Glycoconjugate Journal 15, 671–681 (1998)

The rat α1,3-fucosyltransferase (rFucT-IV) gene encodes both long and short forms of the enzyme which share the same intracellular location

Jennifer M. Aucoin, Omanand Koul, Elizabeth M. Sajdel-Sulkowska, Thia Baboval and Frances I. Smith*

Biomedical Sciences Department, E.K. Shriver Center, Waltham, MA 02254, USA

Fucosyltransferase (FucT) activity has been detected on the surface of mouse germ cells and rat Sertoli cells, and has been postulated to play a role in cell-cell interactions. A recently cloned rat FucT (rFucT-IV) is expressed in the testes, and thus is a candidate for encoding the cell-surface FucT activity. This study maps the 5′-ends of several rFuc-T-IV mRNAs, and these results suggest that initiation of transcription may occur both upstream of the first ATG, as well as between the first two closely spaced, in-frame ATGs. Thus, in certain tissues, notably spleen, significant amounts of both a long and a short form of rFucT-IV would be predicted. This study also determines some basic properties of both the long and short forms of rFucT-IV, and investigates whether the use of alternative ATGs would allow FucT activity to be expressed both on the cell surface and in the Golgi. Plasmids that encode FLAG-epitope-labeled rFucT-IVs that initiate from either of the two ATGs were constructed, and rFucT-IV was expressed either *in vitro* using cell-free rabbit reticulocyte lysate, or after transfection in tissue culture. The results from these studies demonstrate that rFucT-IV is a glycosylated, transmembrane protein with a short cytoplasmic tail, and that either of the two ATGs in the 5′ region of the rFucT-IV gene are capable of acting as functional initiators of translation *in vitro*, to produce enzymatically active glycoproteins. However, no difference in the intracellular localization between the transferase containing a 48 amino acid or a 15 amino acid cytoplasmic tail was detected by immunocytochemistry, as both show the same pattern of Golgi-like staining in several different cell types, with no indication of surface expression. Thus, the additional amino-terminal 33 amino acids of the long form of rFucT-IV do not appear to influence its intracellular location in the cell types investigated.

***Keywords*: fucosyltransferase, recombinant enzyme, intracellular trafficking, transfected cells**

Introduction

Fucosyltransferases (FucTs) are among the group of glycosyltransferases that are postulated to play both a direct and an indirect role in cell-cell interactions. A well-studied role of glycosyltransferases is their involvement in the biosynthesis of cell-surface glycoproteins and glycolipids in the endoplasmic reticulum and Golgi [1] and FucTs are required for the synthesis of SSEA-1 glycoconjugates which have been shown to be involved in cell-cell interactions [2, 3]. In addition to their role in producing cell-surface glycoconjugates, FucTs themselves may also be expressed on the cell surface, and thus also have the opportunity to play a direct role in cell-cell interactions. The presence of FucT activity on the cell surface has been reported in mouse germ cells [4, 5] and rat Sertoli cells [6]. Other glycosyltransferases which have been found on the cell surface include β1,4-galactosyltransferase (GalTase; EC 2.4.1.38; [7]), *N*-acetylgalactosaminylphosphotransferase [8], and α2,6-sialyltransferase [9, 10]. The mechanism by which FucT activity may be localized both in the Golgi and at the cell surface has not yet been elucidated.

We have recently reported the cloning of a rat α1,3-fucosyltransferase gene (rFucT-IV) that is expressed in several tissues including testes [11]. The observation that rFucT-IV is expressed in testes suggests that it is a candidate for encoding the FucT activity found on the surface of Sertoli and germ cells [4–6]. Based on its sequence homology and substrate specificity, rFucT-IV has been classified as a member of the FucT-IV family. All Golgi-resident glycosyltransferases cloned to date are believed to be type-II membrane-bound proteins that contain a short amino-terminal cytoplasmic tail, a non-cleaved signal-anchor domain, a lumenal stem region, and a large carboxyl-terminal

* Corresponding author. Tel: (781) 642-0028; Fax: (781) 893-4018; E-mail: Fsmith@shriver.org

catalytic domain exposed to the Golgi lumen [1]. The predicted amino acid sequence of rFucT suggests that it is a type-II transmembrane protein of 433 amino acids, of which 48 amino acids are postulated to comprise a cytoplasmic tail. However, in addition to the first ATG, there is a second, closely-spaced, downstream, inframe ATG in the rFucT-IV gene. The placement of the two ATGs in the rFucT-IV gene and other members of the FucT-IV family is of special interest, because the α1,3-FucTs of other families so far cloned have an ATG in the position of the downstream ATG only. Moreover, another glycosyltransferase gene, GalTase, that also has two closely spaced, in-frame ATGs, uses two different transcription initiation sites to produce mRNAs that encode either a long or a short form of the protein, originating at either of the two different ATGs [12]. It has been reported that the long and short forms of the protein are differentially distributed within the cell, and that the long form which preferentially goes to the surface plays a direct role in cell-cell interactions [13, 14]. It is not known yet whether more than one isoform of rFucT-IV is expressed in different cell types. Northern blots of rat tissues reveal that multiple, differently sized, transcripts of the rFuc-IV gene are present in brain, spleen, and testes [11]. The basis for these differently sized rFuc-IV mRNAs is not known. However, the extensive use of alternative splicing of upstream exons in other closely related FucT genes [15], and the use of alternative promoters that result in a second transcription initiation site located between two inframe ATGs by other glycosyltransferases such as GalTase [12], must be considered as possibilities. Either one of these mechanisms could lead to the production of rFucT-IVs that differ in their amino-terminal regions, raising the possibility that members of the FucT-IV family could use different isoforms to express FucT activity on the cell surface and thus play a direct role in cell-cell interactions.

To address this possibility, we have mapped the 5′-end of rFucT-IV mRNAs in several tissues in the developing rat. These results show that transcripts encoding the long form of rFucT-IV are the predominant type in most tissues, but that significant amounts of transcripts encoding the short form of rFucT-IV are also present, most notably in spleen. To investigate the properties of these different forms of rFucT-IV, we have constructed several plasmids that express FLAG-epitope-labeled rFucT-IVs. These constructs are used to investigate the properties of the encoded proteins, after expression both *in vitro* using cell-free rabbit reticulocyte translation systems and after transfection of cells in tissue culture. The results from these studies suggest that both of the two ATGs in the 5′ region of the rFucT-IV gene are capable of acting as functional initiators of translation, but that both the long and the short forms of rFucT-IV show the same pattern of Golgi-like intracellular localization in several different cell types. Thus, the additional amino-terminal 33 amino acids of the long form of rFucT-IV do not appear to influence its intracellular location in these cell types, as detected by immunocytochemistry.

Materials and methods

Ribonuclease protection assay

Sprague-Dawley rats of defined age were sacrificed by CO_2 asphyxiation and tissues were removed and frozen at $-70\,^{\circ}C$ until use. The abbreviation "P" is used to denote postnatal, as in P4 for postnatal day 4. Rats were considered to be adults when older than two months. RNA was prepared by homogenizing each tissue sample in at least a $10\times$ volume of RNAzol B (Cinna/MRC, Cincinatti, Ohio) with a Polytron homogenizer, followed by chloroform extraction and RNA precipitation with isopropanol. After washing with 75% ethanol, the RNA pellet was solubilized in water and stored at $-70\,^{\circ}C$.

