Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 309 (2003) 166–174
BBRC
www.elsevier.com/locate/ybbrc

A novel human glycosyltransferase: primary structure and characterization of the gene and transcripts

Taisto Y.K. Heinonen,[a,*] Leena Pasternack,[a] Katri Lindfors,[a] Christelle Breton,[b] Louis N. Gastinel,[c] Markku Mäki,[a] and Heikki Kainulainen[d]

[a] *Paediatric Research Centre, University of Tampere Medical School and Tampere University Hospital, Tampere, Finland*
[b] *Centre de Recherches sur les Macromolécules Végétales, Centre National de la Recherche Scientifique (affiliated with Université Joseph Fourier), Grenoble, France*
[c] *EA 3176 Glycobiologie Végétale et Biotechnologies, Institut des Sciences de la Vie et de la Santé, Université de Limoges, Limoges, France*
[d] *Institute of Medical Technology, University of Tampere, Tampere, Finland*

Received 22 July 2003

Abstract

We report the identification and primary structure of a novel human glycosyltransferase, B3GTL (β3-glycosyltransferase-like). The 498 residue protein consists of a short cytoplasmic N-terminal "tail" (residues 1–4), a single transmembrane domain with type II topology (residues 5–28), a "stem" region (residues 29–260), and a catalytic domain (residues 261–498). The genomes of *Anopheles gambiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans* encode potential orthologs which share 31–39% sequence identity with B3GTL, as well as the following features: a conserved catalytic domain containing a triple aspartate motif (DDD) at its core, a conserved pattern of cysteine residues, a C-terminal KDEL-like motif, and conserved residues and motifs that affiliate this novel group with a family of β3-glycosyltransferases (GT31 in the CAZY classification). The B3GTL gene lacks canonical TATA and CAAT boxes and contains three functional polyadenylation sites. It is transcribed in a wide range of tissues and in TGF-β-treated T84 epithelial cells.

Keywords: Glycosyltransferase; Catalytic domain; cDNA clone; Gene organization

Many eukaryotic proteins and lipids are modified with specific oligosaccharide side chains, called glycans, before they acquire full biological activity. The various glycans display enormous functional diversity, as they are involved in numerous processes in development, normal physiology, and disease [1–3], ranging from cellular recognition, adhesion, and signalling [4–6] to drug clearance [7] and quality control of protein folding [8]. Underlying this functional diversity is a vast diversity of glycan structures, which display varying degrees of complexity and have a large capacity for information storage [9].

The various glycans are synthesized by the sequential addition of monosaccharide units in pathways catalyzed by series of glycosyltransferases, enzymes that function primarily in the ER and the Golgi apparatus. More than 200 glycosyltransferase activities or putative enzyme sequences have already been isolated from human cells, but the full ensemble of the enzymes responsible for establishing the human "glycome" remains to be identified. Only a subset of the total enzyme repertoire is active in any particular cell, where it forms the machinery responsible for the synthesis of glycans specific to that cell.

In epithelial cells, glycosyltransferases are involved in such traditionally recognized functions as cellular adhesion and the synthesis of mucins, specialized and highly glycosylated proteins [10]. Recently, glycosyltransferases have been shown to have more diverse roles in processes ranging from epithelial invagination during vulval morphogenesis in *Caenorhabditis elegans* [11] to signal transduction during ovarian follicle development in *Drosophila melanogaster* [12] and sensory

* Corresponding author. Fax: +358-3-215-8420.
E-mail address: taisto.heinonen@uta.fi (T.Y.K. Heinonen).

0006-291X/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S0006-291X(03)01540-7

hair cell differentiation in the mammalian inner ear [13]. Of particular interest is the "knock-out" mouse, null for β-1,4-galactosyltransferase I, whose phenotype included abnormal differentiation of epithelial cells in the skin and the small intestine [14].

We are interested in finding genes that are involved in the differentiation of epithelial cells along the crypt-villus axis in the small intestine. To this end, we have been using a model system in which cultured T84 human intestinal epithelial cells are induced to differentiate in response to TGF-β [15]. Analysis of gene expression in this system by differential display-PCR led to the discovery of a cDNA clone which encodes a novel protein with sufficient sequence similarity to β3-glycosyltransferases to allow us to tentatively identify it as a novel glycosyltransferase, B3GTL (β3-glycosyltransferase-like), pending biochemical characterization of its activity and function.

Materials and methods

Isolation of cDNA clones. The cDNA clone that led to the discovery of B3GTL was originally identified in cultured T84 cells as described previously [16] and designated 5G1Ca-cl2. The sequence of 5G1Ca-cl2, 446 bp in length, matched the 3′-terminal sequence of a human EST clone (clone id ys88g12, IMAGE id 221926), which was obtained from the IMAGE Consortium and sequenced. The insert contained an open reading frame of 497 codons but lacked an initiator methionine codon. In order to isolate full-length clones, we screened the human heart, brain, and kidney Rapid Screen cDNA Library Panels LHT-1001, LAB-1001, and LKD-1001 (OriGene Technologies, Rockville, Maryland, USA) according to the manufacturer's instructions, using the primers 5′-CATGAGATTGCCCTCTACATCTGG-3′ and 5′-TTCCC AAATCCACAGTAGGAATGG-3′ (Genset Oligos, Paris, France), which amplify a PCR product corresponding to positions 655–946 of the sequence in Fig. 1. The PCR conditions consisted of an initial 2 min incubation at 94 °C followed by 40 cycles of 1 min at 94 °C, 1 min at 60 °C, and 1 min at 72 °C (each primer at 1.4 μM, each dNTP at 0.25 mM final concentration). The library master plates were screened using PfuTurbo DNA polymerase (Stratagene, La Jolla, California, USA) but the "sub-plates" were screened using Dynazyme (Finnzymes Oy, Espoo, Finland). Four independent clones were isolated from the heart library, one from the brain library, and one from the kidney library, and characterized by sequencing.

DNA sequencing. DNA sequencing reactions were performed using BigDye terminator v2.0 mix and run on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, California, USA).

