Journal of Cell Science 109, 675-685 (1996)
Printed in Great Britain © The Company of Biologists Limited 1996
JCS4093

Primate homologues of rat TGN38: primary structure, expression and functional implications

Sreenivasan Ponnambalam[1,2,*], Milena Girotti[3], Marie-Laure Yaspo[1], Charles E. Owen[1], Anthony C. F. Perry[3], Tatsuo Suganuma[4], Tommy Nilsson[1,5], Mike Fried[1], George Banting[3] and Graham Warren[1]

[1]Imperial Cancer Research Fund, 44 Lincoln's Inn Fields, London WC2A 3PX, UK
[2]Department of Biochemistry, University of Dundee, Dundee DD1 4HN, UK
[3]Department of Biochemistry, University of Bristol, Medical School, University Walk, Bristol BS8 1TD, UK
[4]Department of Anatomy, Miyazaki Medical College, Miyazaki, Japan
[5]EMBL, Meyerhofstrasse, Heidelberg, Germany

*Author for correspondence at address 2

SUMMARY

cDNAs encoding the human and macaque homologues of rat TGN38 have been cloned and sequenced. The proteins have a highly conserved N terminus (comprising the signal peptide) and C terminus (comprising part of the lumenal domain, the membrane spanning region and cytoplasmic tail) but vary in the other part of the lumenal domain, which contains the repeat region. Whereas rat TGN38 contains 6 tandem repeats of an 8mer, both primate proteins possess 14 tandem repeats of a 14mer sequence. The human protein, like rat TGN38, is localised primarily to the TGN but is present on the cell surface and returns via endosomes. This behaviour is consistent with conservation of the membrane spanning region and the cytoplasmic tail, which contain the retention and retrieval signals, respectively, for localisation in the TGN. The unexpected differences in the lumenal domain can best be rationalised by the fact that both types of repeat domains have most of the properties of mucins. We suggest that TGN38 homologues are mucin-like molecules that regulate membrane traffic to and from the TGN.

Key words: TGN38, TGN46, *Trans*-Golgi network, Endosome, Mucin

INTRODUCTION

The *trans*-Golgi network (TGN) is a key sorting station for proteins (reviewed by Griffiths and Simons, 1986; Kornfeld, 1992; Bauerfeind and Huttner, 1993). Three pathways of membrane traffic from the TGN identified in mammalian cells are regulated secretion, transport to endosomes and lysosomes, and constitutive exocytosis to the cell surface. Regulated secretion involves protein condensation leading to maturation of secretory granules (reviewed by Bauerfeind and Huttner, 1993). Transport to lysosomes via endosomes appears to occur via clathrin-coated vesicles as exemplified by the mannose 6-phosphate receptor, which delivers lysosomal enzymes by this route (Kornfeld, 1992). Constitutive transport to the plasma membrane is the least well-characterised vesicular route from the TGN (Griffiths and Simons, 1986; Bauerfeind and Huttner, 1993).

The steps along the secretory pathway are mediated by more than one type of vesicle. COP I-coated vesicles are implicated in both anterograde (reviewed by Rothman, 1994; Salama and Schekman, 1995) and retrograde (Letourneur et al., 1994; Salama and Schekman, 1995) transport along the secretory pathway. COP II vesicles appear to mediate transport from the ER to the Golgi apparatus (Barlowe et al., 1994). Significant amounts of β-COP, a constituent of the COP I coat, have been found on the TGN (Oprins et al., 1993; Griffiths et al., 1995) but microinjection studies have shown that this is not involved in transport from the TGN to the cell surface (Pepperkok et al., 1993). Instead, Narula and Stow (1995) have reported that vesicles budding from the TGN label for p200, a peripheral coat protein associated with the constitutive secretory pathway. Intriguingly, TGN38, originally identified as a marker for the TGN (Luzio et al., 1990), was shown to be in close proximity to this coat protein (Narula and Stow, 1995). Furthermore, Howell and co-workers have shown that peptides containing TGN38 cytoplasmic sequences inhibit vesicle budding from TGN membranes in vitro (Jones et al., 1993), suggesting that TGN38 is critically involved in the formation of constitutive exocytic vesicles (Stanley and Howell, 1993).

The TGN38 glycoprotein cycles between the TGN and the cell surface returning via endosomes (Bos et al., 1993; Reaves et al., 1993; Chapman and Munro, 1994). This cyclical movement is mediated by a tyrosine-based signal (SDYQRL) in the cytoplasmic domain (Bos et al., 1993; Humphrey et al., 1993; Wong and Hong, 1993). Mutation of the tyrosine residue abrogates TGN localisation of hybrid proteins containing TGN38-derived cytoplasmic sequences; however, flanking residues within the SDYQRL motif have varying effects on

TGN localisation of hybrid proteins, depending on the system or cell line used in such studies (Bos et al., 1993; Humphrey et al., 1993; Wong and Hong, 1993; Zehavifeferman et al., 1995). A second TGN localisation signal present in the membrane spanning domain (Ponnambalam et al., 1994; Reaves and Banting, 1994b) resembles those found in Golgi-resident enzymes (Nilsson and Warren, 1994). The combination of both signals appears to ensure that most of TGN38 is present in the TGN under steady state conditions (Ponnambalam et al., 1994; Reaves and Banting, 1994b).

In this report we describe the cloning and sequencing of the human and macaque homologues of rat TGN38. Those sequences previously shown to be important for maintaining the steady state distribution of rat TGN38 are conserved in the primate homologues, but a major part of the primate lumenal domain shows considerable sequence divergence from the rodent protein. The divergent domain, which contains a repeating motif in both rodent and primate homologues has features common to highly glycosylated, mucin-like proteins. The role played by TGN38 in the exocytic process might therefore reflect its mucin-like features.

MATERIALS AND METHODS

Materials and screening of cDNA libraries

Chemicals and reagents were obtained from the sources described Ponnambalam et al. (1994). The TGN38 cDNA (Luzio et al., 1990) was used as a probe to screen both phage and plasmid human and macaque cDNA libraries. 4×10^4 human foetal liver cDNA clones and 6×10^4 human foetal thymus clones were screened from poly(dT)-primed cDNA libraries. These libraries were directionally cloned into the pSPORT1 vector (Gibco-BRL) and spotted in duplicate at high density onto nylon filters (Reference Library Database; Zehetner and Lehrach, 1994). Approximately 1×10^6 clones from a human stomach cancer cDNA λgt10 library (in Clontech, CA) was also screened. One complete human liver cDNA (clone ICRFp512N1056) and two partial human thymus cDNAs (clones ICRFp508O02134 and ICRFp508A0679) were isolated and sequenced. A number of clones (>20) were isolated from the human colon cancer cDNA library; most appeared incomplete as judged by restriction digestion analysis and by sequencing of the 5′ ends of the subcloned inserts. The sequence of all human clones were identical.

Approximately 2×10^5 clones from a monkey (*Macara fascicularis*) testis cDNA libraries (Perry et al., 1992, 1994) were screened using a [γ-^{32}P]ATP-labelled oligonucleotide corresponding to nucleotides 20-34 of the cloned monkey TGN46 C terminus PCR product. The cDNA clones isolated from a single hybridising clone was labelled to a high specific activity with [α-^{32}P]dCTP and used to screen a similar number of *Macara fascicularis* epididymal clones. Positive clones were sequenced on both strands using the Sanger chain termination reaction either on a DuPont Genesis 2000 automated DNA analysis system or with the Sequenase version 2.0 kit. The missing 5′-end of the macaque TGN47 cDNA was cloned by RT-PCR. An oligonucleotide (no. 1) corresponding to macaque TGN47 nucleotides 847-866 (5′-AGCTCCTTGTTATCAGAAGG-3′) was used to prime cDNA synthesis using 2 μg of monkey testis poly(A)$^+$ mRNA and a 5′ RACE kit (Clontech, Palo Alto, CA). Purified single-stranded cDNA was used in a PCR reaction using a reverse oligonucleotide complementary to nucleotides 536-555 of macaque TGN47 (5′-AGTGCTGTCTTTCGTGGTCT-3′) and a forward nucleotide (no. 2) based on the sequence of the 5′ UTR of the human TGN46 cDNA (5′-AAGCGCTATCCGAGCAGG-3′). PCR conditions were as follows: 94°C, 1 second; 94°C, 2 minutes: 50°C, 1 minute 30 seconds; 50°C, 2 minutes; 72°C, 1 second; 72°C, 2 minutes 30 seconds; 30 cycles. The resulting PCR fragment was subcloned into pUC18 and clones from four independent reactions were sequenced. To show that the half repeat insertion in macaque TGN47 (residues 191-198) was not a cloning artefact, oligonucleotides no. 1 and no. 2 were used used in a PCR reaction using *Pfu* DNA polymerase (Stratagene, La Jolla, CA). The PCR product was subcloned, sequenced and compared with the original macaque cDNA sequence.

Nucleotide sequences were assembled and analysed using the University of Wisconsin GCG package running on a UNIX computer. Protein sequences were compared with each other using the GAP program running on the standard default penalties to obtain percentage identities and similarities. Multiple sequence analysis and alignment were carried out using the PILEUP and PRETTY programs in the GCG package.

