

Trefoil Factor Family Domains Represent Highly Efficient Conformational Determinants for *N*-Linked *N,N'*-di-*N*-acetylactosediamine (LacdiNAc) Synthesis*

Received for publication, July 14, 2014, and in revised form, September 9, 2014. Published, JBC Papers in Press, September 10, 2014, DOI 10.1074/jbc.M114.596049

David Bonar[‡] and Franz-Georg Hanisch^{‡§1}

From the [‡]Institute of Biochemistry II, Medical Faculty, University of Cologne, Joseph-Stelzmann-Strasse 52, 50931 Köln, Germany and [§]Center for Molecular Medicine Cologne, University of Cologne, Robert-Koch-Strasse 21, 50931 Köln, Germany

Background: Absent microheterogeneity of LacdiNAc *N*-glycan on human gastric TFF2 points to high stringency control mechanism.

Results: Single intact TFF domains of TFF2 control the β 4-GalNAc transfer to terminal GlcNAc residues as conformational determinants.

Conclusion: The role of a hydrophobic patch is hypothesized to form the essential part of the determinant.

Significance: The restricted expression of LacdiNAc on extracellular matrix proteins relates to important biological processes.

The disaccharide *N,N'*-di-*N*-acetylactose diamine (LacdiNAc, GalNAc β 1–4GlcNAc β) is found in a limited number of extracellular matrix glycoproteins and neuropeptide hormones indicating a protein-specific transfer of GalNAc by the glycosyltransferases β 4GalNAc-T3/T4. Whereas previous studies have revealed evidence for peptide determinants as controlling elements in LacdiNAc biosynthesis, we report here on an entirely independent conformational control of GalNAc transfer by single TFF (Trefoil factor) domains as high stringency determinants. Human TFF2 was recombinantly expressed in HEK-293 cells as a wild type full-length probe (TFF2-FL, containing TFF domains P1 and P2), as single P1 or P2 domain probes, as a series of Cys/Gly mutant forms with aberrant domain structures, and as a double point-mutated probe (T68Q/F59Q) lacking aromatic residues within a hydrophobic patch. The *N*-glycosylation probes were analyzed by mass spectrometry for their glycoprofiles. In agreement with natural gastric TFF2, the recombinant full-length and single domain probes expressed nearly exclusively fucosylated LacdiNAc on bi-antennary complex-type chains indicating that a single TFF domain was sufficient to induce transfer of this modification. Contrasting to this, the Cys/Gly mutants showed strongly reduced LacdiNAc levels and instead preponderant LacNAc expression. The probe with point mutations of two highly conserved aromatic residues in loop 3 (T68Q/F59Q) revealed that these are essential determinant components, as the probe lacked LacdiNAc expression. The structural features of the LacdiNAc-inducing determinant on human TFF2 are discussed on the basis of crystal structures of porcine TFF2, and a series of extracellular matrix-related LacdiNAc-positive glycoproteins detected as novel candidate proteins in the secretome of HEK-293 cells.

The initiation of glycosylation is generally determined by structural features in proximity to the glycosylation site; however, evidence has been provided for the existence of additional control mechanisms on the initiation level. A cis-located peptide stretch together with a specific amino acid pattern near *O*-glycosylation sites within the mucin domain were shown to be necessary for one of the rare types of protein-specific *O*-glycosylation, *i.e.* mammalian *O*-mannosylation (1). A similar type of double control was shown to be exerted by the peptide core on peripheral modifications with LacdiNAc of *N*-linked chains (2). Both observations have in common that a peptide stretch located distant from the putative glycosylation site represents a necessary prerequisite for the protein-specific glycosylation. These cis-controlling peptide elements provide another level of protein specificity to the glycosylation event. Other examples of protein sequence- or signal patch-controlled peripheral glycan modifications are found in the modification of *N*-linked chains with mannose 6-phosphate, the modification of neural cell adhesion molecule (NCAM) with polysialic acid, and the extension of *O*-fucose with GlcNAc by Fringe (3).

The peripheral modification of *N*-linked glycans with LacdiNAc² represents a unique protein-specific glycosylation found on a restricted number of glycoproteins and appears to play crucial roles in the regulation of circulatory half-lives of hormones (4) and in cell recognition (5). The modification of *N*-linked glycans was detected in a series of mammalian glycoproteins including the glycoprotein hormone LH (luteinizing hormone) (6), glycodein (7), prolactin-like proteins (8), proopiomelanocortin (9), SorLA/LR11 (10), sialoadhesin (11), tenascin-R (12), and carbonic anhydrase VI (13). On glycodein the LacdiNAc modification was claimed to have contraceptive

* This work was supported by Deutsche Forschungsgemeinschaft Grant HA2092/21-1 (to F.-G. H.).

¹ To whom correspondence should be addressed: Institute of Biochemistry II, Medical Faculty, University of Cologne, Joseph-Stelzmann-Str. 52, 50931 Köln, Germany. Tel.: 49-221-478-4493; Fax: 49-221-478-7788; E-mail: franz.hanisch@uni-koeln.de.

² The abbreviations used are: LacdiNAc, *N,N'*-di-*N*-acetylactosediamine, GalNAc β 1–4GlcNAc β ; TFF2, Trefoil factor 2; hTFF2, human TFF2; FL, full-length; WFA, *W. floribunda*; DC-SIGN, dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin; ECM, extracellular matrix; GM3, NeuAc α 2,3Gal β 1,4Glc β 1,1-ceramide; GM2, GalNAc β 1,4(NeuAc α 2,3)Gal β 1,4Glc β 1,1-ceramide; FSH, follicle-stimulating hormone; LSH, luteinizing hormone; TSH, thyroid stimulating hormone; LDN, LacdiNAc or *N,N'*-di-*N*-acetylactosediamine.

TFF Domain Determines LacdiNAc Formation

functions. As a peripheral modification of *O*-linked chains, LacdiNAc modifications were found on murine zona pellucida glycoprotein ZP3 (14), POMC (15), and on the extracellular matrix glycoproteins nidogen-1, extracellular matrix protein 1, AMACO, α -dystroglycan, and neurofascin (16). The LacdiNAc dihexosamine on *N*- and *O*-linked chains has been shown previously to be the substrate for further modification by fucosylation, sialylation, and sulfation and, as a specific feature of *O*-linked chains, by phosphorylation of the subterminal GlcNAc residue (16).

Two isoforms of the β 4-GalNAc-transferases, T3 and T4 (β 4GalNAc-T3, β 4GalNAc-T4), were described to be responsible for protein-specific LacdiNAc synthesis (17, 18). The group of Baenziger and co-workers (2) provided evidence that for induction of β 4GalNAc-T3 and -T4 activities, a 19-meric peptide sequence within the target protein (LRRFIEQKITKRK-KEKYWP) is necessary and sufficient. This determining cis-located peptide is characterized by a high content of basic amino acids and an α -helical structure. It is located at the C terminus of carbonic anhydrase VI and could induce the modification of *N*-linked chains on a normally unmodified protein (transferrin) after recombinant translocation of the peptide. We recently obtained evidence that the human Trefoil factor 2 (hTFF2 or TFF2) expressed in gastric mucosal secretions carries exclusively fucosylated LacdiNAc antennae on a biantennary *N*-linked chain located at Asn-38 in loop 1 of the N-terminal P1 TFF domain (19). No obvious microheterogeneity of *N*-glycosylation became apparent at this site except for a minor under-fucosylation of the LacdiNAc moieties pointing to a LacdiNAc inducing determinant with high stringency. Strikingly, the amino acid sequences of TFF2 downstream or upstream of the *N*-glycosylation site lack any similarities on the level of primary and secondary structures with the previously published LacdiNAc synthesis-inducing determinant.

Based on this observation we hypothesized that a conformationally stabilized determinant, forming a structural part of TFF domains, may have LacdiNAc-inducing capacities. To prove this assumption we expressed TFF2 as a recombinant full-length probe in HEK-293 cells. HEK-293 cells represent a well established cellular model in the context of LacdiNAc biosynthesis (20) as they express both enzymes, β 4GalNAc-T3 and -T4, involved in LacdiNAc synthesis (17, 18). These two enzymes do not compete with other β 4GalNAc-Ts, the isoenzymes T1 and T2, which are involved in ganglioside biosynthesis by converting GM2 to GD2 (β 4GalNAc-T1) or in the Sda/Cad blood-group synthesis (β 4GalNAc-T2). However, a competition with β 4Gal-T1 (lactose synthase), the enzyme involved in LacNAc synthesis, has certainly to be expected as this enzyme uses the same substrates.

Besides a full-length wild type probe (hTFF2-Fl), we generated a series of truncation probes in HEK-293 cells corresponding to the N-terminal TFF domain P1 (N-P1, carrying the natural *N*-glycosylation site) or to the C-terminal TFF domain P2 (C-P2*, carrying a designed new *N*-glycosylation site). Moreover, we generated and expressed the Cys/Gly mutant variants with defective TFF domains TFF2-Fl-(C52G) and TFF2-Fl-(C42G) and a double point mutant lacking aromatic residues within a putative hydrophobic patch N-P1-(W68Q/F59Q). Based on these glycosylation probes and mass spectrometric

analyses of their *N/O*-glycoprofiles, we could provide evidence that the LacdiNAc-inducing capacity is actually confined to conformationally intact TFF domains and that this capacity is strictly dependent on highly conserved aromatic residues in loop 3 of the TFF domains, which form essential components of a hydrophobic cleft with putative protein binding capacity.