Sequence analysis. DNA sequences were translated using the EditSeq program and protein sequences were aligned using ClustalW in the MegAlign program of Lasergene Software Suite, version 5 (DNASTAR, Madison, Wisconsin, USA). The phylogenetic tree was constructed using MegAlign. Database searches were done using BLAST at (http://www.ncbi.nlm.nih.gov/). O-glycosylation sites were predicted by the "NetOGlyc 2.0 Prediction Server" at (http://www.cbs.dtu.dk/services/NetOGlyc/). Transmembrane domains were predicted with programs at (http://www.hgmp.mrc.ac.uk/Registered/Webapp/pix/). Secondary structure predictions were obtained using the Jnet program at (http://www.compbio.dundee.ac.uk./~www-jpred/submit.html) and programs available at (http://npsa-pbil.ibcp.fr/). Potential binding sites for transcription factors were identified by the MatInspector V2.2 program on the TRANSFAC 4.0 database at (http://transfac.gbf.de/). Only binding sites with matrix similarity scores >0.9 were considered, with the exception of the Sp1 sites which were identified at a cutoff score of >0.8. Binding sites for Smads were identified manually by scanning the sequence for AGAC motifs. Alignment of glycosyltransferase sequences currently listed under family GT31 at (http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html) was performed using ClustalW.

Northern hybridization. The probe corresponds to nucleotides 2899–3283 of the sequence in Fig. 1 and therefore detects transcripts that contain the two longest 3′-UTRs (b and c). The probe fragment was generated by PCR, purified from an agarose gel, labelled using the Rediprime II system and Redivue [α-^{32}P]-dCTP (Amersham–Pharmacia Biotech Europe GmbH, Freiburg, Germany), and hybridized to a human multiple tissue Northern blot (7780-1, Clontech, Palo Alto, California, USA) in ExpressHyb hybridization solution (Clontech).

Analysis of B3GTL mRNA in TGF-β-treated T84 epithelial cells by real-time quantitative PCR. T84 cells were induced to differentiate in culture by TGF-β treatment according to [15] and total RNA was isolated from treated and control (untreated) cells after seven days as described previously [16]. DNase I-treated total RNA was reverse transcribed using SuperScript II (Gibco BRL, Paisley, UK), and the resulting cDNA was subjected to real-time PCR using the LightCycler-FastStart DNA Master Hybridization Probes kit (Roche Molecular Biochemicals, Mannheim, Germany) according to the manufacturer's instructions. To assay for B3GTL mRNA, thermocycling reactions contained 0.5 μM of the unlabeled primers (5′-GCACTGGTGGCTAC AGCT-3′ and 5′- TGGAACGATATGGGAACTTGAT-3′), 0.2 μM of fluorescein-labelled (5′-AAGTCTCCTGACGGCTTCTCTGCTG-3′) and LC Red640-labelled (5′-GACCATTCCTCCTCCTCCCGTG AT-3′) (TIB MOLBIOL, Berlin, Germany) as probes, and 4 mM $MgCl_2$. The cycling conditions consisted of an initial incubation of 10 min at 96 °C followed by 55 cycles of 10 s at 96 °C, 10 s at 57 °C, and 10 s at 72 °C, and were performed on the LightCycler System (Roche Molecular Biochemicals, Mannheim, Germany). To assay for the mRNA encoding the housekeeping enzyme glyceraldehyde-3-phosphate dehydrogenase (GAPDH), the reactions contained 1 μM of each primer (5′-TGGTATCGTGGAAGGACTCATGAC-3′ and 5′-AT GCCAGTGAGCTTCCCGTTCAGC-3′) and 2 mM $MgCl_2$, and the cycling conditions consisted of an initial incubation of 10 min at 95 °C followed by 45 cycles of 20 s at 95 °C, 5 s at 70 °C, and 8 s at 72 °C. The relative amounts of B3GTL and GAPDH cDNA were calculated from standard curves generated by serial dilution of cDNA produced from T84 cells. The amount of B3GTL cDNA in each experimental sample was normalized to the amount of GAPDH cDNA in the same sample. The ratio of B3GTL cDNA in the experimental samples (TGF-β-treated versus control) was calculated from the normalized values.

Results and discussion

The cDNA sequence identifies a novel protein with significant similarity to β3-glycosyltransferases

Fig. 1 presents a summary of sequence data (GenBank Accession No. AY190526) from six independent cDNA clones obtained from three different sources. All clones contain the same open reading frame for a novel protein (identifier: AAO37647), 498 amino acid residues in length, which displays an overall architecture characteristic of glycosyltransferases. It is a type II membrane protein, consisting of a short cytoplasmic domain (residues 1–4), a single transmembrane domain (residues 5–28), and a large C-terminal portion that is presumed