Cloning C termini of TGN38-related sequences

Total RNA was extracted from mouse (NIH3T3), hamster (CHO) and human (HeLa) cells using a guanidium isothiocyanate-based kit (Pharmacia, Milton Keynes, UK); rabbit liver RNA was kindly provided by Dr N. Price (Biochemistry Dept, Bristol). cDNA synthesis was carried out with random hexanucleotide primers and a kit (Pharmacia) and used in a PCR reaction with rat-specific oligonucleotide primers (GB15 and GB17; Reaves et al., 1992) to enable amplification of the C terminus of TGN38 homologues. Cycling parameters were as follows: 94°C, 1 minute; 55°C, 2 minutes; 72°C, 2 minutes; 25 cycles. Resulting PCR fragments were subcloned into the PCRII plasmid (Invitrogen, San Diego, CA) and clones from at least two independent primary PCR reactions were sequenced in each case.

Southern and northern blotting analyses

Tissue from human stomach, macaque testis and rat lung was incubated in 20 mM Tris-HCl, pH 7.6, 1 mM EDTA, 2% SDS and 0.5 mg/ml proteinase K for 1 hour at 42°C. Total nucleic acid was extracted with equal volumes of phenol/chloroform, and ethanol precipitated several times before contaminating RNA was removed by treatment with 6 units/ml of an equivalent mixture of RNase A and RNase T1. A 10 μg sample of purified genomic DNA was digested to completion with *Eco*RI, *Kpn*I or *Hin*dIII and resolved on 0.8% gel in 1× TAE buffer and transferred to GeneScreen Plus membranes (DuPont-NEN, Stevenage, UK). ^{32}P-labelled probes were generated by random priming of gel-purified full-length cDNA inserts of human TGN46 and rat TGN41 or from a 3′ end coding fragment of macaque TGN47 using standard procedures (Sambrook et al., 1989).

A human poly(A)$^+$ northern blot was purchased from Clontech (Palo Alto, CA). Macaque testis and rat lung total RNA were prepared as described by Girotti et al. (1992). Poly(A)$^+$ mRNA was purified using a kit (Pharmacia, Milton Keynes, UK), separated on a 1.1% agarose/6.5% formaldehyde gel and blotted onto Hybond N membrane (Amersham, UK). ^{32}P-labelled probes were generated from full-length cDNA inserts of human TGN46, macaque TGN47 and rat TGN41 as described above.

Antibody production and purification

A peptide with the sequence VPLLATESVKQEEAGVRPC (residues 18-35 of human TGN46 + a cysteine residue) was coupled to BSA via its C-terminal cysteine residue using a bifunctional cross-linker (*m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester; Sigma) and used to immunise rabbits using standard procedures (Harlow and Lane, 1988). The resulting polyclonal antisera were designated GB1 and GB2. Immunoglobulins were precipitated with 50% ammonium sulphate, resuspended in buffer and incubated for 16 hours (at 4°C) with 1 ml of Sepharose 4B beads (Pharmacia, UK) coupled to the peptide. Beads were packed into a column and washed with 10 volumes of PBS and coupled antibodies were eluted with 100 mM glycine, pH 2.3. Aliquots with A_{280} >0.1 were pooled and concentrated to 1 mg/ml. A second peptide with the sequence QTPKDSP-

SKSSAEAQTPEDTPNKSGAEAC (residues 61-88 of human TGN46 + a cysteine residue) was coupled to keyhole limpet haemocyanin and used to produce rabbit polyclonal antibodies. This antiserum was designated as VP5.

Western blotting, cell culture and indirect immunofluorescence microscopy

HeLa cell lysates were prepared and 50 μg of lysate was run on an 8% SDS-PAGE gel and western blotted as described by Brake et al. (1990). Blots were probed with affinity-purified GB1 antiserum in the absence or presence of the antigenic peptide used for antibody production. Detection was done using anti-rabbit alkaline phosphatase-conjugated secondary antibody. HeLa cells were grown and transfected as previously described (Nilsson et al., 1989) and generally processed for immunofluorescence microscopy by methanol fixation (Reaves et al., 1993). However, for detection of internalised labelled transferrin, cells were fixed in paraformaldehyde and permeabilised with TX-100 before processing for immunofluorescence microscopy (Ponnambalam et al., 1994).

Rabbit antiserum were diluted in 0.2% fish skin gelatin/PBS (1:50 for VP5) or (1:100 for GB1) before use. The GTL2 monoclonal antibody was further purified over a MonoQ column using FPLC (Kawano et al., 1994) and used at 5 μg/ml final concentration. Purified OKT9 monoclonal antibody against the human transferrin receptor was a used at 1:100 dilution and was a kind gift from Dr Colin Watts (Biochemistry, Dundee). Secondary antibodies were horse anti-mouse conjugated to Texas Red or FITC and goat anti-rabbit conjugated to FITC or Texas Red (Vector Laboratories, Peterborough, UK) and both were used at a dilution of 1:100. Slides were examined on a Zeiss Axiophot fluorescence microscope using a 63× objective oil immersion lens and a 10× eyepiece. Cells were further visualised using a Bio-Rad MRC-600 confocal laser microscope (Bio-Rad, Hemel Hempstead, UK). Bleed-through from the fluorescein channel into the Texas Red channel was corrected for by collecting both images separately. Final images were compared using the MRC-600 software.

Chloroquine (Sigma; Poole, UK) was used at a final concentration of 100 μM for 3 hours at 37°C. In antibody internalisation experiments, VP5 rabbit antisera was diluted between 1:10 and 1:100 in DMEM/10% FCS and incubated with or bound to cells in the absence or presence (greater than 50-fold molar excess) of competing peptide. Antibody was internalised by warming cells up to 37°C. Rhodamine-transferrin conjugate (Molecular Probes, OR) was used at 20 μg/ml final concentration in DMEM/10% FCS. Quantitation of overlapping patterns was carried out after printing negatives at 1,000× magnification and counting the number of transferrin and TGN46 positive punctate structures in 10 cell profiles.

RESULTS

Cloning and structure of rat TGN38 homologues

A 2.2 kb cDNA encoding the human homologue of rat TGN38 cDNA was isolated from a foetal human liver cDNA library (see Materials and Methods). This cDNA contained a single open reading frame (ORF) with strong homologies to the N-terminal signal peptide and the C-terminal domains of rat TGN38 (Fig. 1). This ORF encodes a polypeptide of 437 residues (Fig. 1) and was designated human TGN46 (humTGN46) following the current nomenclature for these mammalian glycoproteins (Luzio et al., 1990; Reaves et al., 1992).