The ribonuclease assay was performed following the instructions of the Hyb-Speed™RPA kit from Ambion (Austin, Texas). Briefly, a ^{32}P- labeled RNA probe complementary to defined bases of the rFucT-IV gene (see Figure 1, where base $+1$ is the A of the first in-frame ATG) was prepared using the MAXIscript™ *in vitro* transcription kit (Ambion, Austin, Texas), and hybridized to RNA prepared as described above using the rapid 10 min hybridization protocol recommended by Ambion. After digestion by an RNaseA/T1 cocktail, samples were denatured and analysed by electrophoresis on denaturing 5–8% polyacrylamide gels (depending on the size of the protected fragments to be analysed), followed by autoradiography. A sample containing yeast RNA as a negative control was included in every assay. Size markers used were radioactively labeled single-stranded RNA Century™ markers (Ambion, Austin, Texas), or single-stranded DNA obtained from sequencing reactions on plasmids using Sequencase™ (USB, Cleveland, Ohio). When DNA markers were used, size estimates were reduced by 10% to correct for the faster migration of DNA.

Construction of prFucT vectors

A construct expressing the rFucT-IV protein tagged at its 3′-end with the octapeptide FLAG [16] was prepared. The 3′-end of the coding region of the rFucT-IV clone R-59 [11] was amplified with a 5′ primer that overlapped an NcoI site within R-59 (5′-AATTGCTTGAGCT-3′), and a 3′ primer that contained an insertion encoding the FLAG octapeptide immediately prior to the rat FucT-IV stop codon, as well as introducing a terminal EcoRI site (5′-GAATTCTCAC-TTGTCATCGT-CGTCCTTGTA-GTCTCGCTGG-AAC-CAGTCTG-CCAA-3′). The resulting PCR product was digested with NcoI and EcoRI, and ligated onto the 3′-end of the rFucT-IV insert obtained from a BamHI/NcoI digestion of R-59. This insert was then cloned into the expression vector pcDNA3 (In Vitrogen), and sequenced to verify that

no changes had been introduced upon amplification. This construct was named pcDNA3-FucTFLAG (prFucT).

A construct was then prepared which lacked the upstream ATG and thus could produce only the shorter mRNA transcript and encode the short form of rFucT-IV. The 5′ region of prFucT was amplified with a 5′ primer that created a 5′-terminal HindIII site and overlapped the downstream ATG and the 19 nucleotides which precede it (5′-TTAAGCTTGA-CCACTCGTGG-GCGCTGCATG-3′) and a 3′ primer that overlapped an NheI site within the gene (5′-TCTCTAGCTA-GCATCACCGCT-3′). The resulting PCR product was digested with HindIII and NheI, and ligated to the large fragment obtained from a HindIII and NheI digestion of pFucTFLAG. The resulting construct was labeled prFucT^{s}.

Because of possible interference from an upstream, out-of-frame ATG, a construct was also prepared which contained both in-frame ATGs but not the out-of-frame ATG. Two complimentary primers were designed (sense: 5′-AGCTTCCCTC-TATGGCCCC-3′, and antisense: 5′-GGCCGGGCCA-TAGAGGGA-3′) which contained the upstream ATG, the six nucleotides which preceded it and part of an internal EagI site, and when annealed left a 5′ HindIII overhang and a 3′ EagI overhang. The prFucT plasmid was cut with EagI and EcoRI and the small fragment resulting from this digestion was ligated to the annealed primers and to pcDNA3 cut with EcoRI and HindIII. The resulting construct was labeled prFucT^{L}.

Several regions of the rFucT-IV gene also were subcloned in the vector pGEM3zf (Promega, Madison, WI) to enable the production of RPA probes of varying size and composition, as indicated in Figure ID.

Cell culture

The African green monkey kidney cell line COS-1, human epitheloid carcinoma cell line HeLa, human embryonic kidney cell line 293, human glioblastoma U87-MG, and mouse neuroblastoma cell line Neuro-2a were all obtained from the American Type Culture Collection (ATCC). COS-1 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM; Mediatech, Pittsburg, Pennsylvania) supplemented with 10% fetal bovine serum. HeLa and 293 cells were grown in Minimum Essential Medium (MEM; Mediatech, Pittsburg, Pennsylvania) supplemented with 10% fetal bovine serum. U87-MG and Neuro-2a cells were grown in MEM supplemented with 10% fetal bovine serum, sodium pyruvate at 1 mM, and non-essential amino acids at 100 μM.

Cell transfection

Cells were grown to about 70% confluence in 60 mm tissue culture dishes containing poly-D-lysine (Sigma, St. Louis, Missouri) coated coverslips. Cells were transfected with 4 μg plasmid using 12 μl Lipofectamine™ (Gibco-BRL, Grand Island, New York) according to the manufacturer's recommendations. After 6 h, the medium was removed and replaced with fresh medium. About 40 h later, coverslips were removed for immunocytochemistry and the remaining cells were scraped into the medium, and spun down at 4 °C for 5 min at 1500 rpm. The cell pellet was resuspended in 1 ml phosphate-buffered saline (PBS) and spun at 14 000 rpm for 1 min. The supernatant was removed and cell pellet was frozen at − 70 °C.

Immunocytochemistry

Coverslips were rinsed with (PBS) and fixed for 5 min in 4% paraformaldehyde. Cells to be permeabilized were then treated with 0.05% saponin for 10 min, while cells that were to be surface-stained were incubated in 10% normal goat serum in PBS. Cells were incubated with either anti-FLAG (Kodak-IBI, Rochester, New York) or anti-SSEA-1 antibodies (3B11; gift of Dr U. Dräger, [17]), followed by incubation with either rhodamine- or fluoroscein-conjugated anti-mouse IgG or IgM antibodies, respectively. Coverslips were mounted using Gelmount (Biomeda, Foster City, California) containing the anti-fadent para-phenylendiamine and examined using fluorescent and light microscopy.

Western blot analysis

The frozen cell pellet was resuspended in homogenization buffer (50 mM Tris pH 7.4, 2 mM EDTA, 100 μg ml^{-1} phenyl-methylsulfonyl fluoride, 1% aprotinin, and 25 μg ml^{-1} pepstatin), and subsequently boiled for 5 min in loading buffer. Samples were run on either 10% or 12% polyacrylamide gels and electrophoretically transferred to Zeta-probe membrane (BioRad, Hercules, California). The membrane was incubated in 5% milk (Carnation) in Tris-buffered saline (TBS; 50 mM Tris pH 7.4, 150 mM NaCl) for 1 h at room temperature, followed by incubation in anti-FLAG antibody (Kodak-IBI, Rochester, New York 10 μg ml^{-1}) in 5% milk in TBS overnight at 4 °C. The membrane was washed and incubated in peroxidase-conjugated goat anti-mouse IgG antibody in TBS for 30 min at room temperature, then washed three more times. Signal was detected after incubation in Super-Signal™ Solution (Pierce, Rockford, Illinois) for 2–5 min and exposure to HyperfilmECL (Amersham, Arlington Heights, Illinois) for variable lengths of time.