EXPERIMENTAL PROCEDURES

Materials—The applied chemicals were generally purchased from Sigma. The quality was for analysis. Deviations are indicated. DNA, primers, and proteins were handled at +4 °C or on ice, and long time storage was at –20 °C or –80 °C.

Generation of Human TFF2 Glycosylation Probes—The template for the generation of TFF2 constructs was a plasmid containing the cDNA encoding the 129-amino acid comprising human TFF2 precursor protein (GenBankTM accession number BC032820.1). The full-length recombinant construct (TFF2-Fl, p24- p129), the N-terminal TFF domain P1 (N-P1, p24-p78), and the C-terminal TFF domain P2 (C-P2, p74-p129) were generated as secretory fusion proteins without their endogenous N-terminal signal peptide (p1-p23) using the following primers (BioTeZ, Berlin, Germany): full-length and N-P1, C-P2 constructs: Fwd1, AATTCTAGAAGAGAAACCCTCCCC-TGCCAG; Fwd2, AATTCTAGAACCAAAGCAAGAGTCG-GATCAG; Rev1, TGTAGATCTGTAATGGCAGTCTTCCACAGA; Rev2, TGTAGATCTCGACTCTTGCTTTGGGAG-GGG. PCR amplification was conducted by utilizing overhang primers with a non-complimentary 3'-extension including recognition sequences for the restriction enzymes XbaI and BglII. The digested PCR products were cloned into the eukaryotic pCEP-Pu vector V59 designed for the expression of C-terminally tagged (octa-His and STREP) secretory fusion proteins. Site-directed mutations were introduced with the QuikChange II site-directed mutagenesis kit. Cys-42 and Cys-52, respectively, were exchanged for Gly in TFF2-Fl, and a novel *N*-glycosylation site was generated in the otherwise *N*-glycosylation-negative P2 domain by mutation of Asp-87 to Asn-87 and Arg-89 to Thr-89 in one step using the following primers: TFF2-Fl-(C42/G), Fwd CATAACAGGACGAACGGCGGCT-TCCCT; TFF2-Fl-(C52G), Fwd ATCACCAGTGACCAGGTG-TTTGACAAT; C-P2*-(D87N/R89T), Fwd TGCATCATGGAG-GTCTCAAACCGAACAACTGTGGCTACC. The construct N-P1-(W68Q/F59Q) was synthesized by GeneArt-Life Technologies (Invitrogen, Regensburg, Germany) and ligated into V59 vector as above. A schematic view of the generated glycosylation probes is given in Fig. 1.

Cells and Cell Culture—Culture media and cell culture flasks were purchased from Biochrom (Berlin, Germany). Human embryonic kidney cell line HEK-293 (Invitrogen) was cultivated at 37 °C and 5% CO₂, humidified in Dulbecco's minimal essential medium (DMEM), and supplemented with 5% fetal calf serum (FCS), 100 units/ml penicillin, and 100 μ g/ml streptomycin. Lipofectamine 2000 Reagent (Invitrogen) was utilized for all transfection procedures. Selection of transfected cells was established with 5 μ g/ml puromycin (Sigma) as an additional culture media supplement. At a confluence of ~80% cells were triple rinsed with PBS and kept in culture for 3 days in medium lacking

FCS. Supernatants were collected and stored at -20°C until purification of recombinant protein via affinity tag.

Preparation of Proteins from Cell Culture Supernatants—500 ml of supernatants were centrifuged for 15 min at 10000 rpm, $+4^{\circ}\text{C}$. 500 mM NaCl, 20 mM Na_2HPO_4 , 10 mM imidazole, 25 μl of PMSF (1 M in DMSO) were added. The pH was adjusted to 8.0 at $+4^{\circ}\text{C}$ with HCl. Supernatant was then passed through a cellulose filter and applied to the nickel-nitrilotriacetic acid column twice in a row. For detailed description of the procedure refer to Breloy *et al.* (16). Eluted fractions containing the recombinant protein were pooled then concentrated and desalted via Amicon centrifugation units, molecular mass cut-off 10 kDa (Millipore, Darmstadt, Germany).

Gel Electrophoresis and Western Blotting—Aliquots of the recombinantly expressed fusion proteins were treated before Western blot analysis by thrombin digestion to remove the strep tag. Proteins in 5–20% or 15% SDS gels were either stained with silver or blotted onto nitrocellulose (Protran BA 83, Schleicher & Schüll) in a wet blot transfer cell (Bio-Rad) for antibody detection or onto PVDF membranes in a semidry blot chamber (both from Bio-Rad) for lectin detection (16). Fusion proteins were detected with anti-strepII mouse IgG (IBA, Göttingen, Germany). Modification with LacdiNAc was tested either with the monoclonal mouse antibody 273–372 (kindly provided by Dr. C. H. Hokke, Leiden University Medical Clinic, Leiden, The Netherlands; 1:5 diluted culture supernatant, 2 h at ambient temperature) or with biotinylated *Wisteria floribunda* lectin (Sigma, 10 $\mu\text{g}/\text{ml}$, 2 h at ambient temperature). Horse radish peroxidase (HRP)-conjugated rabbit anti-mouse IgG (DAKO, Hamburg, Germany) was applied as a secondary antibody or Strep-tactin-HRP (IBA) for staining of lectin binding. Immunolabeled protein was detected by enhanced chemiluminescence (Roche Applied Science). An aliquot of the full-length probe hTFF2-FI was pretreated with 0.1 M aqueous TFA at 80°C for 1 h to chemically desialinate and partially defucosylate the glycoprotein before *W. floribunda* (WFA) staining.

Proteomics of Wild Type and Mutant Probes—Intact probes (1 μg of protein) were digested with chymotrypsin in 50 mM ammonium bicarbonate, pH 8.0, at 37°C for 4 h using an enzyme to substrate ratio of 1:50 (w/w). After heat denaturation of the protease (10 min at 90°C), the resulting peptide mixture was further digested with trypsin (both enzymes were sequencing grade from Promega (Mannheim, Germany)) in the same buffer at 37°C overnight using a 1:50 enzyme substrate ratio (w/w). After reduction of disulfides and carbamidomethylation of cysteine residues, peptides were analyzed by LC-electrospray ionization mass spectrometry as described (22).

Enzymatic Cleavage and Methylation of N-Glycans—Dried protein was dissolved in 100 μl of 50 mM NH_4HCO_3 and submitted to tryptic digestion at 37°C for ~ 16 h. Trypsin was inactivated at 95°C for 5 min. Digested sample was dried by vacuum centrifugation and solubilized in 50 μl of 50 mM NH_4HCO_3 . An aliquot of 0.5 μl of peptide *N*-glycosidase F (New England Biolabs, Frankfurt, Germany) was added, and *N*-deglycosylation was performed at 37°C for at least 16 h. Liberated *N*-glycans were separated from peptides by solid-phase extraction on C18 cartridges (Agilent Technologies, Waldbronn, Germany). *N*-Glycans in the flow-through were dried by vacuum rotation

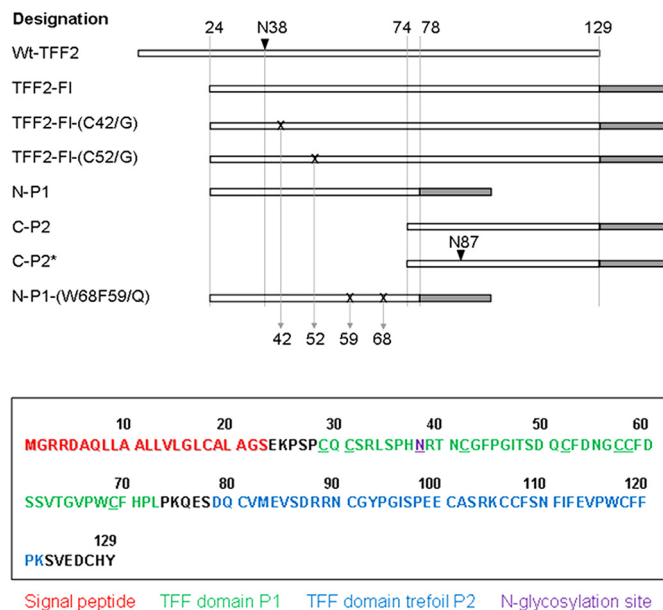


FIGURE 1. TFF2 full-length, truncation, and point-mutated constructs generated in this study. Upper panel, schematic linear views of the various recombinant probes generated and analyzed in this study. Sites of mutation are indicated by X, sites of *N*-glycosylation by arrows, tag regions by gray-shaded areas. Lower panel, the sequence of the full-length probe TFF-FI is shown, and the colors used refer to the BM40 signal peptide (red), to the N-terminal P1 TFF domain (green), and to the C-terminal P2 domain (blue). The sequence of the oligo-histidine and strep tags is RSLVPRGSHHHHHHH SAWSHPQFEK. The *N*-glycosylation sites at Asn-38 (and at Asn-87 in C-P2*) is shown underlined.

and in a desiccator for 1 h in the presence of $\text{P}_2\text{O}_5/\text{KOH}$. Methylation was performed in 100 μl of water-free DMSO containing finely dispersed NaOH (30 min at 22°C) followed by the addition of 50 μl of methyl iodide (30 min at 22°C). After methylation, 300 μl of chloroform was added, and the sample was repeatedly extracted with 200 μl of water. The chloroform phase was dried under N_2 , and methylated glycans were solved in 20 μl of methanol.