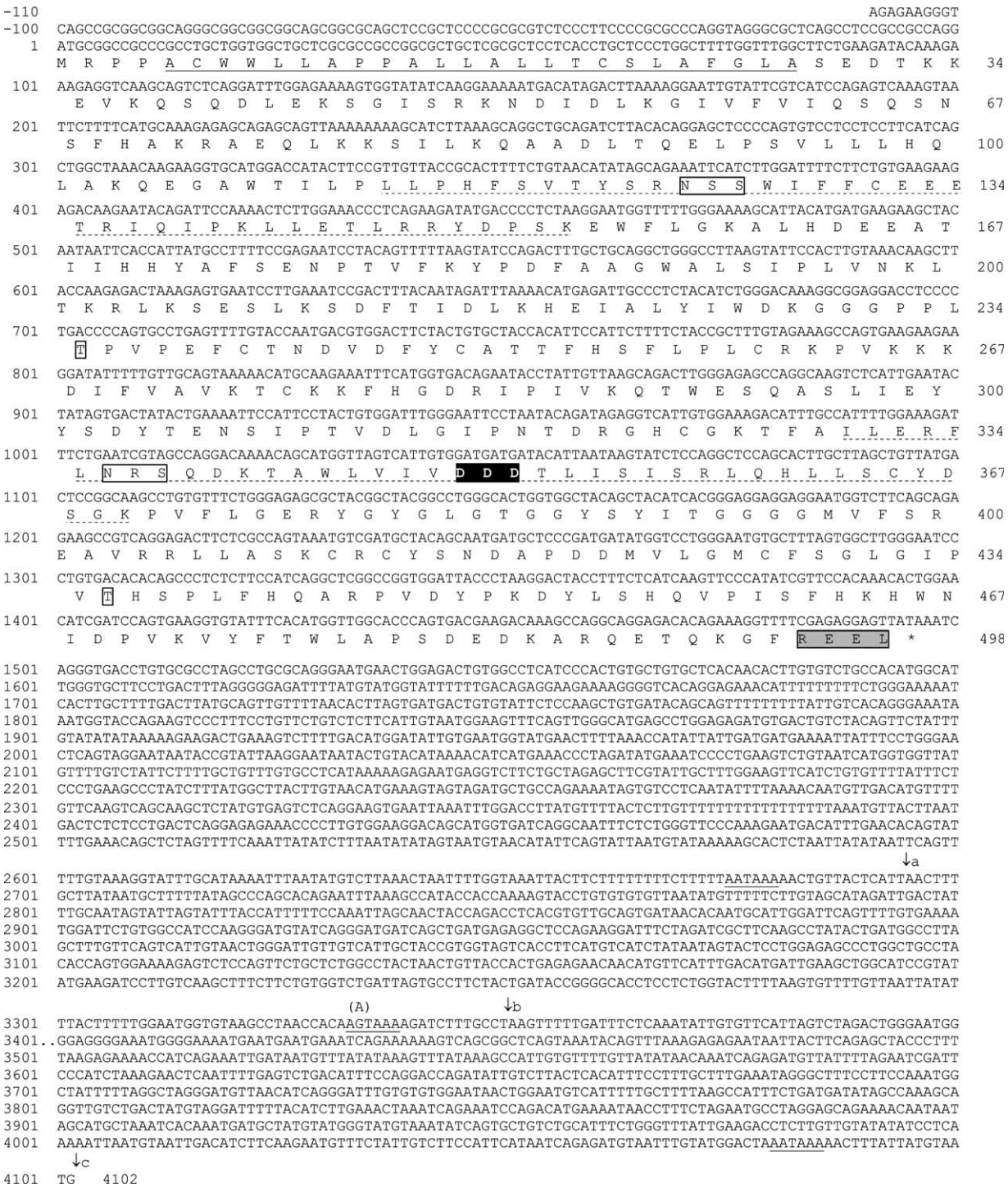

Fig. 1. cDNA sequence and the derived amino acid sequence of B3GTL. Nucleotide numbering is shown on the left, the translated sequence is shown under the nucleotide sequence, and the amino acids are numbered on the right. The transmembrane domain is underlined, potential N- and O-glycosylation sites are boxed in white, the DDD triplet at the core of the catalytic domain is boxed in black, the putative KDEL retention signal is boxed in grey, and the stop codon is indicated by the asterisk. The two regions that share sequence similarity (see text) are indicated by dashed underlines. The three polyadenylation signals are underlined and the three 3′-termini of mRNAs are indicated by the vertical arrows (a, b, and c). A nucleotide polymorphism is indicated above the sequence in brackets.

to reside in the lumen of the Golgi apparatus or the ER. This lumenal portion can be divided into two parts of roughly equal size, a "stem" region (residues 29–260) and a putative catalytic domain (residues 261–498) which was identified by its obvious sequence similarity to the catalytic domains of certain β3-glycosyltransferases, as discussed in more detail below. Based on this similarity, the protein was named B3GTL (β3-glycosyl-

transferase-like). The protein sequence contains two potential N-glycosylation sites (N-124 and N-336), two potential O-glycosylation sites (T-235 and T-436), and a variant (REEL) of the KDEL retention signal in its very carboxyl terminus.

The classification of glycosyltransferases is ultimately based on enzymatic activity, which includes substrate specificity and reaction mechanism, but putative enzymes are classified on the basis of amino acid sequence similarities before their activities are known. Based on the amino acid sequence of its catalytic domain, B3GTL is most closely affiliated with family GT31 in a sequence-based classification that currently includes 66 families of glycosyltransferases (http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html). Family GT31 comprises β3-galactosyltransferases, β3-*N*-acetylgalactosaminyltransferases, and β3-*N*-acetylglucosaminyltransferases that are involved in the synthesis of various complex glycans in mammalian glycoproteins and glycolipids, and several members of the family participate in the synthesis of mucin-type O-glycans. Although members of family GT31 are enzymes with different substrate specificities, their catalytic domains are characterized by certain conserved features. A comprehensive analysis of sequences currently listed under family GT31 reveals 12 amino acid positions that are highly conserved or invariant throughout the family. These 12 positions represent the most highly conserved positions in larger conserved sequence motifs and five such conserved motifs (I–V) can be identified in members of family GT31. The B3GTL sequence contains all 12 highly conserved residues and the five conserved motifs, whose locations are shown in Fig. 2A. In addition to the features conserved widely in family GT31, the catalytic domain of B3GTL shows more extensive sequence conservation with specific family members. It shares 28% and 27% sequence identity, respectively, with radical Fringe (a β3-*N*-acetylglucosaminyltransferase) and core 1 β3-galactosyltransferase (see Fig. 2), which are B3GTL's closest relatives that are associated with a known enzymatic activity.

The core of the catalytic domain in β3-glycosyltransferases contains a highly conserved triplet of aspartate residues (DDD), which is a variant of the more general DXD motif (X denotes any amino acid) found in many families of glycosyltransferases [17–19]. This DXD motif forms a key part of the catalytic domain as it has been shown to be involved in binding of the UDP-donor sugar and the divalent cation (Mn^{2+}) cofactor [20–22]. In the B3GTL sequence, ^{349}DDD^{351} can be identified as the DXD motif of the catalytic core, because it corresponds to the acidic triplet in the other β3-glycosyltransferases, whose catalytic domains can be aligned relatively unambiguously with B3GTL (Fig. 2A). Furthermore, ^{349}DDD^{351} is part of a larger conserved motif (motif II, Fig. 2A) that includes the four highly conserved residues identified in motif II of β3-glycosyltransferases.