Comparison of human TGN46, rat TGN38 and the recently cloned mouse alleles (TGN38a and TGN38b) showed highly conserved domains; however, a unique region in the human

```
        1                                                          50
HumTGN46  MRFVVALVLL NVAAAGAVPL LATESVKQEE AGVRPSAGNV STHPSLSQRP
MacTGN47  ---------- S------A-- SD-A---K-- ---Q--TE-I --Y-TS----
MmTGN38a  ---Q---L-- S--V-R-L-* ****--YKRD -D******* *****SGDSQ
MmTGN38b  ---Q---L-- S--V-R-L-* ****P-YKRD -D******* *****SGDSQ
RatTGN38  -Q-L---L-- S--V-R-L-* ****-ASKPN NT******** *****S-E**

        51                                                         100
HumTGN46  GGSTKSHPEP QTPKDSPSKS SAEAQTPEDT PNKSGAEAKT QKDSSNKSGA
MacTGN47  ------D--- H-SA----R- DP-P--SK-S -S------Q- PE-NP-----
MmTGN38a  NPPNQPSKQS S--LP****- -NQVK-TRP- ***D-QGQKS D-KDQD-T**
MmTGN38b  NPPNQPSKQS S--LPP**E- -NQVK-TRP- ***D-QGQKS D-KDQD-T**
RatTGN38  ***NNPPIQ- S--LPPGVDI -QQVK-NRP- ***D*QRLES D-EGQD-TV-

        101                                                        150
HumTGN46  EAKTQKGSTS KSGSEAQTTK DSTSKSHPEL QTPKDSTGKS GAEAQTPEDS
MacTGN47  ------D-S- -------I-- --S---GA-A K-Q---SS-- -S---ITK--
MmTGN38a  *TLAAVS-KA E--PRTAA-D H-LGD-RRQP EK******** *TD-ELN-TA
MmTGN38b  *TLAAVS-KA E--PPTAA-D H-LGD-RRQP EK******** *TD-ELK-TA
RatTGN38  RTSASVS-GV ES*****A-N LNLDDSKKHP ET******** *AD-KLK-TL

        151                                                        200
HumTGN46  PNRSGAEAKT QKDSPSKSGS EAQTTKDVPN KSGADGQTPK ********DG
MacTGN47  SSK------- ---------- -T-----STS ----EAP-T- DVPNKSGA--
MmTGN38a  RPL-PVNP-L E-*******- DQSS-E-SGK PT-GNS**** **********
MmTGN38b  RPL-PVNP-L E-*******- DQSS-E-SGK PT-GNS**** **********
RatTGN38  QQLLPVDP-Q E-*******- GQKF-K-SGS PT-G-S**** **********

        201                                                        250
HumTGN46  SSKSGAEDQT PKDVPNKSGA EKQTPKDGSN KSGAEEQGPI DGPSKSGAEE
MacTGN47  ---------I ---------- -------AA- ---------- --S---D---
MmTGN38a  ********** ********** G-P-GG-SG* *******K-T EAG-NKAT-D
MmTGN38b  ********** ********** G-P-GG-SGK PT-GDSDK-T EAG-NKAT-D
RatTGN38  ********** ********** DNT-GG-SNK TT-VDSDKTS G-D-NKPTGS

        251                                                        300
HumTGN46  QTSKDSPNKV VPEQPSWKDH SKPISNPSDN KELPKADTNQ LADKGKLSPH
MacTGN47  --P--G-S-- ------R--- ---V------ ---------- S---------
MmTGN38a  DSG-****** ****STKV-L D--T-KISPD T-TS-T-KV- PTE--QKPTL
MmTGN38b  DSG-****** ****STKV-L D--T-KIFPD T-TS-T-KV- PTE--QKATL
RatTGN38  DND-****** ****PTGG-S N--T-KVPS- T-T--I-KV- LTE--QKPTL

        301                                                        350
HumTGN46  AFKTESGE** **ETDLISPP QEEVKSSEPT EDVEPKEAED DDTGPEEGSP
MacTGN47  --------** **D-----S- ---------- ---------- ----------
MmTGN38a  TS------TL AGDS-FSLK- EKGD------ ----T--I-E G--E------
MmTGN38b  TS------TL AGDS-FSLK- EKGD------ ----T--I-E G--E------
RatTGN38  IS------KL AGDS-FSLK- EKGD------ ----T--I-E G--E------

        351                                                        400
HumTGN46  PKEEKEKMSG SASSENREGT LSDSTGSEKD DLYPNGSGNG SAESSHFFAY
MacTGN47  ---------- ---N------ -L--R----- -R-RES---A ----------
MmTGN38a  LE--N--VP- PS----Q--- -T--MKN--- -L-KDS---T ----------
MmTGN38b  LE--N--VS- PS----Q--- -T--MKN--- -L-KDS---T ----------
RatTGN38  LE--N--VL- PS----Q--- -T--MKD--- -H-KDN---T ----------

        401                                                        450
HumTGN46  LVTAAILVAV LYIAHHNKRK IIAFVLEGKR SKVTRRPKAS DYQRLDQKS
MacTGN47  ---------- ---------- ---------- ---------- ---------
MmTGN38a  -----V---- ----Y----- ----A----- ---------- -----NL-L
MmTGN38b  -----V---- ----Y----- ----A----- ---------- -----NL-L
RatTGN38  -----V---- ----Y----- ----A----- ---------- -----NL-L
```

Fig. 1. Comparison between TGN38 homologues. Species TGN38 homologues are designated by 2 or 3 letters followed by 3 upper case letters that indicate the primary intracellular location of the protein (i.e. TGN) and the predicted size of the protein (to the nearest kDa) from the ORF (including the signal peptide). Thus human TGN46 and rat TGN38 are HumTGN46 and RatTGN38, respectively. The exceptions are the mouse alleles (MmTGN38a and MmTGN38b), which are defined according to Kasai et al. (1995). Protein sequences of macaque TGN47 (macTGN47; 445 residues), MmTGN38a (353 residues), MmTGN38b (363 residues), and RatTGN38 (357 residues) are compared with the HumTGN46 sequence. Amino acid identities are marked with a dash (-) and gaps with an asterisk (*). Protein sequences were aligned as described in Materials and Methods. Boundaries between the domains I-V are indicated by the arrowheads above the human sequence (see Results for further details). The nucleotide sequences of the human TGN46 and macaque TGN47 cDNAs have been deposited in the EMBL/GenBank database with X94333 and X94334 accession numbers, respectively.

lumenal domain had little homology with the equivalent region in either the rat protein or the mouse alleles (see Fig. 1). To check whether there were splice variants of human liver TGN46 cDNA, we screened other human cDNA libraries. The nucleotide sequences of isolated cDNA clones were compared with that of human liver TGN46. We obtained two partial clones from a foetal human thymus cDNA library, each of which contained approximately 95% and 50% of the human liver TGN46 ORF shown in Fig. 1. The protein-coding sequences in these foetal thymus cDNAs were identical to the foetal liver sequence (data not shown). These data strongly suggest that the human liver TGN46 cDNA encodes the human homologue of rat TGN38 and no other human protein isoforms are produced. The sequences of the other cDNA clones also excluded possibilities that the liver TGN46 cDNA was an artefact of cDNA synthesis or rearrangement of cloned mammalian cDNAs during propagation in *Escherichia coli.*

To investigate the extent and nature of the protein sequence conservation between the rodent and primate proteins, we isolated and sequenced monkey homologues of rat TGN38. A macaque testis cDNA library was screened using an oligonucleotide complementary to the C terminus coding region of rat TGN38 (see Materials and Methods). A single cDNA clone was used to isolate a further three independent cDNA clones from a macaque epididymal cDNA library; all four cDNA clones were identical throughout their overlapping sequences. Three of the four clones contained mostly 3′ untranslated sequences; the fourth cDNA clone contained the majority of the protein coding sequence. However, sequence analysis revealed that by comparison to the human TGN46 cDNA approximately 300 bp from the 5′ end of this macaque cDNA was missing. This region was cloned by RT-PCR (see Materials and Methods). The predicted macaque protein sequence reveals an ORF of 445 amino acids that encodes a protein of approximately 47 kDa (Fig. 1). Importantly, macaque TGN47 is more highly related to human TGN46 than to either the rat or mouse TGN38 homologues (Fig. 1; Table 1). In fact, the region of difference between the rodent and human proteins is highly conserved between human TGN46 and macaque TGN47. No other TGN47 isoforms or splice variants were isolated from the macaque cDNA library, suggesting that both the human and macaque proteins are the sole homologues of rat TGN38.

We analysed the genomic complexity and mRNA expression from genes encoding rat TGN38, human TGN46 and macaque TGN47 in rat, human and monkey cells. Southern blot analysis was used to identify genomic sequences in all three species that hybridised to these cDNAs (Fig. 2A). Restriction digests of genomic DNAs probed with the cognate ^{32}P-labelled cDNAs revealed a simple pattern of bands. In all three species, a single restriction fragment was detected. These data indicate that the rat TGN38, human TGN46 and macaque TGN47 cDNAs are encoded by a single copy locus in each mammalian species.

Analysis of the mRNA transcripts produced in each species

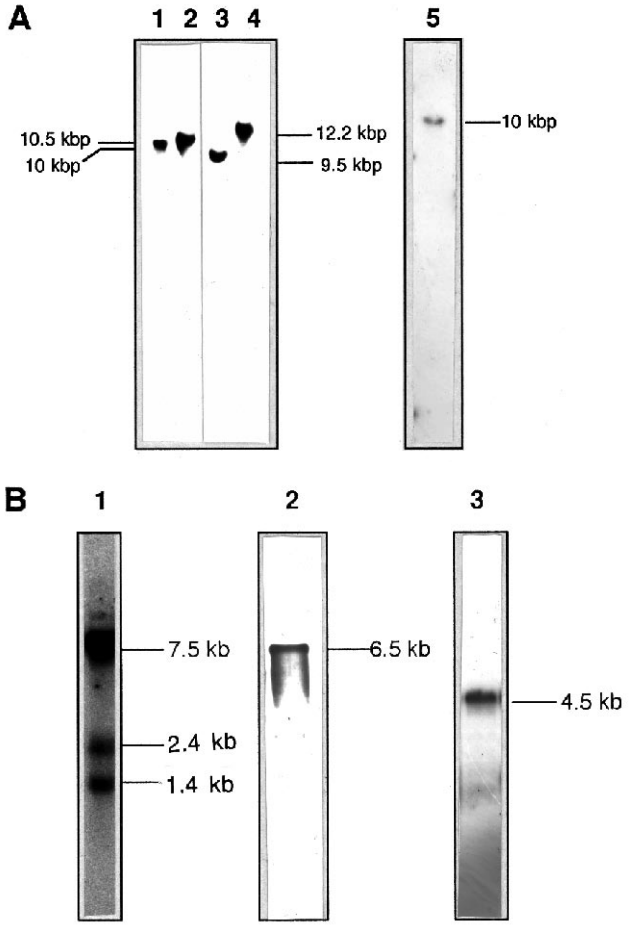
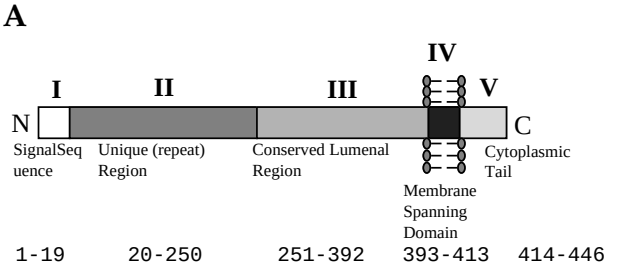


Fig. 2. Analysis of gene organisation and mRNA expression of TGN38 homologues. (A) Southern blot analysis of human genomic DNA digested with *Eco*RI (lane 1) or *Kpn*I (lane 2); monkey genomic DNA digested with *Eco*RI (lane 3) or *Kpn*I (lane 4); and rat genomic DNA digested with *Hin*dIII (lane 5). (B) Northern blot analysis on poly(A)$^+$ mRNA from human liver (lane 1), macaque testis (lane 2) and rat lung (lane 3) poly(A)$^+$ mRNA. Cognate ^{32}P-labelled probes were used for both Southern and northern analyses.