β -Elimination of O-Glycans and Permethylation of Glycan Alditols—The glycan chains were released from the protein by reductive β -elimination. For this purpose the glycoproteins were incubated with 1.0 M NaBH_4 in 50 mM NaOH for 18 h at 50°C . The reaction was stopped by adding 2 μl of acetic acid. Salt was removed with 100 μl of Dowex 50WX8 aqueous suspension (Bio-Rad) in a batch procedure. Excessive borate was co-distilled as methyl ester in a stream of nitrogen by adding several 0.1-ml aliquots of 1% acetic acid in methanol. Permethylation of the glycan chains was performed as described above and in Breloy *et al.* (16).

Matrix-assisted Laser Desorption Ionization (MALDI)-TOF Mass Spectrometry—MALDI mass spectrometry was performed on an UltrafleXtreme instrument (Bruker Daltonics). The permethylated glycans (~ 500 ng) contained in methanol were applied to the stainless steel target by mixing a 0.5- μl aliquot of sample with 1.0 μl of matrix (saturated solution of 2,5-dihydroxy benzoic acid in acetonitrile, 0.1% TFA, 1:2). Alternatively, 0.75 μl of sample were mixed with 0.75 μl of α -cyano-4-hydroxycinnamic acid (Bruker, Bremen, Germany) matrix solution on the MALDI target and subjected to MS analysis as described (16, 21). Peptide samples were analyzed after co-crystallization of aqueous solutions in 0.1% TFA, 50% ace-

TFF Domain Determines LacI_NAc Formation

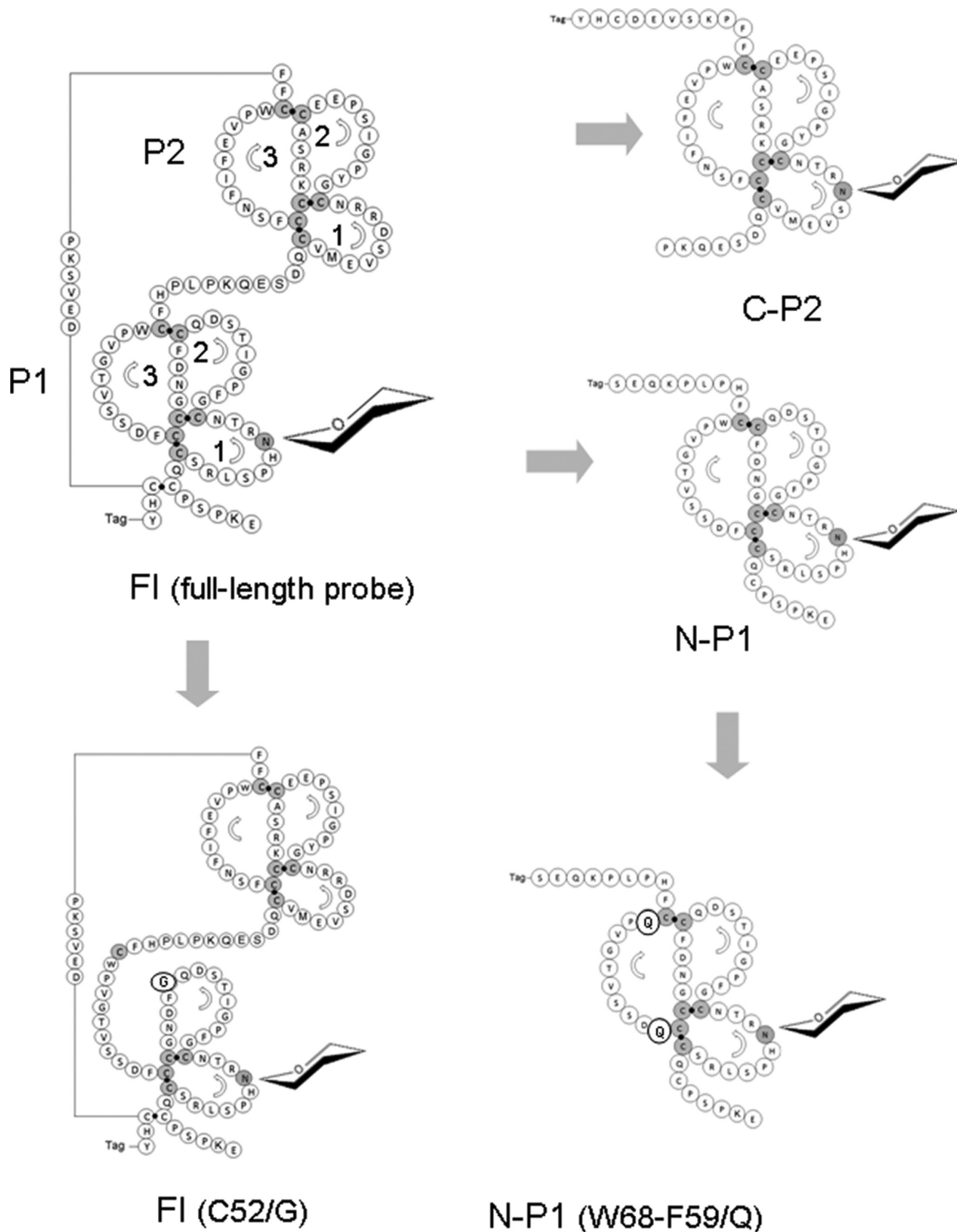


FIGURE 2. Schematic views of recombinant human TFF2 probes expressed in HEK-293 cells. Structural models of TFF2-FI, N-P1, C-P2, and mutant C52G and T68Q/F59Q probes of TFF2 showing amino acid sequences and cysteine disulfide bridges responsible for formation of the three loops in TFF domains P1 and P2. Reading of amino acid sequences is indicated by *arrows*, and the three loops are numbered. Mutated amino acids are *highlighted*. According to proteomic data, the Cys-mutated probes are characterized by aberrant disulfide bridge configurations.

tonitrile and a saturated matrix solution of α -cyano-4-hydroxycinnamic acid in the same solvent. Analyses were performed by positive ion detection in the reflectron mode. Ionization of co-crystallized analytes was induced with a pulsed Smart-beam laser (accumulation of ~ 5000 shots), and the ions were accelerated in a field of 20 kV and reflected at 23 kV. MALDI-MS/MS was performed by analysis of post-source decay fragments in the laser-induced dissociation mode. Fragment annotation of glycans was assisted by application of the Glyco-Workbench tool (23).

Analysis of (Glyco)peptides—(Glyco)peptide analysis of the recombinant proteins was performed by LC-MS/MS of tryptic peptides on an electrospray ionization iontrap, the HCT ultra ETDII PTMDiscovery-System (Bruker-Daltonics) coupled with an online easy-nano-LC system (Proxeon, Odense, Denmark). The sample was separated on an analytical C18 column ($75 \mu\text{m} \times 10 \text{cm}$) using gradient runs from 0 to 35% acetonitrile in 0.1% TFA during 30 min. Ions were scanned with 8100 atomic mass units/s in a range from m/z 300 to 2500 in MS mode and m/z 200 to 3000 in MS/MS mode. MS/MS spectra were generated by collision-induced dissociation fragmentation (16).

Affinity Chromatography on WFA of Glycoproteins in the Secretome of HEK-293 Cells—Culture supernatants of HEK-293 cells transfected with *hZP3* (p29–176) as an internal LacdiNAc-positive standard (16) were collected after 3 days. A volume of 200 ml of supernatant was ultracentrifuged for 30 min at 10,000 rpm (4°C) and dialyzed overnight at 4°C against water. After drying by vacuum rotation, the sample was resolubilized in 500 μl of Tris-buffered saline, pH 7.4, with 100 mM glucose (TBS-Glc) and applied onto a 1-ml column of *W. floribundagarose* (Vector Laboratories, Burlingame, CA) equilibrated with at least 10 ml of the same buffer. After overnight column circulation of the sample at 4°C , the column was washed with 10 column volumes of TBS-Glc at 7 ml/h, and bound proteins were eluted consecutively with 10 ml of TBS containing 100 mM lactose or 25 mM GalNAc. Each fraction was dialyzed and concentrated by centrifugation in Amicon Ultra tubes (Millipore) before mass spectrometric identification of proteins.