The amino acid composition and certain secondary structural features in the vicinity of ^{349}DDD^{351} are consistent with the structure of the catalytic pocket in other glycosyltransferases. ^{349}DDD^{351} is flanked by ^{345}LVIV^{348} and ^{352}TL^{353} and thus conforms to the hhhhDXDXh consensus sequence (h denotes hydrophobic residues) that is conserved in many glycosyltransferase families [17]. Furthermore, the fifth residue N-terminal to ^{349}DDD^{351} is W^{344}, consistent with the conservation of an aromatic residue at the corresponding position in enzymes of family GT31. Secondary structure prediction methods indicate that in B3GTL, the residues immediately preceding (343–348) and those following (353–355) the ^{349}DDD^{351} motif could form short β-strands. This is consistent with data on those DXD-containing glycosyltransferases whose crystal structures have been determined. In these enzymes the DXD motif is flanked on each side by a short β-strand, the first of which consists of apolar or hydrophobic residues and forms a part of the mixed central β-sheet that constitutes the hydrophobic core of the protein [20,21,23].

The "stem" region (amino acids 29–260) of B3GTL contains several interesting features. First, it is predicted to have a high content of secondary structure elements (data not shown), suggesting it may form a structurally ordered, perhaps globular, domain in the protein. Second, it is 232 residues in length and therefore almost exactly the same size as the catalytic domain (236 residues), with which it actually shares weak sequence similarity. This intriguing internal similarity is seen in a comparison of residues 113–153 in the stem with residues 330–370, the region of the catalytic domain that forms a part of the active site (these regions are indicated in Fig. 1). Residues 330–370 include four of the 12 highly conserved residues found in family GT31: W^{344}, D^{349}, D^{351}, and L^{363} (Fig. 2A). All four are conserved in region 113–153, as W^{127}, E^{132}, E^{134}, and L^{146}, respectively, and the sequence immediately N-terminal to the acidic triplet ^{132}EEE^{134} also contains hydrophobic residues and a potential N-glycosylation site, as does the active site in the catalytic domain. The presence of an active site-like sequence in the stem region is curious, and it may reflect actual similarity in function, such as carbohydrate recognition/binding, or in fact the presence of a second active site. There are several examples of glycosyltranferases that have two active sites and two catalytic activities: heparan synthase, hyaluronan synthase, and chondroitin synthase (a member of family GT31). However, the vast majority of glycosyltransferases contain only one active site and one catalytic activity, and in these enzymes the stem region has been suggested to function in subcellular localization [24], acceptor glycoprotein discrimination [25], and oligomerization [26].