Fucosyltransferase enzyme assay

Frozen cell pellets of transfected cells were homogenized in 20 mM MOPS/NaOH buffer (pH7.4, containing 1% Triton X-100). The homogenate was centrifuged (about 12 000 g) and the supernatant was used as the enzyme source. Protein was measured in ELISA plates by the BCA microassay (Pierce, Rockford, Illinois). Fucosyltransferase activity was assayed essentially as described previously [11, 18]. The

incubation medium contained 0.1% CHAPS, 2.5 μmol of MOPS pH7.4, 0.25 μmol of $MnCl_2$, 5 μmol of NaCl, 0.5 μmol of ATP, 104 nmol of *N*-acetyllactosamine (LacNAc), 0.5 nmol of GDP-L- [^{3}H] fucose (American Radiolabeled Products, St. Louis) and between 20 and 50 μg of protein of the enzyme source in a total assay volume of 50 μl. The tubes were incubated for 3 h at 37 °C and the reaction terminated by addition of 0.5 ml of cold water. The mixture was applied to a 1 ml column of AG 1-X8 anion exchange resin (BioRad, Hercules, California) and washed with 1 ml of water. The aqueous effluents containing the fucosylated products were pooled in scintillation vials and mixed with 4.5 ml of ReadySafe scintillation cocktail (Beckman, Palo Alto, California). Blank assays were run in the absence of LacNAc, and the net enzyme activity was calculated after subtraction of counts found in the blank. Enzyme activity is expressed as pmoles of fucose transferred to LacNAc per mg protein.

In vitro coupled transcription/translation reactions

The T7 polymerase-rabbit reticulocyte lysate-based coupled transcription/translation system (TNT™; Promega, Madison, Wisconsin) was used according to manufacturer's instructions in the presence of ^{35}S-labeled methionine (Amersham, Arlington Heights, Illinois). In some reactions, canine pancreatic microsomes (Promega, Madison, Wisconsin) also were added. After incubation at 30 °C for 90 min, reactions were diluted in sample buffer containing SDS and 2-mercaptoethanol, boiled for 5 min, and electrophoresed on 10% or 12% polyacrylamide gels. In some experiments, after the 90 min incubations reactions were treated with proteinase K (20 μg ml^{-1} final concentration) on ice for 30 min before boiling and gel analysis. After electrophoresis, gels were fixed (10% acetic acid, 25% isopropanol) for 20 min, incubated in Amplify (Amersham, Arlington Heights, Illinois) for 20 min, dried, and exposed to film.

Results

Mappings 5′ends of rFucT-IV RNA transcripts

A radiolabeled probe spanning approximately 400 bases of the 5′ region of the rFucT-IV gene (from position − 198 to + 200; RPA probe 397, Figure 1D) initially was used in ribonuclease protection assays to investigate the diversity of the 5′ ends of rFucT-IV RNA transcripts. RNA was prepared from tissues that had previously been shown to express the rFucT-IV gene by Northern blot analysis [11]. Early postnatal tissues were used as these tissues were found to have the highest levels of transcripts for most tissues (data not shown). Results show that all tissues investigated contained a major transcript that protected all of the rFucT sequences in the probe (labeled FL to designate full-length protected, Figure 1A). A second minor species that protected approximately 250 bases of the probe (labeled FT2, Figure 1A) was detected in all tissues except spleen. Additionally, a cluster of shorter transcripts ranging in length from about 150 to 200 bases in length were detected (labelled FT3, Figures 1A and 1B) that were present in highest amount in spleen and in lower amounts in kidney. Although the FT3 bands cannot be seen in lanes 2–5 in Figure 1A, longer exposures and data from other experiments suggest that they are present in trace amounts in these tissues. These shorter FT3 transcripts would not contain the first in-frame ATG of the rFucT-IV gene. Thus, most tissues would be predicted to contain predominantly the long form of rFucT-IV, and spleen would be predicted to contain significant amounts of both long and short forms of rFucT-IV (Figures 1A and 1B).

Additional radiolabeled probes spanning approximately 1800 bases upstream of the first in-frame ATG of the rFucT-IV gene were used in ribonuclease protection assays to map the 5′end(s) of the longest protected fragment (designated FL above). Initial experiments using RPA probe 1264 suggested that this fragment was derived from a major start site(s) approximately 400 bases upstream of the first ATG, and experiments using the smaller RPA probe 170 mapped these sites of 46 and 44 bases upstream of a PstI site (labeled FT1, Figures 1C, D).

In vitro translation studies

In order of facilitate detection of the rFucT-IV protein, the FLAG epitope was added to the 3′ end of the coding region of the rFucT-IV gene to form prFucT (Figure 2). Initial experiments showed that although this construct produced the FLAG-tagged rFucT-IV protein upon transfection into COS-1 cells, it did not produce protein in the *in vitro* system (data not shown), presumably due to the interference of an upstream, out-of-frame, ATG. Therefore, a construct was created which deleted the upstream region containing the out-of-frame ATG. This construct was designated prFucT^L and contains the entire coding region of the rFucT-IV gene. A second construct, prFucT^s, was created which deleted everything upstream within 20 nucleotide base pairs of the downstream ATG, and thus encodes only a truncated rFucT-IV protein.

To determine whether both in-frame ATGs can be recognized *in vitro*, the rFucT-IV constructs were translated in the presence of rabbit reticuloctye lysate. As shown in Figure 3 (lane 2), prFucT^L, which contains both in-frame ATGs, produced predominantly the longer form of the protein (predicted MW 48.8kd), indicating that the first ATG is strongly favored in this system. However, significant amounts of the lower MW form of the protein (predicted MW 45.1kd) are also produced, indicating that the second ATG may also act as a functional initiation codon. Use of the prFucT^s construct produces the short form of the protein, confirming the potential utility of the second ATG (Figure 3, lane 3). However, results obtained *in vitro* do not