RESULTS

Generation and Expression of Full-length, Truncation, and Point-mutated *hTFF2* Probes—The full-length construct (TFF2-FI) generated in this study comprises 106 genuine TFF2 amino acid residues, which are extended N-terminally by a 23-meric BM40 signal peptide and C-terminally by an oligo-histidine/strep-tag peptide (Figs. 1 and 2). Derived from this wild type probe, we generated truncated constructs corresponding either to the N-terminal P1 TFF domain (N-P1, p24–78) or to the C-terminal P2 domain (C-P2, p74–129). Mutation of Asp-87 to Asn and of Arg-89 to Thr in C-P2 resulted in an artificial *N*-glycosylation site in the C-P2 probe (refer to C-P2*). Mutant probes with aberrant disulfide structures in the P1 domain were generated by C52G mutation (expectedly affecting loop 3 formation) and by C42G mutation presumed to result in the formation of a superloop comprising loops 1 and 2. Peptides generated by sequential chymotrypsin and trypsin digestion of recombinant probes (followed by cystine reduction and cysteine alkylation) revealed evidence for at least partial changes of

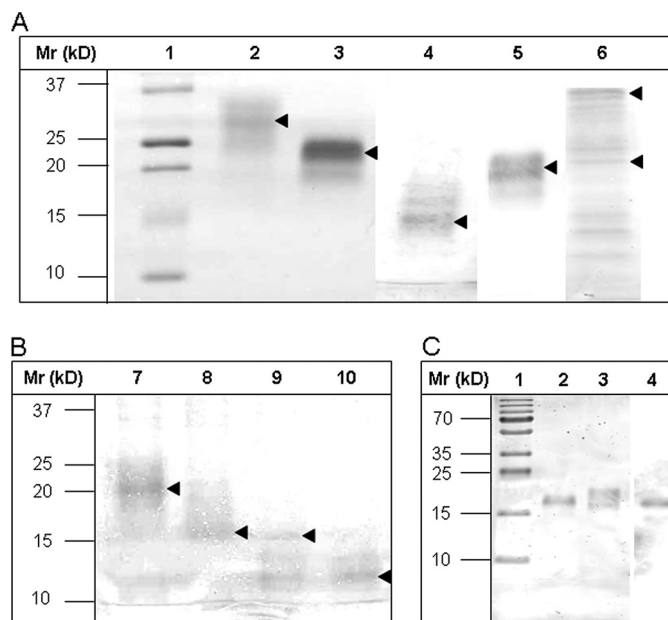


FIGURE 3. Gel electrophoretic separation of TFF2 probes before and after thrombin cleavage of strep-his-tags and Western blot of hTFF2-FI with WFA after desialination/defucosylation. A and B, electrophoretic protein separation was performed in 5–20% gradient gels. A, Lanes 1–6 were loaded with samples: 1, molecular mass standard; 2, hTFF2-FI; 3, N-P1; 4, C-P2; 5, C-P2*; 6, TFF2-FI-(C52G). B, lanes were loaded with native or thrombin-treated probes N-P1 (lanes 7 and 8) or C-P2 (lanes 9 and 10). Proteins in A and B, lanes 2–10, were run in the native, non-reduced state. Arrows mark the positions of recombinant tagged or tag-free proteins. C, hTFF2-FI probe was reduced and electrophoretically separated in a 15% polyacrylamide gel and either silver-stained or blotted onto PVDF membrane before staining with biotinylated lectin (WFA) and Strep-tactin-HRP. Lane 1, molecular mass standard; lane 2, chemically desialinated hTFF2-FI (treatment with 0.1 M aqueous TFA at 80°C for 1 h results also in partial defucosylation); lane 3, untreated hTFF2-FI; lane 4, Western blot of hTFF2 with WFA. Lanes 1–3 show silver-stained proteins.

disulfide bridge configurations in the C52G-mutated probe TFF2-FI-(C52G) compared with the wild type full-length probe TFF2-FI by distinct proteolytic cleavage patterns (data not shown). Striking differences in the peptide patterns are apparent particularly for those TFF2 peptide stretches that link the two domains P1 and P2.

The expression and purity of the probes was proven after affinity isolation via their oligo-histidine tags by mass spectrometric proteomics (data not shown) and by gel electrophoresis (Fig. 3). The tagged probes revealed shifts in their apparent masses when analyzed before and after thrombin cleavage of the tags (Fig. 3).

Immunochemical Evidence for LacdiNAc Expression on *hTFF2* Probes—Natural glycoforms of TFF2 from human stomach had been tested negatively for LacdiNAc expression when using mouse monoclonal antibody 273–372 and shown to be weakly stained with the lectin from WFA (17). Both findings can be explained by the nearly complete substitution of the *N*-linked LacdiNAc termini on TFF2 with fucose $\alpha 3$ -linked to the subterminal GlcNAc. Similar results were obtained with the recombinant TFF2 probes generated in this study, *i.e.* no staining with antibody 273–372 and sporadically very weak staining with WFA (not shown). To prove expression of LacdiNAc on hTFF2-FI, we demasked the glycotope recognized by WFA by chemical desialination and partial defucosylation. As shown in Fig. 3, this treatment resulted in considerably increased stain-

TFF Domain Determines LacdiNAc Formation

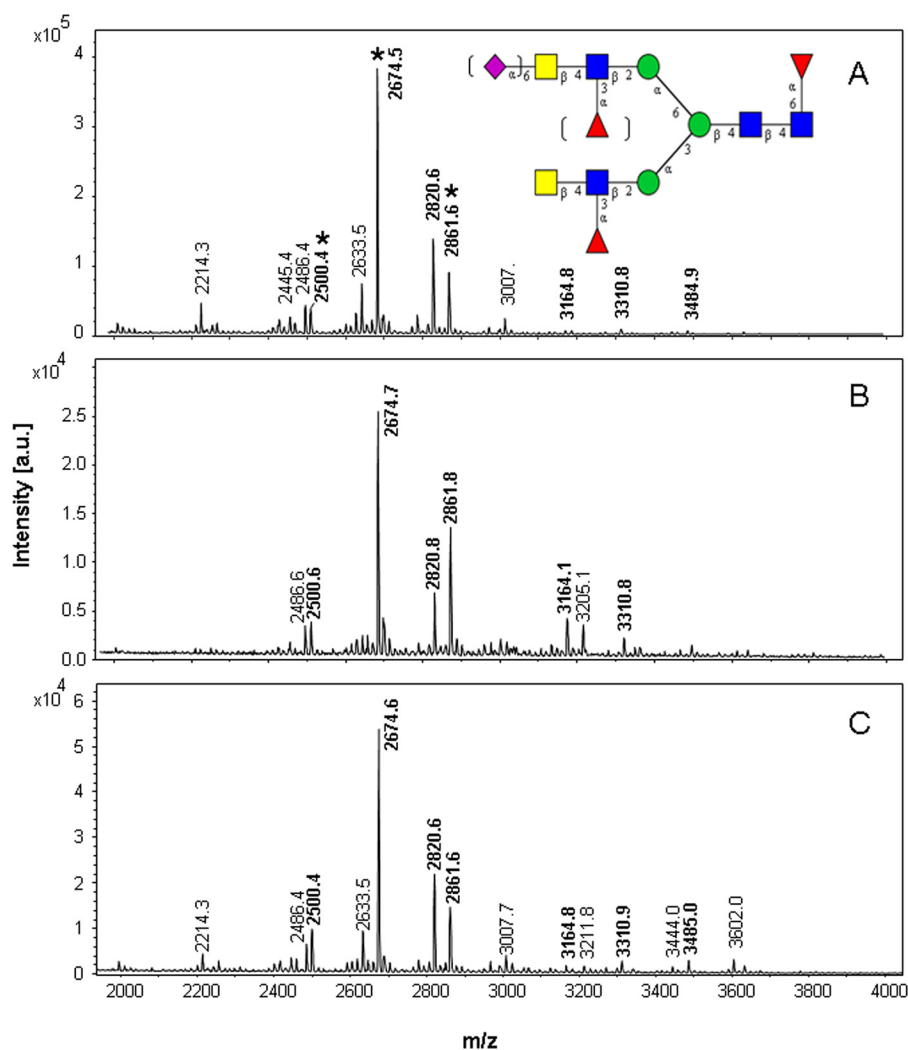


FIGURE 4. MALDI mass spectrometric data of *N*-linked glycans from recombinant TFF2 wild type and truncation probes. A, full-length TFF2 probe (TFF2-FI). B, N-terminal P1 domain probe (N-P1). C, C-terminal P2 domain probe with designed *N*-glycosylation site (D87N/R89T): C-P2*. The inserted glycan structure refers to signals marked by an asterisk. Molecular ions referring to LacdiNAc-positive glycans are written in *boldface*. The signals correspond to the sodium adducts of molecular ions ($M+Na$) of permethylated glycans. *a.u.*, arbitrary units.

ing of a defined protein band at 18 kDa (in agreement with the natural glycoform expressed in human stomach; see Ref. 19). Efficiency of desialination and partial defucosylation were checked by MALDI mass spectrometric analysis of enzymatically released and permethylated *N*-glycans (see below).