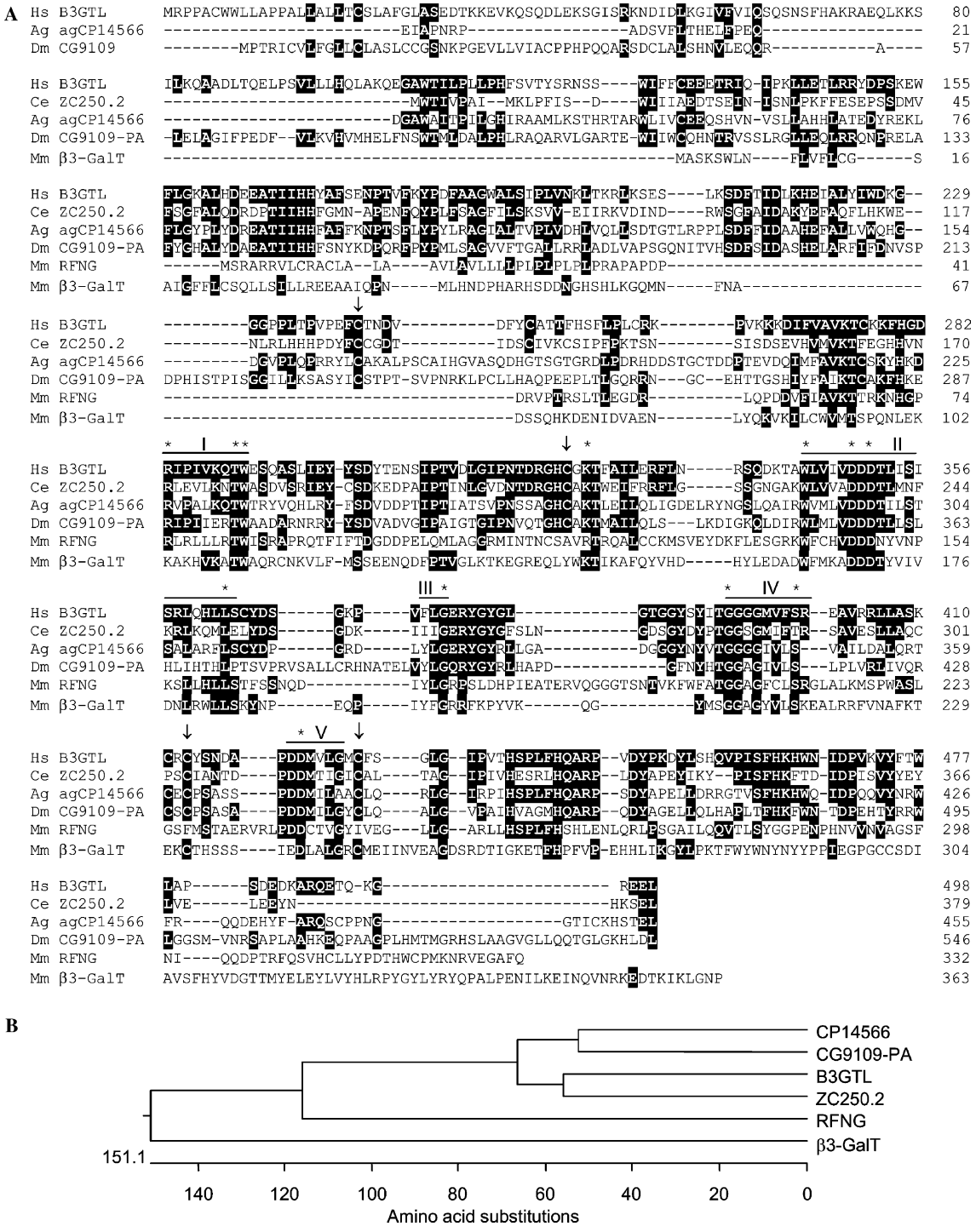

Fig. 2. (A) Multiple sequence alignment of B3GTL and closely related proteins. The species name is shown on the left (Hs, *Homo sapiens*; Ce, *Caenorhabditis elegans*; Ag, *Anopheles gambiae*; Dm, *Drosophila melanogaster*; and Mm, *Mus musculus*), followed by the name of the protein. ZC250.2 (GenBank Accession No. AAB54234), agCP14566 (EAA05853), and CG9109-PA (AAF52326) are hypothetical proteins identified by translation of genomic sequences; RFNG, radical Fringe (AAC53261); β3-GalT, core 1 UDP-galactose:*N*-acetylgalactosamine β1,3-galactosyltransferase (AAF81982). The numbering of amino acid residues is shown on the right, the residues that are identical with B3GTL at the corresponding position are boxed in black, and gaps introduced during the alignment are indicated by hyphens. The cysteine residues conserved among B3GTL, ZC250.2, agCP14566, and CG9109-PA are shown by vertical arrows. The 12 residues and the five motifs (I–V) conserved in β3-glycosyltransferases are indicated by asterisks and overlines, respectively. (B) A phylogenetic tree that relates the six protein sequences in terms of the number of amino acid substitutions that separate them.

Identification of potential orthologs of B3GTL

We have identified a number of homologs of B3GTL in various cDNA entries and translated genomic sequences in GenBank. These homologs form a closely related group of sequences which share with B3GTL much more similarity than the known glycosyltransferases discussed above. This group is confined to multicellular organisms only and includes entries from many mammalian species, the sequences of which are virtually identical to B3GTL, but none of them has had an activity or function ascribed to it yet. It also includes the more distant, hypothetical proteins agCP14566, ZC250.2, and CG9109-PA, identified in the genomic sequences of *Anopheles gambiae*, *C. elegans*, and *D. melanogaster*, respectively. These sequences are similar enough to be expected to share the same or similar function, yet they have undergone sufficient evolutionary divergence to allow the identification of conserved residues and motifs that may pinpoint functionally important parts of the protein.

As shown in Fig. 2A, the sequences of agCP14566, ZC250.2, and CG9109-PA can be aligned relatively unambiguously over their entire lengths with B3GTL, with which they share 37%, 39%, and 31% sequence identity, respectively. The putative catalytic domains (corresponding to B3GTL residues 261–498) are also well conserved, with respective sequence identities of 42%, 40%, and 35%. All four members of this group contain the 12 highly conserved residues and the five sequence motifs that are conserved in β3-glycosyltransferases (as discussed in the previous section). In addition, several invariant residues and highly conserved sequence blocks unique to B3GTL and its homologs can be identified in the catalytic domains as well as in the stem regions. However, the stem regions are somewhat more variable than the catalytic domains and account for most of the variation, in both sequence and length, between these four proteins. This is particularly obvious in the case of ZC250.2 which, at 379 residues, is considerably shorter than the other proteins (455–546 residues).

The pattern of cysteine residues is an important primary structural feature that is conserved in B3GTL, CG9109-PA, agCP14566, and ZC250.2. B3GTL contains a total of 12 cysteine residues, two in the transmembrane domain, four in the "stem" region, and six in the catalytic domain. Of the six cysteine positions in the catalytic domain, three are conserved in all four proteins and all six are conserved between B3GTL and agCP14566. Of the four cysteine positions in the "stem" region, one is conserved in all four proteins and one is conserved between B3GTL, agCP14566, and CG9109-PA in the alignment shown in Fig. 2A. ZC250.2 shows the least conservation of cysteine residues, which is particularly obvious in the stem region, perhaps owing to the fact that it is shorter than the stems in the other proteins. Overall, the cysteine pattern shows remarkable conservation between these four proteins, suggesting that they could share a common pattern of disulfide bridges and therefore other structural similarities. It is noteworthy that two of the cysteine residues (corresponding to C^{413} and C^{427} in B3GTL) are also conserved in core 1 β3-galactosyltransferases [27].

The presence of a motif similar to the KDEL retention signal is another feature shared by B3GTL and its homologs. All four proteins contain a carboxyl-terminal, four-residue, KDEL-like sequence: REEL, STEL, HLDL, and KSEL in B3GTL, agCP14566, CG9109-PA, and ZC250.2, respectively. Out of these, only the REEL of B3GTL fits the (KRHQSA)-(DENQ)-E-L consensus sequence perfectly, but the other three sequences are similar enough to suggest an actual shared motif. The functional significance of this motif is not clear, however. The conventional function of the KDEL signal is to retain soluble proteins in the lumen of the ER, and glycoprotein glucosyltransferase provides an example of a KDEL signal-containing, ER-resident glycosyltransferase [28]. However, most glycosyltransferases are type II membrane proteins found predominantly in the Golgi apparatus. In these glycosyltransferases, a KDEL-like sequence is very rare, having been found in only two [29,30]. Thus the presence of the KDEL-like motif provides another distinguishing characteristic for B3GTL and its homologs.

Structure of the B3GTL mRNA

The structure of the B3GTL mRNA is shown in Fig. 1, which presents a cDNA sequence that summarizes data from six independent cDNA clones. This sequence is likely to represent the full length mRNA because its length matches the length of one of the transcripts detected on a Northern blot, as discussed in more detail below. The 5′-untranslated region (5′-UTR) of the mRNA is 110 nucleotides long, although the 5′-termini of the clones show minor heterogeneity, as discussed below (see Fig. 3).

The cDNA clones contain three types of 3′-untranslated region (3′-UTR) which result from the use of three different cleavage/polyadenylation sites, 1197 bp (3′-UTRa), 1853 bp (3′-UTRb), and 2605 bp (3′-UTRc) following the stop codon. Each site is preceded by an AATAAA polyadenylation signal 18–23 bp before the beginning of the poly(A) tail (Fig. 1). This is the first example of the use of three different 3′-UTRs by a glycosyltransferase gene, and such differential processing may have implications for tissue- or cell type-specific differences in mRNA stability and/or regulation of translation.