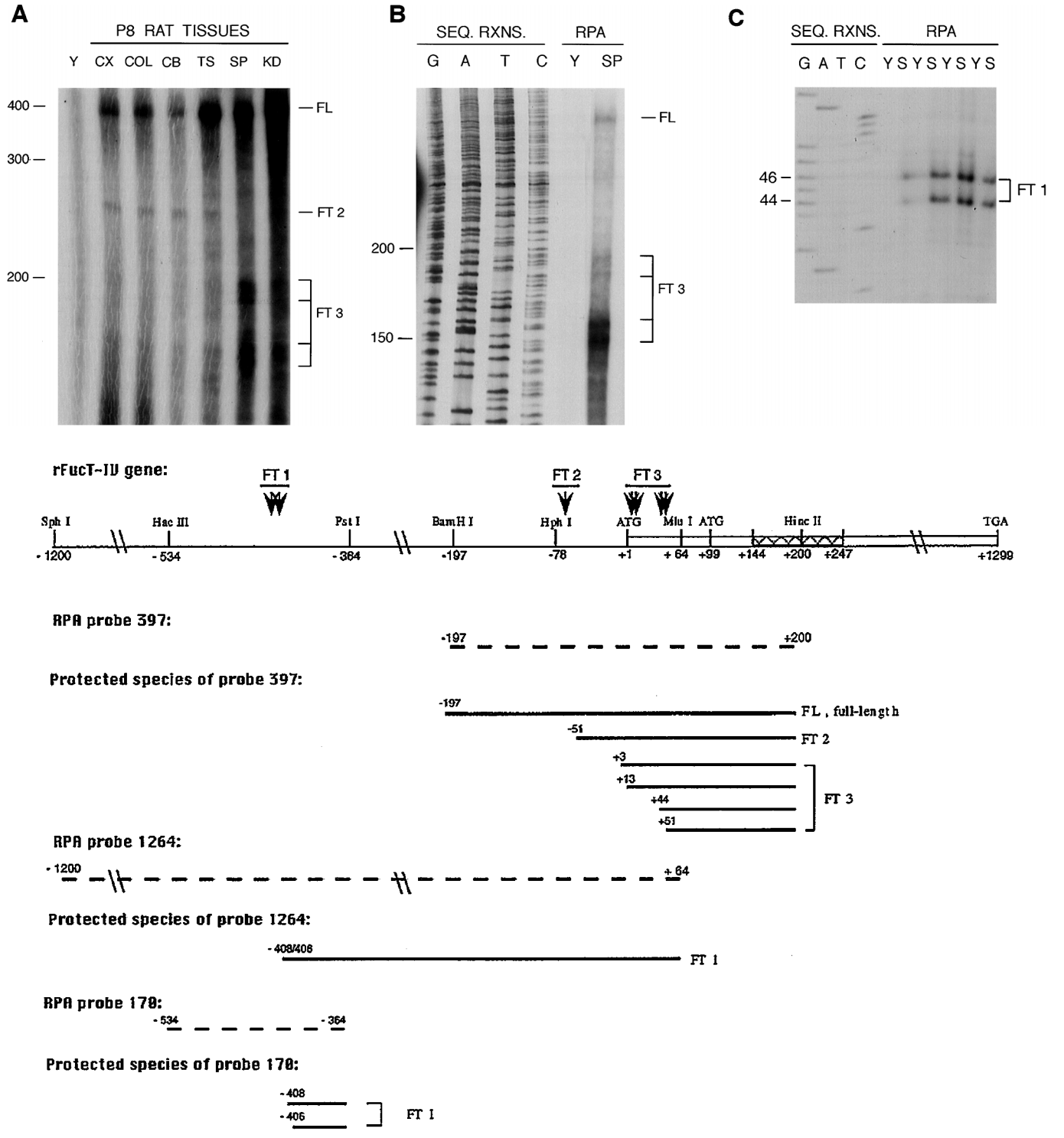

Figure 1. Ribonuclease protection assays detect heterogeneity in the 5′ends of rFuct-IV mRNA transcripts. Radiolabeled RNA probes complementary to indicated regions of the rFucT-IV gene were used to protect homologous sequences in RNA prepared from either yeast (Y), or rat forebrain (CX), colliculus (COL), cerebellum (CB), testes (TS), spleen (SP, or S), or kidney (KD). Panel A: P8 rat tissues were compared using RPA probe 397. RNA size markers were used to estimates sizes. Panel B: RNA was prepared from P11 rat spleen and analyzed using RPA probe 397. DNA sequencing reactions were used as size markers, and the sizes indicated have already been adjusted by 10% to correct for the difference in migration between singled stranded RNA and DNAs. Panel C.RNA was prepared from P30 rat spleen and analyzed in quadruplicate using RPA probe 170. DNA sequencing reactions were used as described in panel B. Panel D. Diagrammatic representation of the rFucT-IV gene, the RPA probes used, the protected fragments detected, and the deduced positions of the 5′ends of the rFucT-IV mRNAs detected in these tissues. The cross-hatched region in the rFuct-IV gene represents the transmembrane region of the encoded protein.

always mirror what is obtained *in vivo*, and so although these results show that both ATGs can function equally well *in vitro*, there may well be other factors that influence their translation initiation efficiency *in vivo*.

In vitro processing studies

Based on the predicted amino acid sequence, rFucT-IV is a type II transmembrane protein with an amino-terminal cytoplasmic tail of 48 or 15 amino acids for the long or short form of the proteins, respectively. Both proteins also contain two potential glycosylation sites in their putative lumenal regions, but it is not known if they are used. Results obtained using *in vitro* translation in the presence of canine microsomes show that both rFucT^{L} and rFucT^{S} are translocated across the membrane and are glycosylated *in vitro*, as evidenced by the appearance of a more slowly migrating band in each lane on polyacrylamide gel analysis (Figure 3, lanes 1 and 4). Limited proteinase K digestion of these samples after incubation with microsomes causes diminution of both the upper and lower bands, and the appearance of a new band of intermediate mobility of the approximate MW as expected for the loss of the cytoplasmic tail, as shown in Figure 4 (lanes 3 and 4, arrowed). This result can be explained by the susceptibility of proteins outside of the microsomes to digestion. Thus, the lower band increases due to digestion of the unprotected, untranslocated protein. The lumenal portion of the translocated, glycosylated, transmembrane protein is protected, but the cytoplasmic tail is cleaved off, giving rise to a shorter protein. Thus, the upper band decreases and gives rise to the intermediate band. These observations support the prediction that rFucT-IV is a type II transmembrane protein.

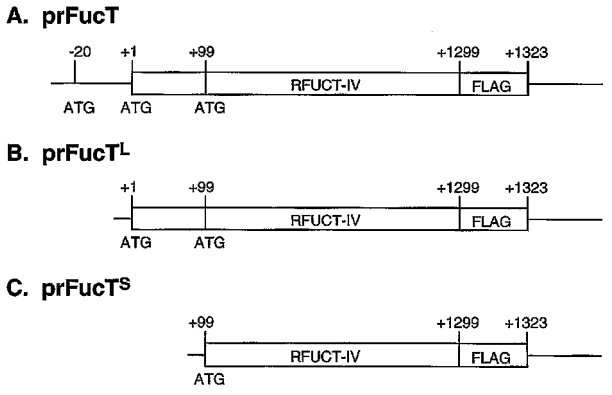

Figure 2. Diagrammatic representation of the coding regions of the three constructs used in this study. All encode the FLAG epitope immediately prior to their termination codon, but differ in the number of upstream ATGs.

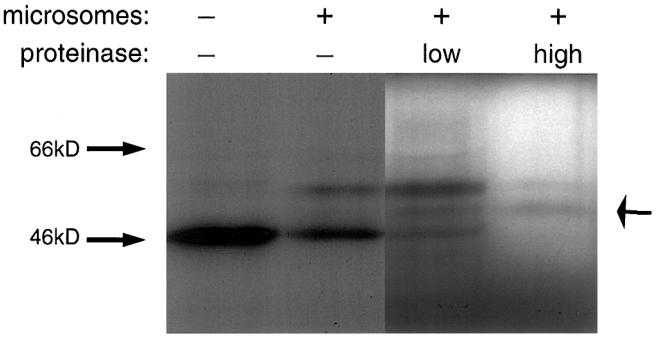

Figure 4. rFuct-IV is a transmembrane glycoprotein. prFuc-T^{L} was transcribed and translated *in vitro* in the presence or absence of microsomes, followed by a subsequent incubation in either no, low or high concentrations of proteinase K. The resulting reaction mixes were separated by polyacrylamide gel electrophoresis, and proteins detected by autoradiography. Due to the partial leakiness of the microsomes, the overall signal drops on digestion with proteinase K, and so a longer exposure was necessary for the detection of the proteins in the digested samples.