N-Glycoprofiling of TFF2 Truncation Probes: Single TFF Domains P1 and P2 Have the Capacity to Induce LacdiNAc Formation—Natural TFF2 (p24–129) consists of two symmetrically positioned TFF domains, the N-terminal glycosylated P1 and the C-terminal P2 domain. We wanted to address the question of whether a single domain exhibits the structural qualities of a determinant that induces LacdiNAc formation. For this purpose the full-length and two truncation probes were generated corresponding to P1 (p24–78) and P2 (p74–129), all exhibiting an *N*-glycosylation site in loop 1. Expression of the three constructs in HEK-293 cells and mass spectrometric analysis of the affinity-purified probes revealed *N*-glycoprofiles similar to the natural gastric protein (19), which was characterized by a core fucosylated bi-antennary *N*-glycan with monofucosylated LacdiNAc antennae (refer to the dominant signal at

m/z 2674.5 in Fig. 4 corresponding to the sodium adduct of the dodecasaccharide F3H3N6; see Table 1). The dominant glycan was accompanied by an under-fucosylated (F2H3N6 at m/z 2500.4) and by a monosialinated species (S1F2H3N6 at m/z 2861.6). After chemical desialination and partial defucosylation, the dominant signal found at m/z 2294.5 (data not shown) corresponds to the previous base peak at m/z 2674.5 (Fig. 4) that had lost two fucose residues in the antennae ($M+Na-32$) as revealed in MS2 by formation of B ions at m/z 260 and 505 (data not shown).

The dominant signals in the spectra shown in Fig. 4 were selected as precursor ions for analysis by post-source decay MALDI-MS/MS (Fig. 5) to elucidate the structural features of the antennae. We demonstrate the presence of antennary dihexosamines HexNAc₂ giving rise to the B2 ion at m/z 505 (derived from the precursor at m/z 2500.4, not shown), monofucosylated HexNAc(Fuc)HexNAc giving rise to formation of the B1 ion at m/z 260, and B2 ions at m/z 679 or 701, respectively, derived from the precursor at m/z 2674.5 (Fig. 5A).

Fragments of the precursor at m/z 2861.6 indicated the presence of monosialylated and monofucosylated antennae:

TABLE 1

N-Linked glycans on recombinant human TFF2 constructs expressed in HEK-293 cells

Glycans with masses in boldface were demonstrated by MS/MS to express terminal LacdiNAc, fucosylated LacdiNAc, or sialinated LacdiNAc. M+Na, M+Na-32, and M+Na-54 molecular ion species shown in Figs. 4 and 6 are listed as M+Na. Semiquantitative data from MALDI-MS of permethylated glycans are presented as ++ (major, base peak), + (>5% of base-peak intensity), (+), (minor species with <5% of base-peak intensity).

Exp. rel. mass M+Na (<i>m/z</i>)	Glycan	hTFF2 constructs					
		Fl	N-P1	C-P2*(D87N)	N-P1 (W68Q/F59Q)	FL-(C52G)	FL-(C42G) ^a
1526.1	H5N2	—	—	—	++	—	—
1730.2	H6N2	—	—	—	++	—	—
1934.3	H7N2	—	—	—	+	—	—
2138.5	H8N2	—	—	—	+	—	—
2342.6	H9N2	—	—	—	+	—	—
2214.3	F2H4N4	+	—	+	—	—	—
2430.8	S1H5N4	—	—	—	—	+	—
2459.4	F2H4N5	—	—	+	—	+	—
2500.4	F2H3N6	+	+	+	—	+	—
2633.5	F3H4N5	+	(+)	+	—	(+)	—
2674.5	F3H3N6	++	++	++	—	+	—
2791.9	S2H5N4	—	—	—	(+)	++	+
2820.6	S1F2H4N5	+	+	+	—	—	—
2861.6	S1F2H3N6	+	+	+	—	—	—
2994.6	S1F3H4N5	(+)	(+)	—	—	—	—
3164.8	F3H3N8	+	+	(+)	—	—	—
3242.0	S2H6N5	—	—	—	(+)	+	—
3310.8	S1F2H4N7	(+)	+	+	—	—	—
3417.1	S2F1H6N5	—	—	—	—	(+)	—
3484.9	S1F3H4N7	(+)	(+)	+	—	—	—
3602.1	S3H6N5	—	—	—	(+)	++	+
3866.2	S2F1H7N6	—	—	—	—	(+)	—
3963.2	S4H6N5	—	—	—	—	+	(+)

^a Only two major glycan species were detectable as sample amounts of FL-(C42G) were significantly lower than those of other TFF2 probes.

NeuAc-HexNAc-HexNAc (B1 ion at *m/z* 376, B3 at *m/z* 866 or 888) and HexNAc(Fuc)HexNAc (B1 ion at *m/z* 260, B2 ion at *m/z* 701 (Fig. 5B). A further major *N*-glycan species was detected at *m/z* 2820.6 and assigned as a biantennary *N*-glycan with one NeuAc-HexNAc (sialyl-LacNAc) and one HexNAc(Fuc)HexNAc antenna (Fig. 5C). Structural assignments of fragments with respect to glycan sequences of the most relevant signals are given in Fig. 5.

***N*-Glycoprofiling of TFF2 Cys/Gly-mutated Probes: TFF Domains Need to Be Conformationally Intact to Induce LacdiNAc Synthesis on Asn-38**—To define the structural requirements of the LacdiNAc inducing determinant, we generated Cys/Gly mutant probes with aberrant disulfide bridge patterns. In the full-length construct TFF2-FL-(C52G) Cys-52 (the third cysteine in the 1–5, 2–4, 3–6 disulfide bridge configuration of TFF domains) is mutated to Gly, which should prevent loop 3 formation in the P1 TFF domain and result in an aberrant disulfide bridge pattern (Fig. 2B). In TFF2-FL-(C42/G), where the second cysteine in the TFF domain-characteristic disulfide configuration is mutated, the missing cysteine fuses loops 1 and 2 to a superloop and should also provoke aberrant disulfide bridge formation (Fig. 2B). The *N*-glycoprofiling of these two mutant probes revealed a strikingly different pattern of major glycans that was dominated by core-fucosylated biantennary glycans with LacNAc antennae (Fig. 6B and Table 1). Only a minor portion (~12% of total glycans derived from TFF2-FL-(C52G) contained LacdiNAc antennae, which contrasts with the pattern of the wild type full-length (TFF2-FL) and the single TFF domain probes (N-P1, C-P2*), where on the average ~84% of the total glycans had been found to contain LacdiNAc antennae. The most abundant *N*-glycan species in these profiles were detected at *m/z* 2792 and 3602, respectively, corresponding to

the biantennary or triantennary *N*-glycans with NeuAc-HexNAc-HexNAc antennae and lacking core-fucosylation.

***N*-Glycoprofiling of N-P1-(W68Q/F59Q)-mutated Probe with an Affected Hydrophobic Patch**—To demonstrate the potential effects of highly conserved aromatic residues in loop 3 on the formation of *N*-linked LacdiNAc, we expressed a double point-mutated probe N-P1-(W68Q/F59Q) in HEK-293 cells and analyzed its *N*-glycoprofiles (Fig. 6C and Table 1). As hypothesized, the MS1 spectrum did not reveal any molecular ion signals that would correspond to the expected masses of LacdiNAc-positive *N*-glycans. Moreover, the *N*-glycoprofile of this probe showed dramatic shifts from complex-type glycans to high mannose type (M5-M9), indicating incomplete processing of the chains. The dominant signals were found at *m/z* 1526.1 (M5) and 1730.2 (M6) representing M-54 ions that arise by loss of sodium methylate. Compared with the profile obtained for the wild type single domain probe N-P1, the double point-mutated probe N-P1-(W68Q/F59Q) exhibits a totally aberrant *N*-glycoprofile, which supports the assumption that these highly conserved aromatic residues may play an essential role in *N*-glycan processing and followup glycosylation.

***O*-Glycoprofiling of TFF2 Probes**—Although no evidence for *O*-linked glycans on the natural human TFF2 was obtained, we detected considerable *O*-glycosylation of the recombinant probes TFF2-FL and N-P1 (Fig. 7). The *O*-glycoprofile was devoid of any LacdiNAc-containing oligosaccharide, as neither the core 2-based tetrasaccharide (*m/z* 1024) nor any of its fucosylated (*m/z* 1198) or sialylated derivatives (*m/z* 1385) were detectable. According to glycopeptide analyses performed by LC-MS of tryptic peptides, we could localize the *O*-glycans to the peptide SLVPR (calculated mass 571.4 Da), which is part of the C-terminal-located oligo-His and strep2-tag sequence.