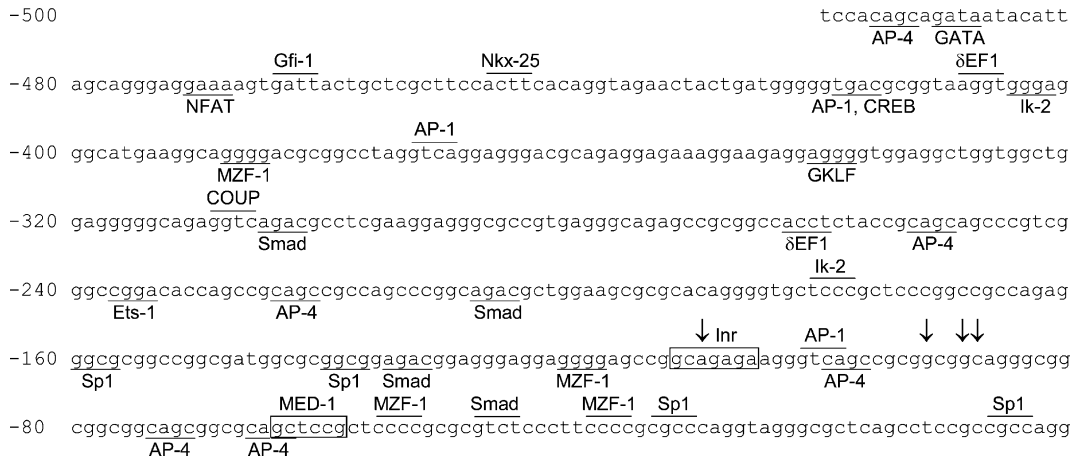

Fig. 3. The proximal promoter and 5′-UTR of the B3GTL gene. Nucleotide numbering, in negative numbers counting upstream from the ATG codon, is shown on the left. Putative transcription start sites are indicated by vertical arrows and the two sequence elements discussed in the text are boxed. Potential binding sites for selected transcription factors are underlined where the recognition sequence has sense strand polarity and overlined where it has antisense polarity. For simplicity, only four nucleotides of the core of the recognition sequence are indicated. The abbreviations for the names of sequence elements and transcription factors are the same as in the text.

The B3GTL gene

The B3GTL gene, which was identified by searching the cDNA sequence against the draft human genome sequence (build 29), is located on chromosome 13 at 13q12.3. The mRNA-coding region spans a distance of 132,188 bp (from the initiator ATG to the 3′-end of the longest 3′-UTR) and 14 introns divide it into 15 exons. The lengths of the exons and introns are shown in Table 1.

The promoter region of the B3GTL gene, tentatively identified in the genomic sequence as the area in the immediate vicinity of the transcription initiation site(s),

Table 1
The exons and introns of the B3GTL gene

Exon	cDNA positions	Length (bp)	Amino acids	Exon/Intron	Intron	Length (bp)	Intron/Exon
				S L[23]			A[24] F
1	1–70	70	1–23	TCC CTG G gt–	1	14,896	–ag CT TTT
				S Q[40]			D[41] L
2	71–120	50	24–40	TCT CAG gt–	2	7851	–ag GAT TTG
				D I[53]			D[54] L
3	121–160	40	41–53	GAC ATA G gt–	3	6193	–ag AC TTA
				T Q[90]			E[91] L
4	161–270	110	54–90	ACA CAG gt–	4	17,728	–ag GAG CTC
				P H[116]			F[117] S
5	271–347	77	91–116	CCG CA gt–	5	755	–ag C TTT TCT
				S K[153]			E[154] W
6	348–459	112	117–153	TCT AAG gt–	6	12,979	–ag GAA TGG
				N K[199]			L[200] T
7	460–596	137	154–199	AAC AA gt–	7	8131	–ag G CTT ACC
				H E[220]			I[221] A
8	597–660	64	200–220	CAT GAG gt–	8	5231	–ag ATT GCC
				L C[260]			R[261] K
9	661–780	120	221–260	CTT TGT gt–	9	2073	–ag AGA AAG
				D R[283]			I[284] P
10	781–850	70	261–283	GAC AGA A gt–	10	7876	–ag TA CCT
				D R[321]			G[322] H
11	851–964	114	284–321	GAT AGA G gt–	11	1958	–ag GT CAT
				I S[355]			I[356] S
12	965–1064	100	322–355	ATA AG gt–	12	30,746	–ag T ATC TCC
				G G[395]			M[396] V
13	1065–1184	120	356–395	GGA GG gt–	13	6065	–ag A ATG GTC
				H Q[443]			A[444] R
14	1185–1329	145	396–443	CAT CAG gt–	14	5605	–ag GCT CGG
15	1330–4102	2773	444–498 (+3′-UTR)				

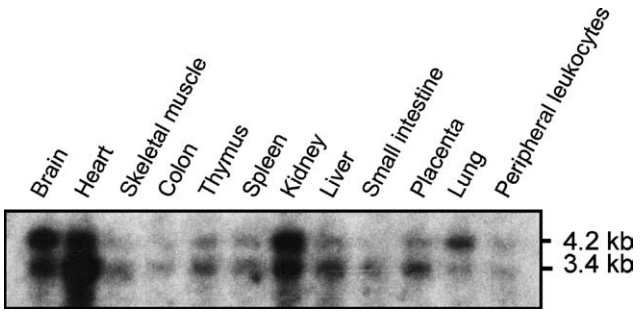

Fig. 4. B3GTL mRNAs in human tissues analyzed by Northern hybridization. The tissue source of mRNA is indicated at the top and the sizes of the transcripts are indicated on the right. The probe is specific to the two longest 3′-UTRs.

is shown in Fig. 3. This region and the 5′-UTR are unusually (G + C)-rich, with a (G + C) content of 75%. One putative transcription initiation site was identified at position -110 (relative to the ATG initiator codon), based on the 5′-termini of two independent cDNA clones. The sequence at -110 (GCAGAGA) matches the loose consensus sequence of the transcriptional initiator element Inr, Py Py A_{+1} N (A/T) Py Py [31], at four of the seven positions, including the three most critical nucleotides: A at +1, A or T at +3, and a pyrimidine (Py) at −1. The sequence in the vicinity of this initiation site does not contain canonical TATA or CAAT boxes, but contains three potential binding sites for the general transcription factor Sp1 (at positions -164, -144, and -89). These sites are noteworthy because Sp1 is known to be involved in the regulation of many TATA-less promoters, which often contain multiple binding sites for this factor and generally use an Inr to position the start of transcription [31,32]. A downstream element, MED-1 (GCTCCG), found in a class of TATA-less promoters that utilize multiple transcription initiation sites [33], is present at -64. This is of interest because the B3GTL gene may use multiple transcription start sites, located between nucleotide positions -110 and -88 bp, as shown in Fig. 3.