B

Human	FVLEGKRSKVTRRPKASDYQRLDQKS
Macaque	----------
Rabbit	-A------R--------E-----L-L
Rat (TGN38)	-A----------NL-L
Hamster	-A----------NL-L
Mouse	-A----------NL-L

Fig. 3. Domain analysis of TGN38 homologues. (A) TGN38 homologues contain 5 domains: (I) the signal sequence; (II) the unique repeat domain; (III) the conserved region of the lumenal domain; (IV) the membrane spanning domain; and (V) the cytoplasmic tail. The residues (1-446) that comprise each domain of the human protein are indicated below each domain. (B) Comparison of the amino acid sequences of the cytoplasmic domains of the TGN38 homologues from different mammalian species.

was performed by northern blotting. Purified poly(A)$^+$ mRNA was hybridised to the cognate ^{32}P-labelled cDNA probe (Fig. 2B). A specific transcript of 4.5-5 kb that hybridised to the rat TGN38 cDNA was detected in rat liver (Fig. 2B, lane 3). A transcript of the same size is also detected in rat heart, lung, brain, testis and pancreas (data not shown). Relatively large mRNAs of approximately 7.5 kb in human liver (Fig. 2B, lane 1) and 6.5 kb in macaque testis (Fig. 2B, lane 2) were detected by northern blotting analyses.

Smaller transcripts of 2.4 kb and 1.4 kb were also detected in human liver (Fig. 2B, lane 1). Sequences of human TGN46 cDNA clones from different human cDNA libraries (S. Ponnambalam and G. Warren, unpublished observations) and comparison with the rat TGN38 cDNA nucleotide sequence showed that the coding region of the rat and human transcripts did not exceed 1.3 kb yet the mRNA transcripts were between 4.5 kb to 7.5 kb in size. This implies the presence of extensive amounts of either 5′ and/or 3′ untranslated nucleotide sequence in each of these species, with significantly more in primates than in rodents. It is of note that whilst little 5′ untranslated sequence has been found in mRNAs of the TGN38 homologues from all three species, significant amounts (~1.6 kb in rat, ~3.5 kb in macaque and ~1.5 kb in human) of 3′ untranslated region sequences have been cloned. Thus the different transcripts detected in human liver that hybridise to the TGN46 probe are probably derived from the alternate splicing of the 3′ untranslated region of the primary RNA transcript. Screening both poly(dT) and random-primed cDNA libraries produced cDNA clones with the same ORF (see Materials and Methods). Surprisingly, there is greater nucleotide conservation in the 3′ untranslated regions of rat TGN38 with human TGN46 or macaque TGN47 mRNAs than in the 5′ coding regions of the cDNAs (M. Girotti, S. Ponnambalam and G. Banting, unpublished observations). No significant sequence homology was found between the 3′ untranslated regions and any sequences in the database.

Domains of human TGN46 and macaque TGN47

Each TGN38 homologue can be represented by five domains (Fig. 3A). Domain I contains the signal peptide and the region following it; domain II contains the unique region that is divergent between primate and rat proteins. Domain III is a conserved region of the lumenal domain; IV and V are the membrane spanning and cytoplasmic domains, respectively. The human and macaque polypeptides show 82% overall identity (87% similarity; see Table 1). In contrast, the rodent TGN38 homologues have less than 50% overall identity (about 60% similarity) to either human TGN46 or macaque TGN47 (Table 1). However, when each domain in these homologues is analysed separately, the percentage of homologies differs widely. In domain I, the 19 residue signal peptide sequence has greater than 60% identity between all the homologues (not shown). Homologies and features of the unique region (domain II) are discussed in the next section. Domain III consists of approximately 130 residues before the membrane spanning domain in the primate proteins and displays 65% identity with the equivalent region in rodent TGN38. These percentage homologies are similar to those values for rat and human homologues of other integral membrane proteins found in the EMBL/GenBank database.

The membrane spanning (IV) and cytoplasmic (V) domains are almost identical between the primate proteins and rodent TGN38 (Fig. 1). Comparison of the sequences of the cytoplasmic domains of TGN38 homologues from different species reveals a high degree of conservation (Fig. 3B). The SDYQRL motif implicated in retrieval to the TGN (Wong and Hong, 1993) is conserved in all species examined; the rabbit TGN38 homologue contains a conserved substitution of glutamic acid for aspartic acid within this motif (Fig. 3B). The only notable difference between the cytoplasmic domains of these different homologues are at the extreme C terminus where three out of four of the C-terminal amino acids differ.

Repeat domain in human TGN46 and macaque TGN47

A specific region, designated domain II (Fig. 3A), located just downstream of the signal peptide in human TGN46 and

Table 1. Amino acid sequence comparison between rat tgn38 and its primate homologues

Homologies	HumTGN46	MacTGN46	RatTGN38	MmTGN38a	MmTGN38b
Full length					
HumTGN46	–	82 (87)	49 (60)	48 (62)	48(62)
MacTGN46	82 (87)	–	48 (58)	49 (61)	48(62)
RatTGN38	49 (60)	48 (58)	–	75 (82)	76 (83)
MmTGN 38a	48 (62)	49 (61)	75 (82)	–	98 (99)
MmTGN38b	48 (62)	47 (62)	76 (83)	98 (99)	–
Conserved					
HumTGN46	–	93 (96)	66 (72)	65 (72)	65 (72)
MacTGN46	93 (96)	–	65 (72)	65 (72)	65 (72)
RatTGN38	66 (72)	–	–	90 (93)	90 (93)
MmTGN38a	65 (72)	65 (72)	90 (93)	–	98 (99)
MmTGN38b	65 (72)	65 (72)	90 (93)	98 (99)	–
Non-conserved					
HumTGN46	–	74 (80)	20 (36)	26 (45)	25 (45)
MacTGN46	74 (80)	–	25 (38)	23 (42)	18 (38)
RatTGN38	20 (36)	24 (38)	–	53 (66)	56 (68)
MmTGN38a	26 (45)	23 (42)	53 (66)	–	98 (99)
MmTGN38b	25 (45)	18 (38)	56 (68)	98 (99)	–

% Identity and similarity (in parentheses) between any two TGN38 family members. Homologies between any two pairs of TGN38 homologues over the full length of each protein, conserved regions or non-conserved is tabulated. The conserved regions are: human (251-437), macaque (259-446), rat (174-357), mouse TGN38a (169-353) and mouse TGN38b (179-363). Non-conserved regions are: human (20-250), macaque (20-258), rat (20-173), mouse TGN38a (20-168) and mouse TGN38b (20-178). Sequence comparisons were carried out as described in Materials and Methods.

A

Repeat Number	Human Repeats	Macaque Repeats
1	TKSHPEPQTPKDSP	TKSDPEPHTSADSP
2	SKSSAEAQTPEDTP	SRSDPEPQTSKDSP
3	NKSGAEAKTQKDSS	SKSGAEAQTPEDNP
4	NKSGAEAKTQKGST	NKSGAEAKTQKDSS
5	SKSGSEAQTTKDST	SKSGSEAQITKDSS
6	SKSHPELQTPKDST	SKSGAEAKTQKDSS
7	GKSGAEAQTPEDSP	SKSGSEAQITKDSS
8	NRSGAEAKTQKDSP	SKSGAEAKTQKDSP
9	SKSGSEAQTTKDVP	SKSGSETQTTKDST
10	NKSGADGQTPKDGS	SKSGAEAPTTKDVP
11		NKSGAD......GS
12	SKSGAEDQTPKDVP	SKSGAEDQIPKDVP
13	NKSGAEKQTPKDGS	NKSGAEKQTPKDAA
14	NKSGAEEQGPIDGP	NKSGAEEQGPIDGS
15	SKSGAEEQTSKDSP	SKSDAEEQTPKDGP
Cons.	SKSGAEAQTPKDSP	SKSGAEAQTPKDSS
	N T(S)N P	

B

Heptad vertex no.	HumTGN46	MacTGN47
1	TQSQNKNKSQSQGQNKSQNQSQNQNQSQ	THSQSQNKSQSKSQSKSQSPNSDPKAESEP
2	KTKTKTKTKTKTKTRTKTKTKTKTKGKT	KTRTKTKTKIKTKIKTKTKTKSQNQNQSQ
3	SPSPSQSQSTSPSPGQSTSPSPSPSSS	SSSSSPSQSTSQSTSQSTSTSKIKTKGKT
4	HKSEGKGKGKHKGESKGKGKGKGKGIGK	DADKGEGKGKGKGKGKGKGKGSPSPSPSP
5	PDADADAGSDPDADADSDADADADADAD	PDPDADADSDADSDADSDADAGKGKGIDK
6	ESETESESESESESESEVDGEVEGEGES	ESESENESESESESESESEVDADADADAD
7	PPAPASATATLTAPAPAPGSDPKSEPEP	PPPPAPASASASASAPTTAPGEVEAEGEG