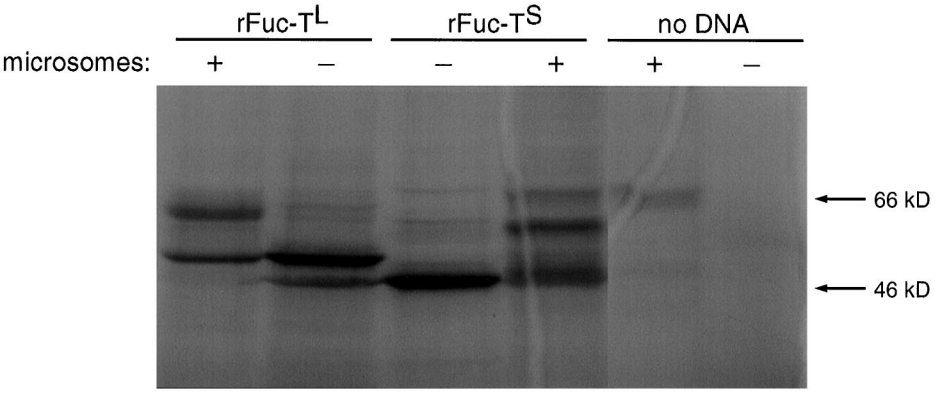

Figure 3. Both ATGs can fuction as initiators of translation *in vitro* to produce glycosylated proteins. prFuc-T^{L} and prFuc-T^{S} were transcribed and translated *in vitro* in the presence or absence of microsomes, and the resulting reaction mixes were separated by polyacrylamide gel electrophoresis

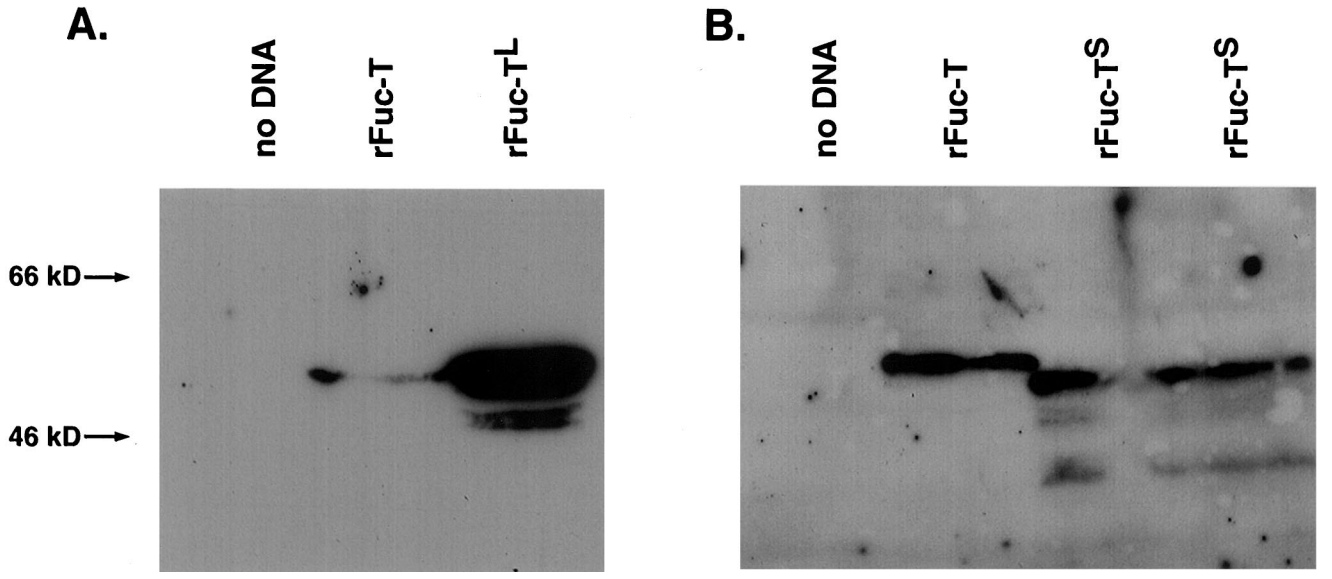

Figure 5. Both ATGs can initiate translation in transfected cells. COS-1 cells were transfected with rFuc-T plasmids and the cell lysates were analyzed by Western blot using anti-FLAG antibody. Panels A and B show two separate experiments.

Table 1. Fucosyltransferase activity is expressed by all three plasmids on transfection into COS-1 cells

Plasmid	*Enzyme activity (pmoles mg^{-1} protein)*[a]
prFuc-T	288 ± 38
prFuc-T^{L}	100 ± 8
prFuc-T^{S}	81 ± 6
None	10 ± 8

[a] pmoles of fucose transferred to LacNAc per mg protein. Results are presented as the mean ± standard deviation for duplicate determinations.

In vivo transfection studies

The above results indicated that both in-frame ATGs can function as initiators of protein translation in a rabbit reticulocyte *in vitro* system. To determine whether both in-frame ATGs can be used in whole cells, COS-1 cells were transfected with the three constructs and cell lysates were collected 48 h later and analyzed for protein expression by Western blot using anti-FLAG antibody. Results in Figure 5A show that both the prFucT and prFucT^{L} constructs produce predominantly the long form of the protein, and results in Figure 5B show that prFucT^{S} construct produces the short form of the protein. No band that aligned with the rFucT^{S} form of the protein could be detected when the longer constructs were used, indicating that considerably less of the shorter form is present in these cells than was observed in the *in vitro* reticulocyte lysate system. The minor lower bands seen in this figure are believed to be due to proteolysis.

To determine whether the FLAG-tagged forms of rFucT-IV still retained enzyme activity, fucosyltransferase activity was measured by the ability of transfected cell lysates to transfer radiolabeled fucose to the substrate LacNAc. Enzyme activity was present in COS-1 cells transfected by any of the three prFucT plasmids (Table 1).

Immunocytochemistry results

COS-1 cells transfected with any of the three plasmids containing the FLAG-tagged rFucT-IV (prFucT, prFucT^{L}, or prFucT^{S}) express SSEA-1 on the cell surface. These results confirm that addition of the FLAG epitope onto the carboxy-terminus, or shortening of the amino-terminus, does not abrogate α1,3-fucosyltransferase activity. Results obtained for cells transfected with prFucT^{L} are shown in Figure 6A, and similar results were obtained with the other two plasmids (data not shown). Staining of permeabilized cells with anti-FLAG antibody after transfection with any of these three plasmids shows a perinuclear staining, with a Golgi-like pattern of intracellular localization, as may be expected for a glycosyltransferase, with no obvious surface staining above background levels. Again, all three plasmids, produced indistinguishable results, and so only results for prFucT^{L} are shown in Figure 6B. Furthermore, no anti-FLAG antibody reaction above background levels with non-permeabilized transfected cells was detected (data not

shown), implying very low or lack of surface expression of recombinant rFucT-IV in COS-1 cells. As it has been reported that intracellular localization may differ for the same protein depending on cell type, several different cell lines were transfected and either stained for surface expression of SSEA-1 antigen or FLAG-epitope, or stained for intracellular FLAG-epitope. Cell lines used were human glioblastoma (U87-MG), human epitheloid (HeLa), human embryonic kidney (293), African green monkey kidney (COS-1), and mouse neuroblastoma (Neuro2a). All cell lines transfected with any of the three plasmids showed the same result: surface SSEA-1 positive, with a Golgi-like intracellular localization of FLAG-epitope, with no surface expression of FLAG-epitope detected. Some examples of these results for cells transfected with prFucT^{L} are shown in Figure 6. Thus, changing the cell type or the length of the amino-terminus of the rFucT-IV protein did not appear to alter its intracellular targeting, as detected by the methods employed in our studies.