The proximal promoter region harbors potential binding sites for the following transcriptional regulators (Fig. 3): GATA binding proteins (GATA), nuclear factor of activated T-cells (NFAT), the proto-oncoprotein Gfi-1, the cardiac transcription factor Nkx-25, c-AMP response element binding protein (CREB), activator protein-1 (AP-1), the transcriptional repressor δEF1 (δEF1), Ikaros factor (Ik-2), myeloid zinc finger protein-1 (MZF-1), gut-enriched Kruppel-like factor (GKLF), chicken ovalbumin upstream binding protein (COUP), Ets-1, Smads, and the general transcriptional activator AP-4. This list includes several differentiation-specific factors known to regulate gene expression in blood cells, cardiac cells or epithelial cells, and potentially allows the B3GTL gene to be regulated by diverse signals. In the context of the present study, where the amount of B3GTL mRNA was found to increase in TGF-β-treated T84 cells (see below), the architecture of the promoter suggests a potential mechanism of transcriptional activation in response to TGF-β. The promoter contains several potential binding sites for Smads, proteins that transduce TGF-β signals and activate transcription by interacting with Sp1 [34,35] or AP-1 [36] at the promoter of the target gene.

The B3GTL mRNA is expressed in many human tissues and in TGF-β-treated T84 epithelial cells

Northern hybridization analysis of adult human mRNA revealed that the B3GTL gene is transcribed in all the tissues that were examined (Fig. 4). A probe specific to the two longer 3′-UTRs detected two transcripts, 4.2 and 3.4 kb in length. The 4.2 kb transcript corresponds to the full length cDNA sequence, containing 3′-UTRc, shown in Fig. 1 (total length 4102 nucleotides, not including the poly(A) tail). The 3.4 kb transcript carries 3′-UTRb (total length 3460 nucleotides). The tissues show considerable differences in mRNA levels and the relative levels of the two transcripts also differ in a tissue-specific manner (Fig. 4). These data indicate that the B3GTL mRNAs are subjected to strong tissue- or cell type-specific regulation.

We are interested in the modulation of gene expression in T84 intestinal epithelial cells by TGF-β, which induces these cells to differentiate in culture from secretory cells into absorptive enterocytes [15]. Analysis of mRNA in this system by real-time quantitative PCR indicated that the amount of B3GTL mRNA increased 3.1-fold in TGF-β-treated cells relative to control cells after seven days (the ratios from three independent experiments were 2.4, 3.8, and 3.1). This increase is consistent with the architecture of the promoter of the B3GTL gene, which suggests a potential mechanism of signal transduction and transcriptional activation in response to TGF-β (as discussed above), and implies that B3GTL may play some role in the TGF-β-induced differentiation of T84 epithelial cells in culture. An understanding of the exact role of B3GTL in epithelial cells, and in other cells and tissues, will have to await characterization of its activity, substrate(s), and cellular function(s).

Acknowledgments

This study has been funded by the Commission of the European Communities, specific RTD programme "Quality of Life and Management of Living Resources", QLK1-CT-1999-00037 "Evaluation of the prevalence of the coeliac disease and its genetic components in the European population". It does not necessarily reflect the Commission's views and in no way anticipates the Commission's future policy in this area. This work is also supported by the Academy of Finland Research Council for Health (funding decision numbers 73489 and 201361), the Päivikki and Sakari Sohlberg Foundation, the Foundation of the

Friends of the University Children's Hospitals in Finland, the Medical Research Fund of Tampere University Hospital, and the Foundation for Paediatric Research in Finland.