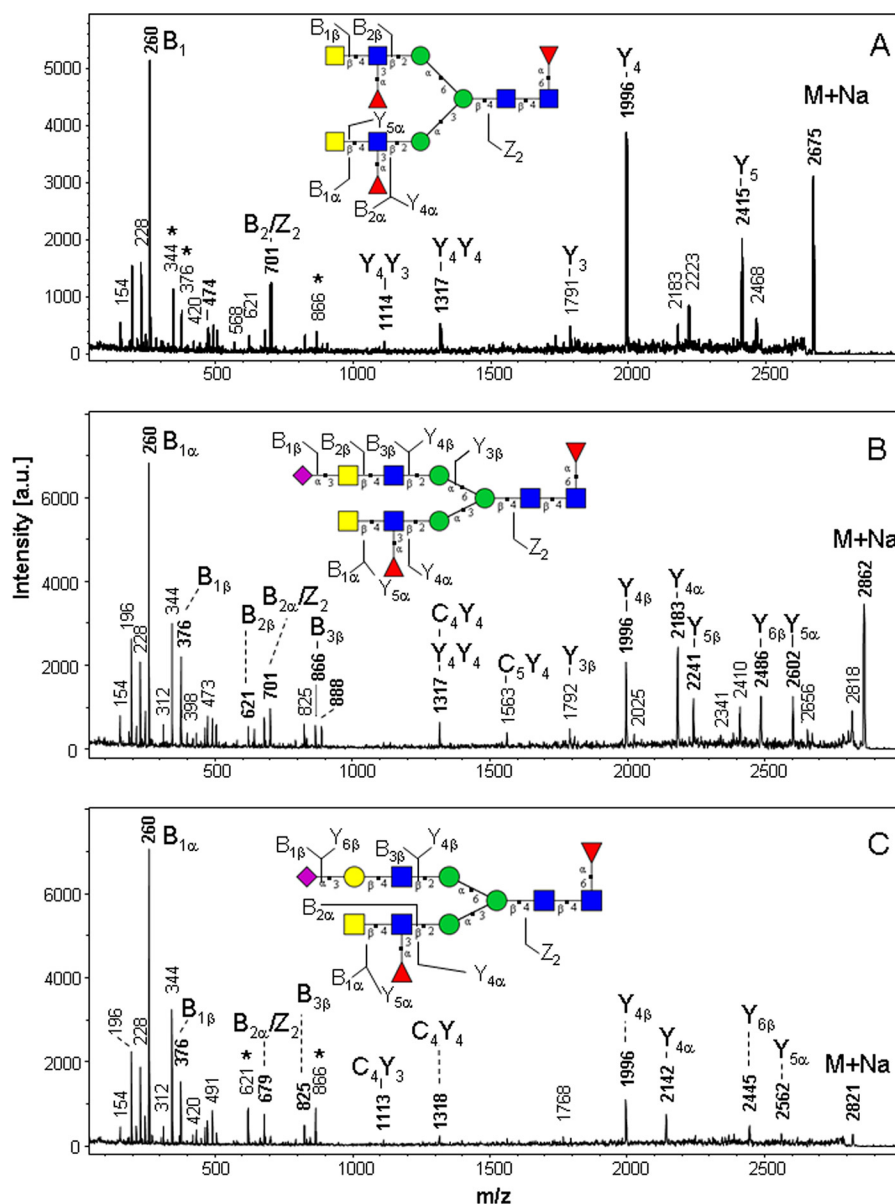


FIGURE 5. Post-source decay-MALDI mass spectrometric fragmentation analysis of permethylated *N*-glycans from recombinant TFF2 probes. A, N-P1-derived glycan at *m/z* 2675. B, N-P1-derived glycan at *m/z* 2862. C, N-P1-derived glycan at *m/z* 2821. Fragments were designated according to the nomenclature of Domon and Costello using the GlycoWorkbench annotation tool (23) and assigned to the inserted structural model. For the sake of clarity only a selection of single fragmentation ions (*B*, *Y*, *C*, *Z*) were assigned. Masses marked by an asterisk refer to fragment ions derived from interfering precursor ions. Note that fragment ions can be either proton or sodium adducts. *a.u.*, arbitrary units.

WFA Affinity Chromatographic Enrichment of LacdiNAc-positive Glycoproteins from the Secretome of Human HEK-293 Cells—To identify new candidate proteins forming substrates of the glycosyltransferases β 4GalNAc-T3 or -T4, we analyzed affinity-enriched fractions of glycoproteins from the secretome of HEK-293 cells that bound to immobilized WFA and elutable with GalNAc or lactose (Table 2). In replicate experiments we identified by mass spectrometric proteomics a series of 12 potentially LacdiNAc-positive secreted *N*- or *N,O*-glycoproteins mostly related to the extracellular matrix, whereas 8 glycoproteins were found in only one replicate. Among the former were galectin-3-binding protein, peroxidase, two nidogen isoforms, several laminin isoforms, the metalloproteinase TIMP1, fibulin-1, fibrillin-2, clusterin, cochlin, and collagen 6 α 1. For nidogen-1 we previously reported diHexNAc-positive *O*-glyco-

peptides as potentially LacdiNAc-expressing (16). For some of the proteins listed in Table 2 it cannot be ruled out that their binding to the affinity matrix was mediated by protein-protein interactions, as for example in the case of galectin-3-binding protein and its known laminin-binding interactors fibronectin, nidogen-1, and collagen 6.

DISCUSSION

Structural Features of TFF Domains That Could Underlie the Observed Protein Specificity—Summing up the known, structurally proven LacdiNAc-positive proteins (7 *O*-linked and 11 *N*-linked mammalian glycoproteins) and potentially LacdiNAc-positive proteins enriched by WFA-affinity chromatography from the secretome of HEK-293 cells (see above), the total

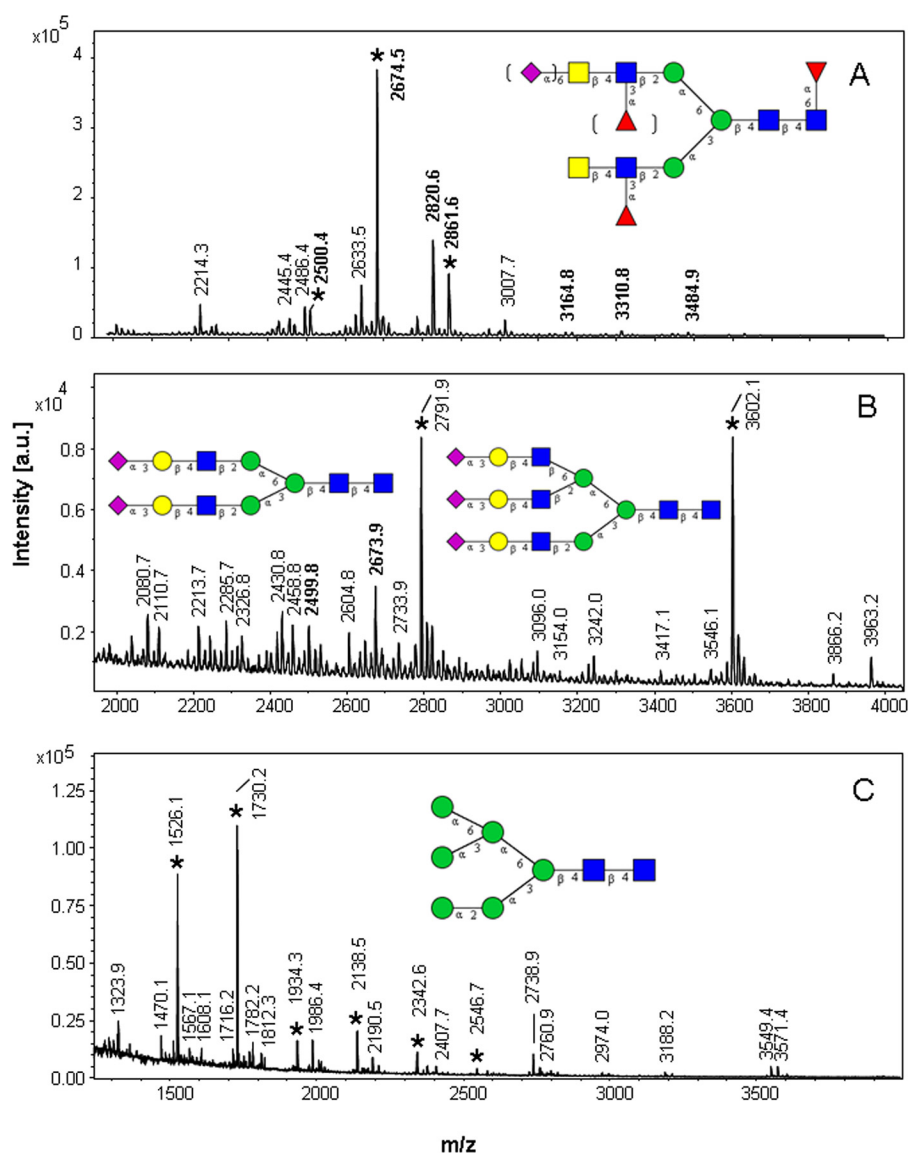


FIGURE 6. MALDI mass spectrometric data of *N*-linked glycans from recombinant TFF2 Cys/Gly- and T68Q/F59Q-mutated probes. *A*, full-length probe (TFF2-FI). *B*, TFF2-FI-(C52G). *C*, N-P1-(W68Q/F59Q). The inserted glycan structure refers to major signals marked by an asterisk that characterize the respective *N*-glycoprofile. Molecular ions referring to LacdiNAc-positive glycans are written in *boldface*. The signals correspond to the sodium adducts of molecular ions ($M+Na$) of permethylated glycans. *a.u.*, arbitrary units.

number of these may be estimated to range at ~ 40 glycoproteins. Although this is still a limited number, the question arises whether all these proteins may have a common structural element that could serve as a determinant with LacdiNAc-inducing capacity. Most of the known LacdiNAc-positive *N,O*-glycoproteins have neither a CA6-related cis-controlling peptide element in their primary structures nor do they contain TFF domains. These considerations raise the question of whether particular structural features of TFF domains may underlie the observed phenomenon. We, therefore, aligned the primary structures of human TFF2 with those of other mammalian TFF domain-containing proteins to reveal highly conserved amino acids (Fig. 8). Beyond the characteristic cysteine patterns we found highly conserved positions in loop 2 (corresponding to the hTFF2 positions Ile47 and Thr48) and in loop 3 (corresponding to the hTFF2 positions Phe-108 and Trp-117). In particular the latter two aromatic residues were claimed to stabi-

lize the fold of the domain and to play a role in protein binding (24).

