995) *PNAS USA* **92,** 8259–8263.
10. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Science* **6,** 1179–1189.
11. Rost, B., and Sander, C. (1993) *J. Mol. Biol.* **232**(2), 584–599.
12. DeBry, R. W., and Seldin, M. F. (1996) *Genomics* **33,** 337–351.
13. Church, G. M., and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1991–1995.
14. Burshtyn, D. N., Yang, W., Yi, T., and Long, E. O. (1997) *J. Biol. Chem.* **272**(20), 13066–13072.
15. Feller, S. M., Ren, R., Hanafusa, H., and Baltimore, D. (1994) *Trends Biochem. Sci.* **19**(11), 453–458.
16. Fujioka, Y., Matozaki, T., Noguchi, T., Iwamatsu, A., Yamao, T., Takahashi, N., Tsuda, M., Takada, T., and Kasuga, M. (1996) *Molec. Cell. Biol.* **16,** 6887–6899.
17. Blaxter, M. (1998) *Science* **282**(5396), 2041–2046.