References

[1] T. Muramatsu, Essential roles of carbohydrate signals in development, immune response and tissue functions, as revealed by gene targeting, J. Biochem. 127 (2000) 171–176.
[2] J.W. Dennis, M. Granovsky, C.E. Warren, Glycoprotein glycosylation and cancer progression, Biochim. Biophys. Acta 1473 (1999) 21–34.
[3] J. Jaeken, G. Matthijs, Congenital disorders of glycosylation, Annu. Rev. Genom. Hum. Genet. 2 (2001) 129–151.
[4] D.J. Moloney, V.M. Panin, S.H. Johnston, J. Chen, L. Shao, R. Wilson, Y. Wang, P. Stanley, K.D. Irvine, R.S. Haltiwanger, T.F. Vogt, Fringe is a glycosyltransferase that modifies Notch, Nature 406 (2000) 369–375.
[5] N.E. Zachara, G.W. Hart, The emerging significance of O-GlcNAc in cellular regulation, Chem. Rev. 102 (2002) 431–438.
[6] S. Hakomori, The glycosynapse, Proc. Natl. Acad. Sci. USA 99 (2002) 225–232.
[7] M.B. Fisher, M.F. Paine, T.J. Strelevitz, S.A. Wrighton, The role of hepatic and extrahepatic UDP-glucuronosyltransferases in human drug metabolism, Drug Metab. Rev. 33 (2001) 273–297.
[8] C.M. Cabral, Y. Liu, R.N. Sifers, Dissecting glycoprotein quality control in the secretory pathway, Trends Biochem. Sci. 26 (2001) 619–624.
[9] G. Reuter, H.-J. Gabius, Eukaryotic glycosylation: whom of nature or multipurpose tool?, Cell. Mol. Life Sci. 55 (1999) 368–422.
[10] J. Dekker, J.W.A. Rossen, H.A. Buller, A.W.C. Einerhand, The MUC family: an obituary, Trends Biochem. Sci. 27 (2002) 126–131.
[11] D.A. Bulik, G. Wei, H. Toyoda, A. Kinoshita-Toyoda, W.R. Waldrip, J.D. Esko, P.W. Robbins, S.B. Selleck, sqv-3, -7, and -8, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis, Proc. Natl. Acad. Sci. USA 96 (2000) 968–973.
[12] D. Zhao, D. Clyde, M. Bownes, Expression of fringe is down regulated by Gurken/Epidermal Growth Factor Receptor signalling and is required for the morphogenesis of ovarian follicle cells, J. Cell Sci. 113 (2000) 3781–3794.
[13] N. Zhang, G.V. Martin, M.W. Kelley, T. Gridley, A mutation in the Lunatic fringe gene suppresses the effects of a Jagged2 mutation on inner hair cell development in the cochlea, Curr. Biol. 10 (2000) 659–662.
[14] M. Asano, K. Furukawa, M. Kido, S. Matsumoto, Y. Umesaki, N. Kochibe, Y. Iwakura, Growth retardation and early death of β-1,4-galactosyltransferase knockout mice with augmented proliferation and abnormal differentiation of epithelial cells, EMBO J. 16 (1997) 1850–1857.
[15] T. Halttunen, A. Marttinen, I. Rantala, H. Kainulainen, M. Maki, Fibroblasts and transforming growth factor β induce organization and differentiation of T84 human epithelial cells, Gastroenterology 111 (1996) 1252–1262.
[16] K. Lindfors, T. Halttunen, P. Huotari, N. Nupponen, M. Vihinen, T. Visakorpi, M. Maki, H. Kainulainen, Identification of novel transcription factor-like gene from human intestinal cells, Biochem. Biophys. Res. Commun. 276 (2000) 660–666.
[17] C.A.R. Wiggins, S. Munro, Activity of the yeast MNN1 α-1,3-mannosyltransferase requires a motif conserved in many other families of glycosyltransferases, Proc. Natl. Acad. Sci. USA 95 (1998) 7945–7950.
[18] C. Breton, E. Bettler, D.H. Joziasse, R.A. Geremia, A. Imberty, Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases, J. Biochem. 123 (1998) 1000–1009.
[19] C. Breton, A. Imberty, Structure/function studies of glycosyltransferases, Curr. Opin. Struct. Biol. 9 (1999) 563–571.
[20] L.N. Gastinel, C. Cambillau, Y. Bourne, Crystal structures of the bovine β4 galactosyltransferase catalytic domain and its complex with uridine diphosphogalactose, EMBO J. 18 (1999) 3546–3557.
[21] B. Ramakrishnan, P.K. Qasba, Crystal structure of lactose synthase reveals a large conformational change in its catalytic component, the β1,4-galactosyltransferase-1, J. Mol. Biol. 310 (2001) 205–218.
[22] U.M. Unligil, J.M. Rini, Glycosyltransferase structure and mechanism, Curr. Opin. Struct. Biol. 10 (2000) 510–517.
[23] U.M. Unligil, S. Zhou, S. Yuwaraj, M. Sarkar, H. Schachter, J.M. Rini, X-ray crystal structure of rabbit *N*-acetylglucosaminyltransferase I: catalytic mechanism and a new protein superfamily, EMBO J. 19 (2000) 5269–5280.
[24] S. Kitazume-Kawaguchi, N. Dohmae, K. Takio, S. Tsuji, K.J. Colley, The relationship between ST6Gal I Golgi retention and its cleavage-secretion, Glycobiology 9 (1999) 1397–1406.
[25] P. Legaigneur, C. Breton, A. El Battari, J.-C. Guillemot, C. Auge, M. Malissard, E.G. Berger, C. Ronin, Exploring the substrate recognition of the human β-galactoside α2,6-sialyltransferase, J. Biol. Chem. 276 (2001) 21608–21617.
[26] K.J. Colley, Golgi localization of glycosyltransferases: more questions than answers, Glycobiology 7 (1997) 1–13.
[27] T. Ju, K. Brewer, A. D'Souza, R.D. Cummings, W.M. Canfield, Cloning and expression of human core 1 β1,3-galactosyltransferase, J. Biol. Chem. 277 (2002) 178–186.
[28] D.C. Tessier, D. Dignard, A. Zapun, A. Radominska-Pandya, A.J. Parodi, J.J. Bergeron, D.Y. Thomas, Cloning and characterization of mammalian UDP-glucose glycoprotein: glucosyltransferase and the development of a specific substrate for this enzyme, Glycobiology 10 (2000) 403–412.
[29] Y. Wang, L. Shao, S. Shi, R.J. Harris, M.W. Spellman, P. Stanley, R.S. Haltiwanger, Modification of epidermal growth factor-like repeats with O-fucose. Molecular cloning and expression of a novel GDP-fucose protein O-fucosyltransferase, J. Biol. Chem. 276 (2001) 40338–40345.
[30] F.K. Hagen, K. Nehrke, cDNA cloning and expression of a family of UDP-*N*-acetyl-D-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase sequence homologs from *Caenorhabditis elegans*, J. Biol. Chem. 273 (1998) 8268–8277.
[31] S.T. Smale, Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes, Biochim. Biophys. Acta 1351 (1997) 73–88.
[32] L. Weis, D. Reinberg, Accurate positioning of RNA Polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins, Mol. Cell. Biol. 17 (1997) 2973–2984.
[33] T.A. Ince, K.W. Scotto, A conserved downstream element defines a new class of RNA polymerase II promoters, J. Biol. Chem. 270 (1995) 30249–30252.
[34] A. Moustakas, K. Pardali, A. Gaal, C.-H. Heldin, Mechanisms of TGF-β signaling in regulation of cell growth and differentiation, Immunol. Lett. 82 (2002) 85–91.
[35] P. ten Dijke, M.-J. Goumans, F. Itoh, S. Itoh, Regulation of cell proliferation by Smad proteins, J. Cell. Physiol. 191 (2002) 1–16.
[36] Y. Zhang, X.-H. Feng, R. Derynck, Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-β-induced transcription, Nature 394 (1998) 909–913.