Fig. 4. Primate repeat analysis. Repeating amino acid motifs in (A) human TGN46 (from +54 to +249) and macaque TGN47 (from +54 to +257) were aligned and a consensus sequence displayed for each set of repeats. (B) The human and macaque tandem repeats were analysed by plotting the region of tandem repeats vertically onto positions 1-7 of a heptad repeat such as those found in coiled-coil domains. Charged residues on adjacent vertices are underlined.

macaque TGN47, shows significant differences to that of rat TGN38 (see Fig. 1). This region in human TGN46 (from +20 to +250) and macaque TGN47 (from +20 to +258) only exhibits 20-26% overall identity in comparison to the equivalent region in rodent TGN38 (see Table 1). These values are slightly above the threshold value of 15-20% identity for two unrelated proteins from human and rat species. The equivalent regions in rodent TGN38 homologues contain either 6 (rat and mouse TGN38a) or 7 (mouse TGN38b) direct tandem repeats of an 8mer (Luzio et al., 1990; Kasai et al., 1995). In contrast, human TGN46 contains 14 direct repeats of a 14mer sequence (Fig. 4A). The equivalent region in macaque TGN47 contains 14 complete repeats of the 14mer sequence and one partial repeat. To ensure that the half-repeat insertion in macaque TGN47 (residues 191-198, Fig. 4A) was not due to a cloning artefact, PCR analysis of macaque mRNA using *Pfu* DNA polymerase was carried out (see Materials and Methods). The PCR amplification product did not show any difference from the original macaque cDNA clone (data not shown).

Alignment of the 14mer direct repeats in the extracellular/lumenal domain of the human and macaque TGN47 proteins gives an almost identical consensus sequence (Fig. 4A). The consensus sequence consists of a core sequence (KSGAEAQTPKDS) flanked by a more variable residue (S, N, T or P) on both sides (see Fig. 4A). The number of residues (14) within each repeat raised the possibility that they might form a helical or coiled-coil structure consisting of repeating 7mer motifs. The sequence of this region in the human and macaque proteins was plotted out on vertices of a heptad repeat such as those found in coiled-coil domains. Positions 1-7 of the vertices are numbered vertically in Fig. 4B. Surprisingly, adjacent vertices at positions 2 and 6 formed a highly charged face. Vertex 2 is made up of a series of alternating lysine and threonine residues (Fig. 4B). In contrast, vertex 6 is made up of alternating serine and glutamic acid residues (Fig. 4B). Note that in macaque TGN47, the half-repeat insertion after the tandem repeat number 10 (Fig. 4A) results in a shift of adjacent charged vertices to positions 3 and 7 (Fig. 4B).

Localisation of human TGN46

Two polyclonal antisera (GB1 and VP5) were raised against peptides corresponding to a unique sequence just downstream of the signal peptide and within the repeat domain of human TGN46, respectively (see Materials and Methods). These antibodies did not react with rat TGN38 as assessed by immunofluorescence. Both antibodies recognised a protein of ~150 kDa by western blotting and this was blocked by pre-incubation with their respective peptides (data not shown). Immunofluorescence analysis of HeLa cells using the VP5 (Fig. 5A) and GB1 (Fig. 5B) antisera produced a juxtanuclear staining pattern characteristic of the Golgi apparatus. These staining patterns were abolished by prior incubation of the antibodies with the respective peptides (data not shown). Similar results were obtained when the GB1 antiserum was used in immunofluorescence analysis of monkey (COS) cells (data not shown).

HeLa cells were simultaneously stained with VP5 polyclonal antibodies to human TGN46 (Fig. 5C) and a monoclonal antibody to β(1,4)-galactosyltransferase (GalT) (Fig. 5D). The monoclonal antibody, GTL2, recognises a conserved epitope present on both rat and human GalT (Kawano et al., 1994). This Golgi-resident enzyme is located in the *trans*-Golgi cisterna and *trans*-Golgi network in HeLa cells (Nilsson et al., 1993). The staining pattern for human TGN46 (Fig. 5C) co-localised with that for GalT (Fig. 5D).

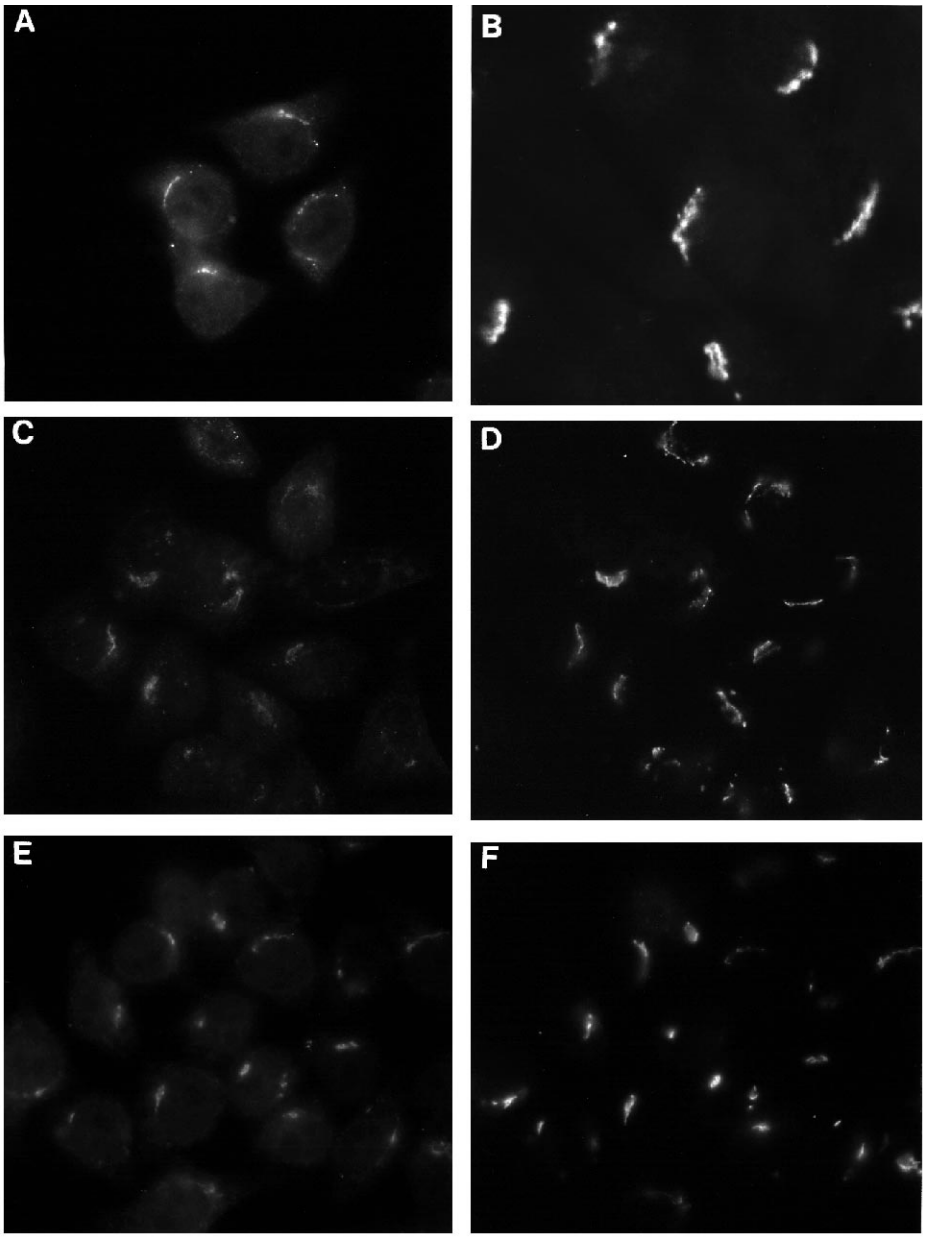


Fig. 5. Location of TGN46 in HeLa cells. Parental HeLa cells (A-D) or the SA:48 cell line stably expressing a VSV-G epitope-tagged SialylT (E, F) were fixed, permeabilised and labelled with peptide antisera VP5 (A) or GB1 (B) against human TGN46 or double-labelled with the VP5 antisera to human TGN46 (C,E) and either the GTL2 mAb to GalT (D) or the P5D4 mAb to the epitope-tag on SialylT (F). Bound antibodies were visualised using secondary antibodies coupled to FITC (A-C,E) or rhodamine (D,F). ×800.

A stably transfected HeLa cell line (SA:48; see Materials and Methods for description) that expresses an epitope-tagged sialyltransferase (SialylT) was stained for TGN46 (Fig. 5E) and SialylT (Fig. 5F). This epitope-tagged enzyme is located in the *trans*-Golgi cisterna and TGN (Rabouille et al., 1995). The staining pattern for TGN46 coincided with that for SialylT.