Discussion

Results presented in this paper reveal that there is considerable variation in the 5′ends of transcripts of the rFucT-IV gene. In most tissues, the dominant transcripts begin at approximately 406–408 bases before the first in-frame ATG, and a minor transcript begins around 50 bases before the first ATG. Both these transcripts encode the long form of rFucT-IV. In spleen, an additional major cluster of transcripts whose 5′ends is mapped between the first two in-frame ATGs is also detected, and these transcripts encode the short form of rFucT-IV. This cluster of transcripts also is detected as a minor component in kidney, and only in trace amounts in the other tissues examined. Thus, if one assumes equal efficiency of translation of these transcripts, the postnatal tissues examined in this study may be expected to express predominantly the long form of rFucT-IV, and only spleen would be expected to have comparable amounts of long and short forms of rFucT-IV at the developmental

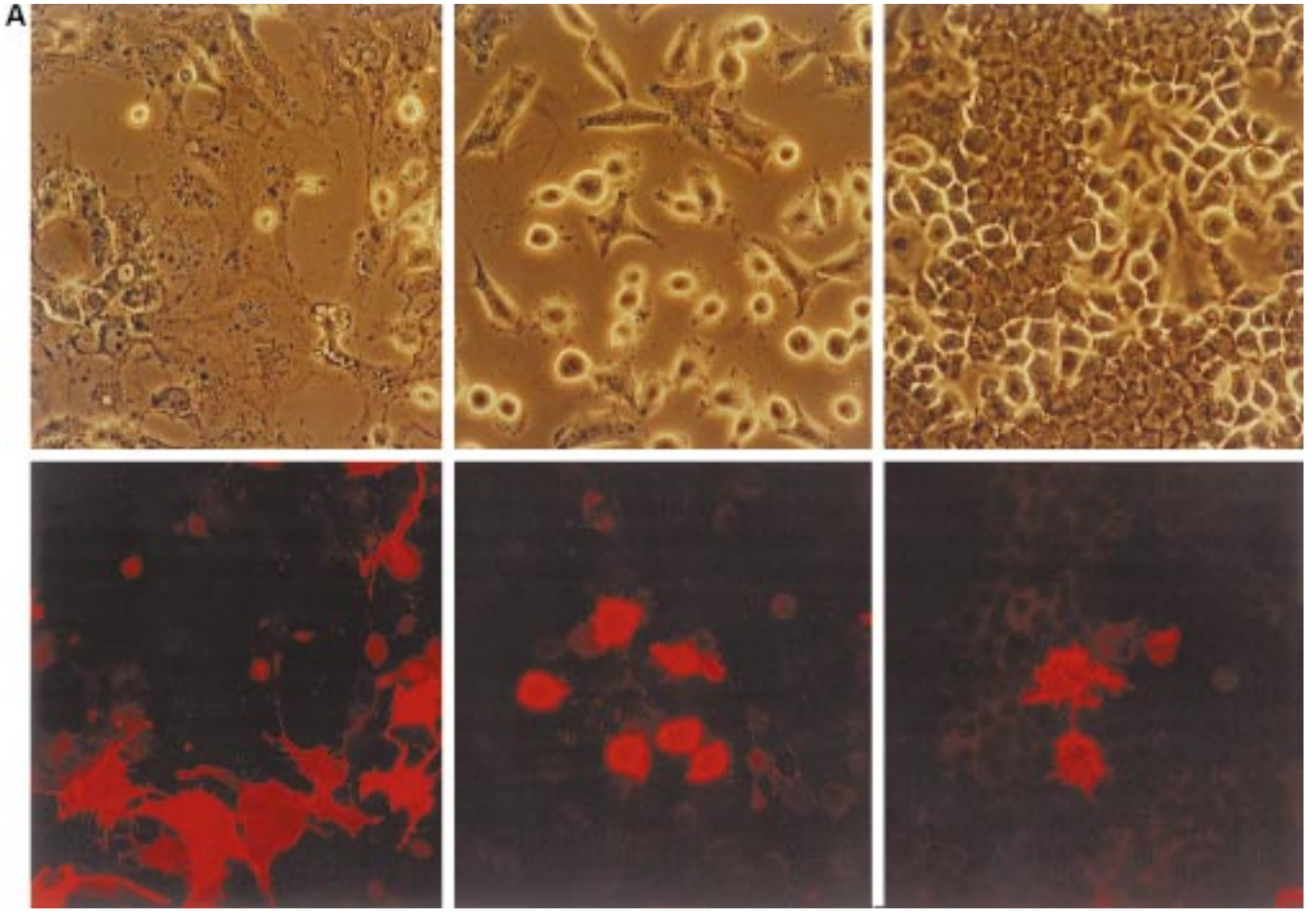

Figure 6. Immunocytochemistry of cells transfected with prFuc-T^{L}. Panel A. Surface staining of cells using anti-SSEA-1 antibody. × 400 magnification. Panel B. Staining of permeabilized cells using anti-FLAG antibody. × 1000 magnification. From left to right in both panels: results for COS-1, HeLa, and Neuro2a cell lines are shown.