Sequence alignments of the LacdiNAc-positive neuropeptide hormone β -subunits (FSH, LSH, TSH) with TFF2 revealed considerable similarities, in particular with respect to the cysteine positions relative to the *N*-glycosylation site (Fig. 9). The cysteines at -7 and $+4$ were conserved, whereas the cysteine at $+20$ in TFF2 was shifted to $+27$ (FSH, LSH) or $+29$ (TSH). In TFF2 domain P1 cysteines at Asn-38 -7 , $+4$, and $+20$ correspond to TFF domain cysteines I, II, and V, and the disulfide bridge of cysteines I and V contributes to formation of loop 1. The disulfide bridge pattern in the FSH β -subunit is strikingly different, as six disulfide bridges are formed (Cys-21–Cys-69, Cys-35–Cys-84, Cys-38–Cys-122, Cys-46–Cys-100, Cys-50–Cys-102, Cys-105–Cys-112), and no overlap with the disulfide bridge pattern of TFF2 becomes obvious, as Cys-35 in FSH β corresponds to cysteine I of the TFF2 domain and Cys-69

TFF Domain Determines LacdiNAc Formation

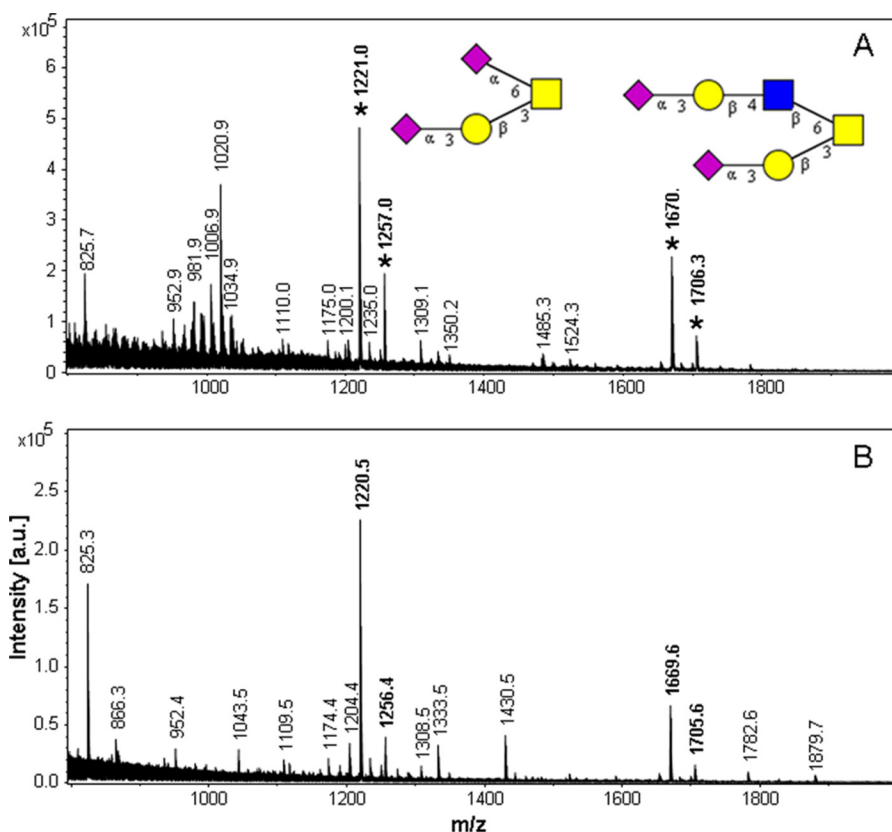


FIGURE 7. MALDI-MS data of O-glycan alditols from recombinant TFF2 probes. A, full-length probe (TFF2-FI). B, N-terminal P1 domain probe N-P1. The inserted glycan structures correspond to the signals marked by an asterisk (M+Na at m/z 1256: S2H1N1; at m/z 1706: S2H2N2). Signals at M-36 correspond to mono-undermethylated species were detected as proton adducts. *a.u.*, arbitrary units.

TABLE 2

WFA affinity chromatography of HEK-293 secretome

The list summarizes those proteins identified in fractions eluted with 0.1 M lactose and 0.1 M GalNAc. Proteins designated in boldface were found in both technical replicates. GO data refer to proteins forming ECM components or being functionally ECM-related (ECM), to secreted proteins (sec), and to glycosylated proteins (N, O).

Designation	ECM component or ECM-related	Secreted protein	Glycosylated (N- or O-linked)
LG3BP (galectin-3-binding protein)	ECM	sec	N
PXDN (peroxidase)	ECM	sec	N
C1R (C1 component)	serine protease	sec	N
SMOC1 (SPARC-related modular calcium binding protein 1)	ECM	sec	N
VTNC (vitronectin)	ECM	sec	N,O
NID1 (nidogen-1)	ECM	sec	N, O
NID2 (nidogen-2)	ECM	sec	N, O
LAMA1 (laminin- α 1)	ECM	sec	N
LAMA4 (laminin- β 4)	ECM	sec	N
LAMAB1 (laminin- β 1)	ECM	sec	N
LAMAC1 (laminin- γ 1)	ECM	sec	N
TIMP1 (tissue metalloproteinase 1)	ECM	sec	N
NPTX1 (neuronal pentraxin 1)	synaptic regeneration	sec	N
FBN2 (fibrillin-2)	ECM	sec	N
FBLN1 (fibulin-1)	ECM	sec	N
CLUS (clusterin)	extracell. chaperone	sec	N
LTBP4 (latent transforming growth factor β -binding protein 4)	ECM	sec	N
AGRIN	ECM	sec	N
COCH (cochlin)	ECM	sec	N
COL6A1 (collagen 6 α 1)	ECM	sec	N
FN1 (fibronectin)	ECM	sec	N, O
γ -Glutamyl hydrolase	γ -glutamyl peptidase	sec	N
Sulphydryl oxidase 1	protein disulfide isomerase	sec	N

(FSH β) to cysteine V (TFF2), for example. A subset of the disulfide bridge pattern in FSH β (Cys-21–Cys-69, Cys-46–Cys-100, and Cys-50–Cys-102) constitutes a cysteine knot motif similar to that found in the growth factor superfamily.

TFF Domains as Conformational Determinants—TFF domains were identified in this study as conformational determinants

inducing LacdiNAc formation on N-linked glycans of human TFF2. The LacdiNAc-inducing capacity was so strong that even though β 4Gal-T1 (lactose synthase) expression in HEK-293 cells is in the same range as β 4GalNAc-T4 expression (Model Organism Protein Expression Database), the latter overrides the former by increasing the N-linked LacdiNAc/LacNAc ratio