Recycling of human TGN46

Two approaches were used to show that human TGN46 recycles between the TGN and the plasma membrane in the same manner as rat TGN38. The first was to bind antibodies to the lumenal domain as it appeared at the cell surface after leaving the TGN. HeLa cells were incubated for 3 hours at 37°C with diluted VP5 antiserum, washed and then processed for immunofluorescence microscopy using only secondary antibodies conjugated to FITC. As shown in Fig. 6A, the antibody was found in a juxta-nuclear reticulum that partly coincided with GalT (Fig. 6B) and in punctate structures throughout the cell cytoplasm. This pattern was not the consequence of fluid-phase endocytosis of free antibodies, since similar results were obtained when the cells were pre-incubated at 4°C with the VP5 antiserum and then washed before internalisation was initiated by shifting the temperature to 37°C (data not shown). In addition, no staining was observed when the antiserum was pre-treated with the peptide to which it was made (data not shown).

The punctate structures containing the anti-VP5 antibody varied in size from 300 to 500 nm as measured by confocal laser microscopy. They were identified using rhodamine-transferrin as a marker for the early part of the endocytic pathway (Hopkins et al., 1990). Between 40 and 50% of these punctate structures also contained rhodamine-transferrin (arrowheads; Fig. 6C and D) strongly suggesting that endosomes are inter-

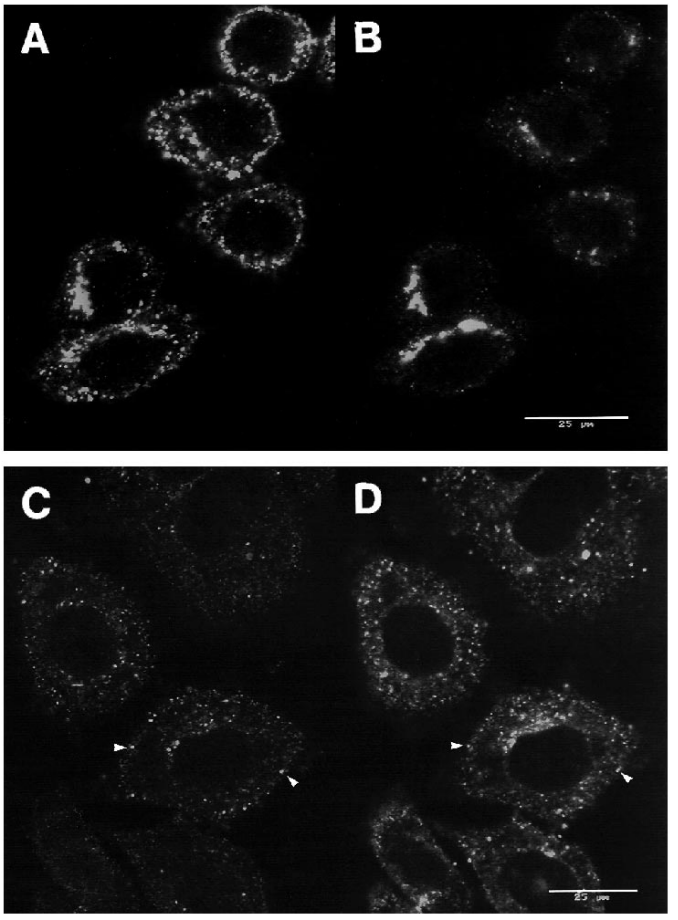


Fig. 6. Internalisation of antibodies to human TGN46. HeLa cells were incubated with VP5 antisera to TGN46 in the absence (A,B) or presence of rhodamine-labelled transferrin (C,D) for 3 hours and visualised by confocal microscopy. After fixation and permeabilisation, VP5 antibodies were detected by secondary antibodies coupled to FITC (A,C). Cells shown in A and B were double-labelled with the GTL2 monoclonal antibody to GalT, which was detected using secondary antibodies coupled to Texas Red (B). Rhodamine-transferrin was detected directly (D). Arrowheads (C,D) denote co-localisation of punctate structures. Bars in B and D, 25 μm.

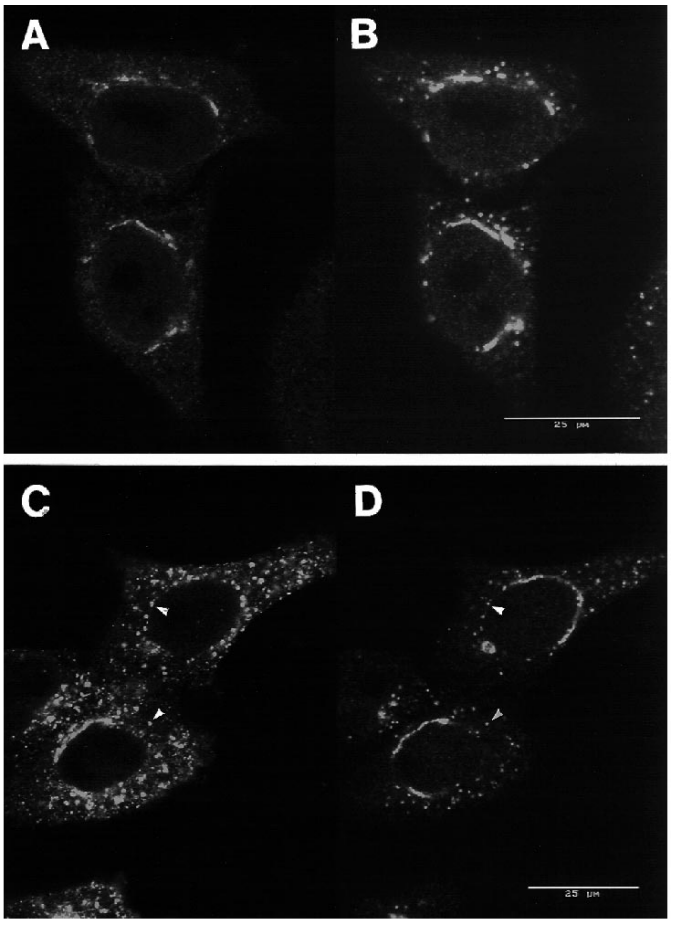


Fig. 7. Effect of chloroquine on the distribution of human TGN46. HeLa cells were treated with chloroquine for 3 hours and visualised by confocal microscopy. GalT (A), human TGN46 (B,D) or human transferrin receptor (C) were detected using either the GTL2 (A), VP5 (B,D) or OKT9 (C) antibodies, respectively. Bound mouse or rabbit antibodies were detected by secondary antibodies conjugated to FITC (A, C) or Texas Red (B,D), respectively. Arrowheads (C,D) denote co-localisation of punctate structures. Bars in B and D, 25 μm.

mediates on the pathway taken by human TGN46 from the plasma membrane back to the TGN.

The second approach used to show recycling of human TGN46 was to treat HeLa cells with chloroquine. This lysosomotropic drug blocks transport of rat TGN38 from endosomes back to the TGN (Chapman and Munro, 1994; Reaves and Banting 1994a). As shown in Fig. 7B, after a 3 hour treatment with 0.1 mM chloroquine, human TGN46 was still found to some extent in the Golgi apparatus, as shown by its co-localisation with GalT (Fig. 7A), but a substantial amount was now found in punctate, cytoplasmic structures (Fig. 7B). These were identified by carrying out the same experiment but double-labelling the cells for TGN46 and the transferrin receptor after fixation and permeabilisation. Comparison of the staining pattern for the transferrin receptor (Fig. 7C) and human TGN46 (Fig. 7D) showed that up to 50% of the TGN46 was present in structures that also contained the transferrin receptor.

DISCUSSION

The protein sequences of the human and macaque homologues of rat TGN38 are highly conserved (>80% identity; see Table 1) in four out of the five domains. Domains I (signal sequence) and III (a lumenal region) exhibit greater than 60% identity between the rat, mouse and primate proteins. Domain IV (membrane spanning) contains only one conserved substitution between the rodent and primate proteins (Val to Ile), indicating conservation of the TGN localisation signal in this region. Domain V (cytoplasmic tail) is also highly conserved between different mammals with the tyrosine-based TGN retrieval signal almost identical in all species examined.

Both human TGN46 and macaque TGN47 proteins contain a unique region in the lumenal domain when compared with the rodent homologues (Fig. 1). This region is located just downstream of the signal peptide and is composed of a series of 14mer direct tandem repeats. The equivalent region in rodent TGN38 is composed of a series of 8mer direct tandem

repeats. This region in rodent and primate proteins displays 20-26% identity (Table 1). Such values are in a 'twilight zone' for relatedness between two protein sequences (S. Subbiah, personal communication). These differences between the rodent and primate proteins cannot be explained by the existence of more than one primate isoform. The sequences of two TGN46 human thymus cDNAs were identical to the human liver sequence. More than 20 clones were isolated from another cDNA library and the 5′ sequences were identical to those of human liver TGN46. Screening of macaque epididymal and testis cDNA libraries from non-siblings yielded four overlapping cDNAs. Genomic Southern blot analysis indicates that rat TGN38 and the human and macaque homologues are encoded by a single genetic locus and precludes the possibility that the sequences presented here are products of processed pseudogenes or individual members of an extended gene family within each species.