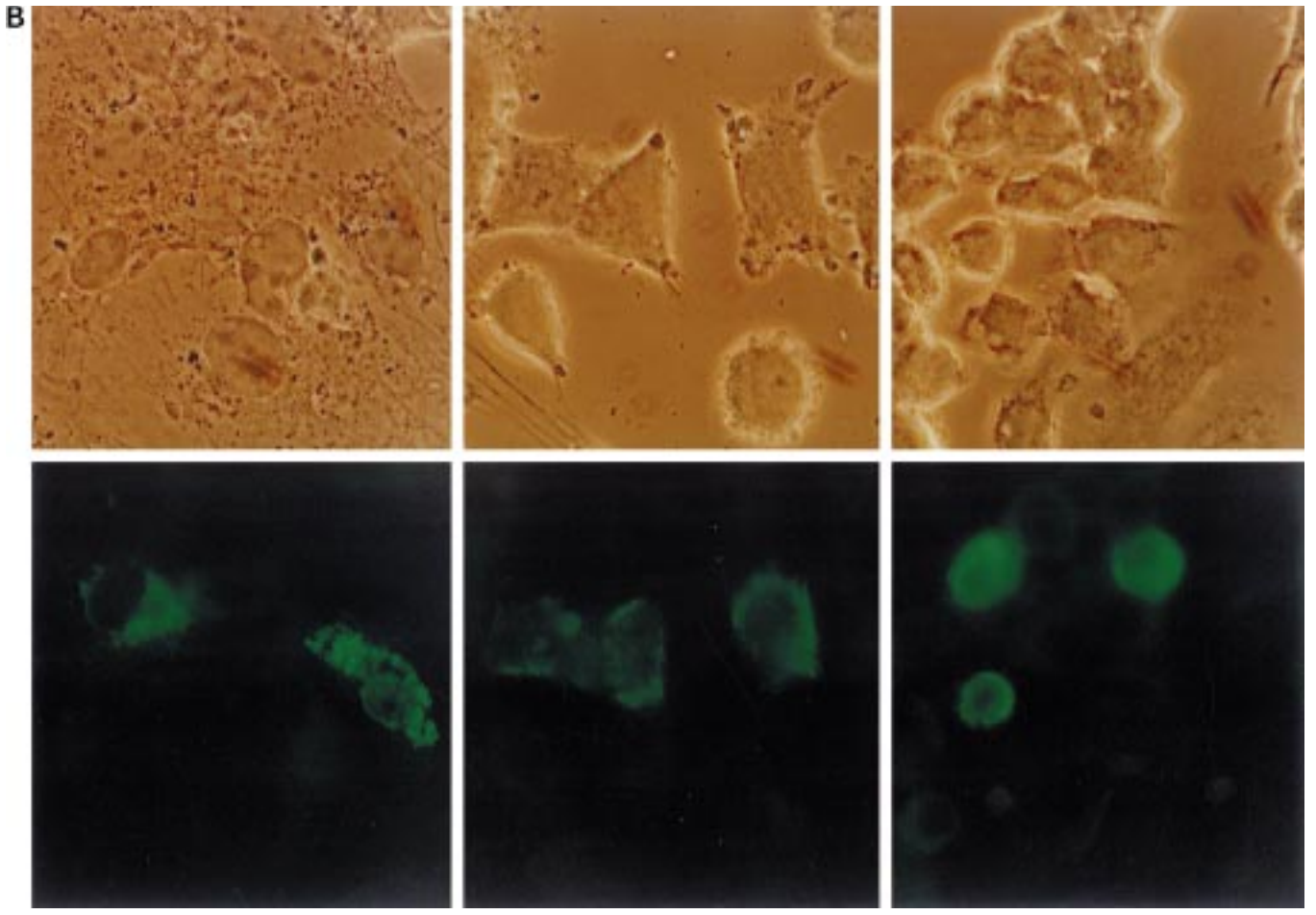

Fig. 6. Continued.

times examined. The mechanisms responsible for the generation of these different transcripts is not yet clear. The 5′ ends of the longest FT1 transcripts are mapped downstream of a very purine-rich region, with no obvious promoter elements. The second minor transcript found in most tissues also is not near any obvious promoter elements, but is located in the vicinity of two possible alternative splice acceptor sites (GAG at positions − 151 and − 144). Since it has been noted that the 5′ intron of genes is removed more slowly than other introns within a gene [19], if band FT2 does indeed arise from splicing of a longer precursor, the relative intensity of the band at the FT2 position must be interpreted with caution. The cluster of transcripts between the two ATGs (FT3 bands) is located in a GC-rich region that contains multiple motifs (consensus sequence GGGCGG, in either orientation) that are characteristic of SP1 transcription factor binding sites [20]. Housekeeping gene promoters frequently contain clusters of SP1 binding sites, and frequently have multiple initiation sites within these regions, making it likely that these rFucT-IV transcripts are due to multiple initiations of transcription. Thus, the short form of rFucT-IV most likely arises from the use of an alternative promoter between the two ATGs. This arrangement could allow differential regulation of the long and short forms of rFucT-IV. Such a situation has been examined in detail in the case of the GalTase gene, in which the long form of GalTase is driven by a housekeeping promoter in most tissues, and a second highly regulated promoter is used in mammary tissue to produce large amounts of the short form of GalTase during lactation [21].

Results obtained in this study indicate that both long and short forms of rFucT-IV are active enzymatically. A comparison of the amino acid sequences of FucT family members indicates that all α1,3-FucTs, except for the FucT-IV family, have an ATG in the position of the downstream ATG only, predicting short cytoplasmic tails of 14–22 amino acids. (The FucT-IV family have an additional upstream ATG encoding an additional 33 amino acids for rFucT-IV, an additional 34 amino acids for murine FucT-IV, and an additional 125 amino acids for human FucT-IV). Thus, in analogy with the other FucT families, the use of the downstream ATG in rFucT would be predicted to produce

a functional enzyme, and reults presented here using a truncated form of the rFucT-IV gene confirm this prediction. The preservation of this downstream ATG in all FucT-IV family members in human, mouse and rat [11, 22–25], may suggest that the second ATG is the one most likely to be used. However, on transcription into mRNA, both ATGs in the 5′ region of the rFucT gene give reasonably good matches to the Kozak consensus sequence for protein initiation CC(A/G) CCAUGG [26]: CCTCTAUGG and GCTGCAUGG for the upstream and downstream AUGs, respectively. Based on the 'scanning model' [26], protein synthesis is expected to initiate at the first appropriate AUG encountered. Western blot analysis after transfection of COS-1 cells with constructs bearing either both ATGs or only the downstream ATG confirms this prediction. Thus, when both AUGs are present, the most 5′-AUG is used. When only the downstream AUG is present, then it is used with apparently equal efficiency.

In vitro studies shown here inidate that rFucT-IV is a glycosylated, transmembrane protein, with a short cytoplasmic tail, as predicted from hydrophobicity plots of rFucT-IV. Transfection of several different cell lines with the rFucT-IV constructs reveals that both the long and short forms of the protein have enzymatic activity. As would be expected for transmembrane glycosyltransferases, immunocytochemistry detected the presence of the recombinant rFucT-IVs in a perinuclear, Golgi-like pattern within the cell. No surface expression was detected in any cell line tested, for any construct. Thus, neigher the long nor the short forms of rFucT-IV traffic to the surface in sufficient quantity to be detected in our assays, suggesting that the terminal amino acids of the long form of the rFucT-IV do not influence the localization of this protein. The signal for retention in the Golgi most likely resides in rFucT-IV′s membrane spanning domain, as has been shown for the glycosyltransferases GalTase [27–31], α2,3-sialyltransferase [32–34], and *N*-acetylglucosaminyltransferase I [35, 36]. Although it is known that the transmembrane domain is responsible for Golgi retention of these glycosyltransferases, the mechanism of retention is not yet known and they share little if any sequence similarity in the transmembrane region. Proposed mechanisms of retention include association with tubulins [37], the oligomerization or interaction of proteins with each other in the Golgi membrane [28, 38–40], and the correlation of shorter transmembrane length with Golgi retention [31, 32, 41–43]. Little if any sequence similarity is found between the cytoplasmic regions of the different glycosyltransferases, including rFucT-IV and GalTase.

Although the studies presented herein did not find any evidence for surface expression of rFucT-IV, one cannot rule out the possibility that rFuc-TIV is present on the cell surface either in very low amounts, or only in certain specialized cell types. Even in the GalTase system, there is controversy over how much and which form of GalTase is expressed on the cell surface. Some investigators found evidence of one or both forms of GalTase on the cell surface [14, 30, 44, 45] while others reported that it was not there [27, 29, 31]. Several explanations for the failure of certain groups to localize GalTase to the cell surface have been proposed [46]. One such explanation is that only a relatively small portion of the long form of GalTase goes to the surface, and that unless one is actively searching for it, it may go unrecognized. An indirect way that the long form of GalTase was implicated as the cell surface form was to prepare constructs of the short form (SGT) and the long form (LGT) both of which lacked the catalytic portion of GalTase [47]. These constructs were transfected into cells which normally use GalTase on the surface to mediate cell-adhesion events. Cells transfected with the inactivated LGT were found to lose cell-adhesion, while transfection with inactivated SGT had no effect. Thus, the inactivated LGT was able to block cell adhesion by competing with the normal LGT for transport to the cell surface. This assay allowed the functional detection of GalTase on the cell surface in a situation in which very small amounts were present. No function has yet been demonstrated for surface FucT, and so such an approach cannot be used to demonstrate the presence of small amounts of rFucT-IV on the cell surface.