The mouse TGN38 alleles are also encoded by a single genetic locus (Kasai et al., 1995). The amino acid differences between the two mouse alleles are all found in the extracellular domain (Fig. 1; Kasai et al., 1995). Eleven nucleotide point substitutions scattered over the coding region of the two alleles result in six amino acid changes of which five are silent. There are also two deletions of 6 and 24 bp, respectively, in the extracellular/lumenal domain of mouse TGN38a (see Fig. 1). These allelic differences within the mouse species highlight the divergence of the amino acid sequence of the extracellular domain in all four mammalian species. All available evidence points to a single gene encoding the TGN38 proteins expressed in mammals.

Relatively large mRNAs (4.5-7.5 kb) are transcribed from each gene. The ORFs of the TGN38 homologues from rat and primates are 1.1 and 1.3 kb, respectively. The large mRNA size is due to a long 3′ untranslated region of 3-6 kb. Three transcripts of differing sizes (4.5 kb, 3.3 kb and 1.8 kb) are also detected in mouse cells and tissues and these differences have been ascribed to the length of the 3′ non-coding region (Kasai et al., 1995). Although the 3′ untranslated region of these mammalian mRNAs have not revealed significant homologies with sequences in the database, marked conservation of nucleotide sequences within this region (M. Girotti et al., unpublished observations) indicate that it may be functionally important for mRNA export or translation.

Despite these sequence differences, the membrane trafficking properties of the molecules appear to be the same. This is expected as the sequences of the membrane spanning and cytoplasmic domains are almost identical between the primate and rodent proteins (see Fig. 1). These highly conserved domains contain the TGN retention and retrieval signals, respectively. Immunofluorescence studies show that the human TGN46 localises to a juxtanuclear reticulum in HeLa and COS cells. In HeLa cells it co-localises with both GalT and SialylT, more than 70% of which is present in the TGN (Nilsson et al., 1993; Rabouille et al., 1995). This provides good evidence that human TGN46 is also resident in the TGN. Additionally, expression of human TGN46 in rat NRK cells results in targeting of the protein to a juxtanuclear compartment resembling the Golgi apparatus (S.P. and G.W., unpublished data). We had previously shown that rat TGN38 had the correct location in HeLa cells (Ponnamabalam et al., 1994; Rabouille et al., 1995). Thus, TGN localisation signals appear to function correctly in heterologous cell lines.

Two lines of evidence show that human TGN46 recycles between the TGN and the cell surface. Firstly, antibodies to the extracellular/lumenal domain of TGN46 were bound at the cell surface and internalised. Antibody-TGN46 complexes were detected in punctate structures of 300-500 nm in diameter, similar to those found using antibodies to rat TGN38 (Reaves et al., 1993; Chapman and Munro, 1994). Between 40 and 50% of these structures co-localised with rhodamine-transferrin, identifying many of them as early endosomes. Some of the antibody-TGN46 complexes also co-localised with GalT, defining a route from the cell surface to the TGN most likely via endosomes. The second line of evidence came from experiments in which HeLa cells were treated with chloroquine, a drug that prevents recycling of rat TGN38 from endosomes to the TGN (Chapman and Munro, 1994; Reaves and Banting, 1994a). Human TGN46 accumulated in punctate structures, more than 50% of which co-localised with the transferrin receptor. In contrast to rat TGN38 (Chapman and Munro, 1994), a considerable portion of human TGN46 was found in the Golgi apparatus after the 3 hour incubation with chloroquine. Either the cycling time for human TGN46 is slower than for rat TGN38 or the block by chloroquine is not as complete and some of the recycling protein is reaching the TGN. Changes in incubation time or chloroquine concentration had little effect, so it will be important to carry out kinetic experiments on the recycling rate. The punctate structures containing antibody-TGN46 complexes but neither transferrin or the transferrin receptor have not been identified. They may represent a later endosomal compartment that must be traversed before the protein can reach the TGN. A more detailed morphological analysis using immunoelectron microscopy will be needed to resolve this point.

Structural analysis of the 14mer tandem repeats in the primate TGN38 homologues reveals a possible heptad-like motif, similar to those found in intermediate filaments and coiled-coil domains, which might facilitate antiparallel homodimers to assemble within this compartment. Although the incidence of proline in the repeat region is relatively high and may disrupt formation of putative α-helices, the striking arrangement of charged residues at adjacent vertices (based on a heptad repeat) cannot be coincidental and argues for a novel structural motif loosely based on those found in coiled-coil domains. Strong support for this is revealed by similarity of part of the 14mer repeat (KSGAEA) to a motif found in the C-terminal domain of mammalian neurofilament proteins (KSPAEA). Phosphorylation of the neurofilament motif on the serine residue is implicated in regulating the cytoplasmic assembly of these proteins (reviewed by Liem, 1993). A priority now is to mutate this motif in human TGN46 and study its effect on both structure and location.

The extracellular domain of all TGN38 homologues are notable for the high incidence of negatively charged amino acids, prolines, serines, threonines and other glycosylation sites. It is thus likely that the protein forms an extended, highly glycosylated structure in solution. Differences in primary structure can be rationalised by careful examination of common structural features within the tandem repeats of the three TGN38 homologues. The common features have also been found in highly glycosylated proteins such as the mucins

(reviewed by Devine and McKenzie, 1992; Burchell and Taylor-Papadimitriou, 1994). These proteins contain direct tandem repeats 16-23 amino acids in length (Devine and McKenzie, 1992; Burchell and Taylor-Papadimitriou, 1994). The TGN38 homologues have slightly shorter repeats of 8 or 14 residues. The mucin repeats of homologues from different species do not exhibit strict sequence or length homology (e.g. MUC1; reviewed by Devine and McKenzie, 1992), as now reported for the TGN38 homologues, and have a similar high preponderance of serine, threonine and proline residues.

The tandem repeats in the TGN38 homologues have a large number of serines and threonines, which are potential O-linked glycosylation sites. Additionally, there are five potential N-linked glycosylation sites just within the repeat domain. Rat TGN38, which has an apparent molecular mass of 85-95 kDa, is extensively glycosylated (Luzio et al., 1990; Jones et al., 1993). The human homologue has an apparent molecular mass of ~150 kDa on SDS-PAGE gels, suggesting extensive O- and N-glycosylation. Interestingly, the heavily glycosylated mucin type I integral membrane glycoprotein (MUC1) contains tyrosine-based motifs in its cytoplasmic domain that have been implicated in protein sorting in polarised cells (Pemberton et al., 1992). Again, this resembles the tyrosine-based signal found in the cytoplasmic domain of the TGN38 homologues. Taken together, these features suggest that these TGN38 homologues are distantly related to the mucin family of glycoproteins with a mucin-like lumenal domain.

How can any homology to mucin-like glycoproteins be reconciled with roles for TGN38 proteins in vesicular traffic from the TGN? Two possibilities exist for TGN38 function. The first possibility is that these proteins are ubiquitous receptors for ligands. As these TGN38 homologues cycle between different endocytic and exocytic organelles, ligand(s) may be bound and released at a later compartment along the route. In support of this, a tetrapeptide sequence (EAQT) within the TGN46 14mer repeat is also found within a repeating motif in the human P-selectin ligand (PSGL-1). The PSGL-1 ligand is a novel mucin-like type I transmembrane glycoprotein that contains 15 10mer direct tandem repeats (Sako et al., 1993).

Alternatively, and perhaps more likely, they may play a role in the packaging of cargo for export from the TGN as suggested by in vitro studies (Jones et al., 1993; Stanley and Howell, 1993). There is growing interest in the possibility that lectins play a role in sorting and concentrating proteins for export (Fiedler and Simons, 1995). Increasing protein concentration increases the likelihood of precipitation. TGN38 homologues could prevent this by acting as a type of molecular packing material. This property would result from the elongated, flexible, highly glycosylated nature of the protein. A calculated molecular length of 20-40 nm (S. Subbiah, personal communication) would permit anti-parallel homodimers to fill the intra-cisternal and intra-vesicular spaces by bridging them. Removal of TGN38 might even trigger the condensation process that leads to secretory granules. Further work will be needed to explore this or other functions of TGN38 proteins, but it is becoming increasingly clear that what started as a TGN marker is turning into an essential and novel component of the trafficking machinery.

We are extremely grateful to Alan Prescott (Dundee) for help with the confocal microscopy; and Joy Burchell and Joyce Taylor-Papadimitriou (ICRF), Kazuhisa Nakayama (Tsukuba, Japan) and Henrik Clausen (Copenhagen, Denmark) for communicating results prior to publication and discussions. S. Subbiah (Stanford), Tony Clarke (Bristol) and Paul Luzio (Cambridge) are thanked for discussions and insights into protein structure and sequence homologies. We thank Rose Watson and Felicia Hunte for technical help, ICRF Peptide Synthesis Unit and Dr Graham Bloomberg (BBSRC Molecular Recognition Centre, Bristol) for peptide syntheses. We also thank Sylvie Urbe, Francis Barr, Tom Misteli, Catherine Rabouille and the Banting and Warren labs for helpful advice and discussions. This work was supported by an MRC Senior Fellowship Career Development Award (S.P.), the MRC and CRC (M.G. and G.B.), and Imperial Cancer Research Fund (S.P., M.L.Y, C.O., T.N., M.F. and G.W.).

REFERENCES

Barlowe, C., Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama N., Rexach, M. F., Ravazzola, M., Amherdt, M. and Schekman, R. (1994). COPII - a membrane coat formed by sec proteins that drive vesicle budding from the endoplasmic-reticulum. *Cell* **77**, 895-907.