Another explanation that has been suggested for the lack of detection of glycosyltransferases on the surface of certain cells is that cell surface localization is dependent on interaction with certain proteins that are present only in certain specialized cell types. Since both the long and short forms of GalTase possess the same membrane-spanning region that has been shown to be sufficient for Golgi localization [26–30], the movement of glycosyltransferases from the Golgi to the membrane may require interaction with receptors which recognize the additional amino-terminal residues of the long form. Cells which do not normally produce surface glycosyltransferases may not have such receptors available, and thus transfection of these cells with constructs encoding the surface form of the protein would not result in the transport of this protein to the surface. Since FucT activity has been detected on the surface of mouse germ cells and rat Sertoli cells, it is possible that it requires interaction with receptors that are present in these specialized cells, but lacking in the cell types that were used in our experiments. It is also possible that surface expression of FucT activity on rat Sertoli cells in due to the expression of an as yet unknown FucT gene.

Acknowledgments

We thank Dr. R.H. McCluer for helpful advice throughout this project. This work was supported by the National Institutes of Health grant HD05515.

References

1 Paulson JC, Colley KJ (1989) *J Biol Chem* **264**: 17615–18.
2 Hakomori S (1983) In *Handbook of Lipid Research* (Hanahan DH, ed) vol. 3, pp 327–79. New York: Plenum Press.
3 Muramatsu T (1992) In *Cell Surface Carbohydrates and Cell Development* (Fukada M, ed) pp 239–56. Boca Raton, FL: CRC Press.
4 Ram PA, Cardullo RA, Millette CF (1989) *Gamete Res* **22**: 321–2.
5 Cardullo RA, Armant DR, Millette CF (1989) *Biochemistry* **28**: 1611–17.
6 Raychoudhury SS, Millette CF (1993) *Mol Repr Dev* **36**: 195–202.
7 Roth S, McGuire EJ, Roseman S (1971) *J Cell Biol* **91**: 536–47.
8 Balsamo J, Pratt RS, Emmerling MR, Grunwald GB, Lilien J (1986) *J Cell Biochem* **32**: 235–41.
9 Paietta E, Hubbard AL, Wiernik PH, Diehl V, Stockert R (1987) *Cancer Res* **47**: 2461–7.
10 Taatjes DJ, Roth J, Weinstein J, Paulson J (1988) J Biol **267**: 4084–96.
11 Sajdel-Sulkowska EM, Smith FI, Wiederschain G, McCluer RH (1997) *Glycoconjugate J* **14**: 249–58.
12 Russo RN, Shaper NL, Shaper JH (1990) *J Biol Chem* **265**: 3324–31.
13 Youakim A, Hathaway HJ, Miller DJ, Gong X, Shur BD (1994) *J Cell Biol* **126**: 1573–83.
14 Youakim A, Dubois DH, Shur BD (1994) *Proc Nat Acad Sci* **91**: 10913–17.
15 Cameron HS, Szczepaniak D, Weston BW (1995). *J Biol Chem* **270**: 20112–22.
16 Hopp TP, Pricket KS, Price V, Libby RT, March CJ, Ceretti P, Urdal DL, Conlon PJ (1988) *Biotech* **6**: 1205–10.
17 Ko AI, Drager UC, Harn DA (1990) *Proc Natl Acad Sci* **87**: 4159–63.
18 Wiederschain G, Koul O, McCluer RH (1993) *Glycoconj J* **12**: S9, 442.
19 Kozak M (1996) *Mammalian Genome* **7**: 563–74.
20 Briggs MR, Kadonaga JT, Bell SP, Tjian R (1986) *Science* **234**: 47–52.
21 Rajput B, Shaper NL, Shaper JH (1996) *J Biol Chem* **271**: 5131–42.
22 Potvin B, Kumar R, Howard DR, Stanley P (1990) *J Biol Chem* **265**: 1615–22.
23 Goelz SE, Hession G, Goff D, Griffiths B, Tizard R, Newman B, Chi-Rosso G, Lobb R (1990) *Cell* **63**: 1349–56.
24 Lowe JB, Kukowska-Latallo JF, Nair RP, Larsen RD, Marks RM, Macher BA, Kelly RJ, Ernst LR (1991) *J Biol Chem* **266**: 17467–77.
25 Gersten KM, Natsuka S, Trinchera M, Petrynia B, Kelly RJ, Hiraiwa N, Jenkins NA, Gilbert DJ, Copeland NG, Lowe JB (1995) *J Biol Chem* **270**: 25047–56.
26 Kozak M (1984) *Nucleic Acids Res* **12**: 857–72.
27 Nilsson T, Lucocq JM, Mackay D, Warren G (1991) *EMBO J* **10**: 3567–75.
28 Aoki D, Lee N, Yamaguchi N, Dubois C, Fukada MN (1992) *Proc Natl Acad Sci* **89**: 4319–23.
29 Russo RN, Shaper NL, Taatjes DJ, Shaper JH (1992) *J Biol Chem* **267**: 9241–7.
30 Teasdale RD, D'Agostarno G, Gleeson PA (1992) *J Biol Chem* **267**: 4084–96.
31 Masibay AS, Balaji PV, Boeggeman EE, Qasba PK (1993) *J Biol Chem* **268**: 9908–16.
32 Munro S (1991) *EMBO J* **10**: 3577–88.
33 Colley KJ, Lee EU, Paulson JC (1992) *J Biol Chem* **267**: 7784–93.
34 Wong SH, Low SH, Hong W (1992) *J Cell Biol* **117**: 245–58.
35 Burke J, Pettit JM, Schachter H, Sarkar M, Gleeson PA (1992) *J Biol Chem* **267**: 24433–40.
36 Tang BL, Wong SH, Low SH, Hong W (1992) *J Biol Chem* **267**: 10122–8.
37 Yamaguchi N, Fukuda MN (1995) *J Biol Chem* **270**: 12170–6.
38 Weisz OA, Swift AM, Machamer CE (1993) *J Cell Biol* **122**: 1185–96.
39 Machamer CE (1993) *Curr Opin Cell Biol* **5**: 602–12.
40 Nilsson T, Warren G (1994) *Curr Opin Cell Biol* **6**: 517–21.
41 Bretsher MS, Munro S (1993) *Science* **261**: 1280–1.
42 Pelham HS, Munro S (1993) *Cell* **75**: 603–5.
43 Munro S (1995) *Biochem Soc Trans* **23**: 527–30.
44 Lopez LC, Youakim A, Evans SC, Shur BD (1991) *J Biol Chem* **266**: 15984–91.
45 Nguyen TTM, Hinton DA, Shur BD (1994) *J Biol Chem* **269**: 28000–9.
46 Evans SC, Youakim A, Shur BD (1995) *BioEssays* **17**: 261–8.
47 Evans S., Lopez LC, Shur BD (1993) *J Cell Biol* **120**: 1045–57.

Received 5 August 1997, revised 28 January 1998, accepted 3 February 1998