Bauerfeind, R. and Huttner, W. B. (1993). Biogenesis of constitutive secretory vesicles, secretory granules and synaptic vesicles. *Curr. Opin. Cell Biol.* **5**, 628-635.

Bos, K., Wraight, C. and Stanley, K. (1993). Tgn38 is maintained in the trans-golgi network by tyrosine-containing motif in the cytoplasmic domain. *EMBO J.* **12**, 2219-2228.

Brake, B., Braghetta, P., Banting, G., Luzio, J. P. and Stanley, K. K. (1990). A new recombinant DNA strategy for the molecular cloning of rare membrane proteins. *Biochem. J.* **267**, 631-637.

Burchell, J. and Taylor-Papadimitriou, J. (1993). Mucins. In *Guidebook to the Extracellular Matrix and Adhesion Proteins* (ed. T. Kreis and R. Vale), pp. 73-74. Oxford University Press.

Chapman, R. E. and Munro, S. (1994). Retrieval of TGN proteins from the cell surface requires endosomal acidification. *EMBO J.* **13**, 2305-2312.

Devine, P. L. and McKenzie, F. C. (1992). Mucins: structure, function, and associations with malignancy. *BioEssays* **14**, 619-625.

Fiedler, K. and Simons, K. (1995). The role of N-glycans in the secretory pathway. *Cell* **81**, 309-312.

Girotti, M., Jones, R., Emery, D. C., Chia, W. and Hall, L. (1992). Structure and expression of the rat epididymal secretory protein I gene : an androgen-regulated member of the lipocalin superfamily with a rare splice donor site. *Biochem. J.* **281**, 203-210.

Griffiths, G. and Simons, K. (1986). The trans-Golgi network: sorting at the exit site of the Golgi complex. *Science* **234**, 438-443.

Griffiths, G., Pepperkok, R., Krinjse-Locker, J. and Kreis, T. (1995). Immunocytochemical localization of β-COP to the ER-Golgi boundary and the TGN. *J. Cell Sci.* **108**, 2839-2856.

Harlow, E. and Lane, D. (1988). *Antibodies. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, New York.

Hopkins, C. R., Gibson, A., Shipman, M. and Miller, K. (1990). Movement of internalized ligand-receptor complexes along a continuous endosomal reticulum. *Nature* **346**, 335-339.

Humphrey, J. S., Peters, P. J., Yuan, L. C. and Bonifacino, J. S. (1993). A tyrosine-containing sequence within the cytoplasmic domain of TGN38 confers localization to the trans-Golgi network. *J. Cell Biol.* **120**, 1123-1135.

Jones, S. M., Crosby, J. R., Salamero, J. and Howell, K. E. (1993). A cytosolic complex of p62 and rab6 associates with TGN38/41 and is involved in budding of exocytic vesicles from the trans-Golgi network. *J. Cell Biol.* **122**, 775-788.

Kasai, K., Takahashi, S., Murakami, K. and Nakayama, K. (1995). Strain-specific presence of two TGN38 isoforms and absence of TGN41 in mouse. *J. Biol. Chem.* **270**, 14471-14476.

Kawano, J. I., Ide, S., Oinuma, T. and Suganuma, T. (1994). A protein-specific monoclonal antibody to rat liver beta-1-]4 galactosyltransferase and its application in immunohistochemistry. *J. Histochem. Cytochem.* **42**, 363-369.

Kornfeld, S. (1992). Structure and function of the mannose 6-phosphate insulin-like growth factor II receptors. *Annu. Rev. Biochem.* **61**, 307-330.

Letourneur, F., Gaynor, E. C., Hennecke, S., Demolliere, C., Duden, R., Emr, S. D., Reizman, H. and Cosson, P. (1994). Coatomer is essential for

retrieval of dilysine-tagged proteins to the endoplasmic reticulum. *Cell* **79**, 1199-1207.

Liem, R. K. H. (1993). Molecular biology of neuronal intermediate filaments. *Curr. Opin. Cell Biol.* **5**, 12-16.

Luzio, J. P., Brake, B., Banting, G., Howell, K. E., Braghetta, P. and Stanley, K. K. (1990). Identification, sequencing and expression of an integral mebrane protein of the trans-Golgi network (TGN38). *Biochem. J.* **270**, 97-102.

Narula, N. and Stow, J. L. (1995). Distinct coated vesicles labeled for p200 bud fromtrans-Golgi network membranes. *Proc. Nat. Acad. Sci. USA* **92**, 2874-2878.

Nilsson, T., Jackson, M. R. and Peterson, P. A. (1989). Retrieval of transmembrane proteins to the endoplasmic reticulum. *Cell* **58**, 707-718.

Nilsson, T., Pypaert, M., Hoe, M. H., Slusarewicz, P., Berger, E. G. and Warren, G. (1993). Overlapping distribution of two glycosyltransferases in the Golgi apparatus of HeLa cells. *J. Cell Biol.* **120**, 5-13.

Nilsson, T. and Warren, G. (1994). Retention and retrieval in the endoplasmic reticulum and Golgi apparatus. *Curr. Opin. Cell Biol.* **6**, 517-521.

Oprins, A., Duden, R., Kreis, T., Geuze, H. J. and Slot, J. W. (1993). β-COP localizes mainly to the cis-Golgi side in exocrine pancreas. *J. Cell Biol.* **121**, 49-59.

Pemberton, L., Taylor-Papadimitriou, J. and Gendler, S. J. (1992). The role of the cytoplasmic tail domain in actin interactions and in the polarized localization of cell-associated mucin, MUC1. *Biol. Cell* **3**, 266A.

Pepperkok, R., Scheel, J., Horstmann, H., Hauri, H. P., Griffiths, G., Kreis, T. E. (1993). b-cop is essential for biosynthetic membrane transport from the endoplasmic reticulum to the Golgi complex in vivo. *Cell* **74**, 71-82.

Perry, A. C. F., Jones, R., Niang, L. S. P., Jackson, R. M. and Hall, L. (1992). Genetic-evidence for an androgen-regulated epididymal secretory glutathione-peroxidase whose transcript does not contain a selenocysteine codon. *Biochem. J.* **285**, 863-870.

Perry, A. C. F., Hall, L., Bell, A. E. and Jones, R. (1994). Sequence analysis of a mammalian phospholipid-binding protein from testis and epididymis and its distribution between spermatozoa and extracellular secretions. *Biochem. J.* **301**, 235-242.

Ponnambalam, S., Rabouille, C., Luzio, J. P., Nilsson, T. and Warren. G. (1994). The TGN38 glycoprotein contains two non-overlapping signals that mediate localization to the trans-Golgi network. *J. Cell Biol.* **125**, 253-268.

Rabouille, C., Hui, N. H., Watson, R., Hunte, F., Berger, E., Warren, G. and Nilsson, T. (1995). Mapping the distribution of Golgi enzymes involved in the construction of complex oligosaccharides. *J. Cell Sci.* **108**, 1617-1627.

Reaves, B. and Banting, G. (1994a). Vacuolar ATPase inactivation blocks recycling to thetrans-golgi network from the plasma-membrane. *FEBS Lett.* 345, 61-66.

Reaves, B. and Banting, G. (1994b). Over-expression of TGN38/41 leads to mislocalisation of γ-adaptin. *FEBS Lett.* **351**, 448-456.

Reaves, B., Wilde, A. and Banting, G. (1992). Identification, molecular characterization and immunolocalization of an isoform of the trans-golgi network (TGN)-specific integral membrane-protein TGN38. *Biochem. J.* **283**, 313-316.

Reaves, B., Horn, M. and Banting, G. (1993). TGN38/41 recycles between the cell surface and the TGN: Brefeldin A affects its rate of return to the TGN. *Mol. Biol. Cell* **4**, 93-105.

Rothman, J. E. (1994). Mechanism of intracellular protein transport. *Nature* **372**, 55-63.

Sako, D., Chang, X. J., Barone, K. M., Vachino, G., White, H. M., Shaw, G., Veldman, G. M., Bean, K. M., Ahern, T. J., Furie, B., Cumming, D. A. and Larsen, G. R. (1993). Expression cloning of a functional glycoprotein ligand for P-selectin. *Cell* **75**, 1179-1186.

Salama, N. R. and Schekman, R. W. (1995). The role of coat proteins in the biosynthesis of secretory proteins. *Curr. Opin. Cell Biol.* **7**, 536-543.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York.

Stanley, K. K. and Howell, K. E. (1993). TGN38/41 - A molecule on the move? *Trends Cell Biol.* **3**, 252-255.

Wong, S. H. and Hong, W. (1993). The SXYQRL sequence in the cytoplasmic domain of TGN38 plays a major role in trans-Golgi network localization. *J. Biol. Chem.* **268**, 22853-22862.

Zehavifeferman, R., Burgess, J. W. and Stanley, K. K. (1995). Control of p62 binding to tgn38/41 by phosphorylation. *FEBS Lett.* **368**, 122-124.

Zehetner, G. and Lehrach, H. (1994). The reference library system - sharing biological material and experimental data. *Nature* **367**, 489-491.

(Received 22 September 1995 - Accepted 4 December 1995)