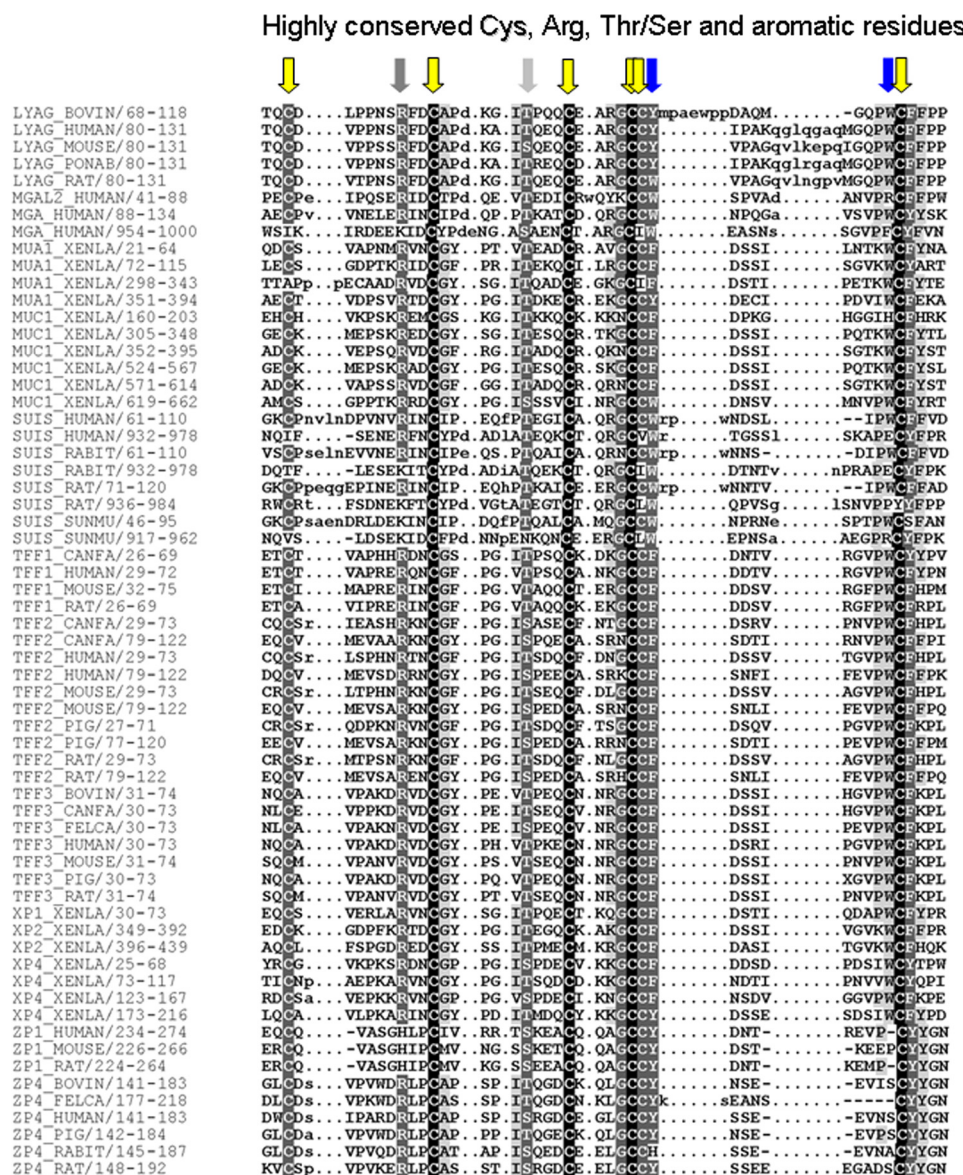


FIGURE 8. Multiple sequence alignments of hTFF2 with TFF domain-containing proteins. Sequence alignment of mammalian TFF peptides (TFF), lysosomal α -glucosidase (LYAG), maltase glucoamylase-like protein (MGAL2), maltase glucoamylase (MGA), *Xenopus* integumentary mucin A.1 (MUA1), and C.1 (MUC1), sucrose-isomaltase (SUIS), putative gastrointestinal growth factors from *Xenopus* (XP), and *zona pellucida* glycoproteins (ZP) in ClustaW (1.83). Arrows indicate highly conserved amino acid residues: cysteines (yellow), aromatic residues in loop 3 (blue), other residues (light or dark gray).

to ~10. The capacity of LacdiNAc induction was not confined to the native two-TFF domain structure of TFF2 but was seen also when monomeric (single) P1 or P2 domains of TFF2 were expressed as *N*-glycosylation probes N-P1 or C-P2*. Mutation of Cys-52 to Gly resulted in a nearly complete loss of LacdiNAc inducing capacity, and similar effects were seen in the Cys-42 mutant. The β 4-GalNAc transfer-inducing TFF domain needs to be in close proximity to the *N*-glycosylation site, as there was no compensating effect exerted by an intact P2 domain on LacdiNAc formation in the Cys/Gly-mutated P1 domain. An even more pronounced effect on *N*-glycosylation and LacdiNAc formation was revealed for the double point-mutated probe, which lacked components of a hydrophobic patch. To understand the strong effects seen with the latter mutant probe, three-dimensional views of the crystal structures of the struc-

turally closely related porcine spasmodic protein (pSP, pTFF2) were inspected.

The Potential Role of Hydrophobic Patches—In pSP aromatic residues and a series of other hydrophobic residues were found to line a cleft (Fig. 10), which has been suggested to bind mucin glycoproteins, and these aromatic rings were claimed to play also a role in stabilization of the fold of the domain (24). Two of these aromatic residues are highly conserved in most TFF domains throughout mammalian species and even in *Xenopus* mucins (Fig. 8). These are Phe-59 (corresponding to Phe-36 in the expressed protein) and Trp-68 (or Trp45) in the P1 domain (see also Phe-108 and Trp-117 in P2) both flanking loop 3. According to the crystal structure of porcine spasmodic protein (25–27), the aromatic residues corresponding to Phe-59/Trp-68 in the TFF2 P1 domain are located within the strand 1

TFF Domain Determines LacdiNAc Formation

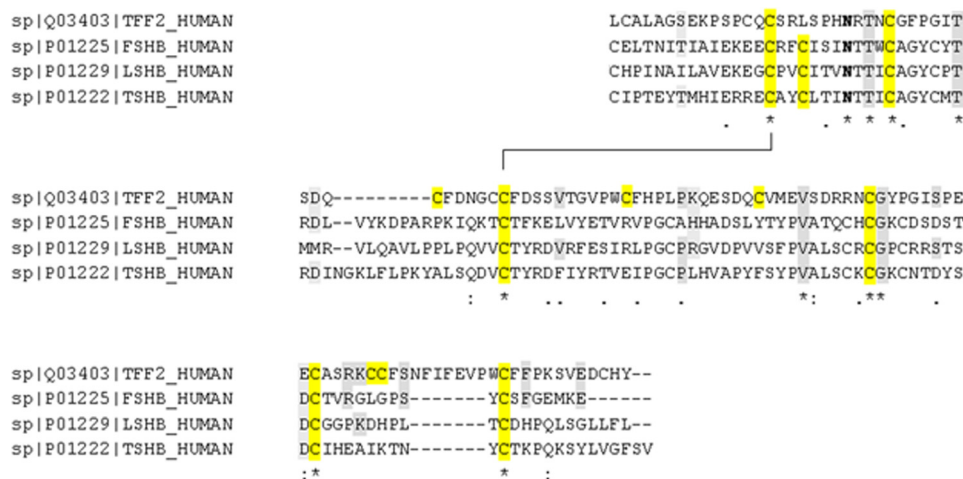


FIGURE 9. **Multiple sequence alignments of hTFF2 with human neuropeptide hormones.** Amino acid sequences of human neuropeptide hormone subunits FSH β , LSH β , and TSH β and of hTFF2 were aligned in ClustalW (1.83) as above in Fig. 8. Conserved amino acid residues are highlighted: cysteines (yellow), other residues (gray).

(Phe-59) and strand 2 (Trp-68) of the two-stranded antiparallel β -sheet at the central core region of loop 3 (Fig. 10). These aromatic residues together with other aromatic/hydrophobic residues clustering around the β -turn in loop 2 (Gly-43, Phe-44, Gly-46, Ile-47, Phe-53, Val-63, Gly-65, Val-66) form a hydrophobic binding pocket that might be involved in protein-protein interactions. In accordance with this, the double mutant TFF2-(W68Q/F59Q) was devoid of any LacdiNAc expression and showed a shift from complex to high mannose-type *N*-glycosylation.

Similar Hydrophobic Patches in Other LacdiNAc-expressing Proteins—The above discussed structural features of TFF2 resemble the cystine-knot modules of the neuropeptide hormones, where a number of hydrophobic residues are exposed in a loop-like structure between the two highly twisted double-stranded β -sheets (28).

TFF domains are also found in *hZP1* (Fig. 8) but not in the LacdiNAc-positive *hZP3* and most of the other structurally defined or proposed LacdiNAc-positive glycoproteins. According to the above discussed features of a conformationally stabilized hydrophobic patch, it can be assumed that similar patches may be involved in the control of LacdiNAc formation on other glycoproteins. Two highly conserved regions of the N-terminal ZP-N domain on chicken ZP3, for example, expose hydrophobic residues (Tyr-111 and Tyr-141, both conserved in human ZP3) and do not overlap with sites that can be glycosylated (29).

In human nidogen-1, previously reported to express LacdiNAc (16), an 11-stranded β -barrel is characterized by a patch on the barrel surface, which is conserved in all metazoan nidogens. Site-directed mutagenesis shows that the residues in the conserved patch are involved in the binding of perlecan (30). Among these surface-exposed residues two invariant tyrosines (Tyr-431 and Tyr-440) and one valine (Val-433) may contribute to the formation of a hydrophobic patch.

Potential Role of Cis-oriented Peptide Elements—Inspection of the conformational pSP model further reveals that the only surface-exposed basic residues are in close proximity of the *N*-glycosylation site within loop 1 (Asn-38 -5, -1, +1). An involvement of these residues in the binding of β 4GalNAc-T3 or -T4 is

highly unlikely due to steric hindrance by the glycan substrate. Although previous work had suggested that basic residues within a cis-located α -helical peptide may contribute to a LacdiNAc-inducing determinant, we could not obtain any supporting evidence for the existence of such a peptide determinant in the TFF domains of TFF2. Our findings do not rule out the contribution of CA6-related basic peptide determinants to the formation of LacdiNAc on other glycoproteins. Actually, we could confirm, at least for artificial peptides with a high content of basic residues (oligo-His tag), that LacdiNAc synthesis is induced at adjacent *O*-glycosylation sites in *hZP3* probes.³ However, the more distant oligo-histidine tag in hTFF2-FI or N-P1 did not induce LacdiNAc synthesis on the *N*-glycosylation site (refer to the LacdiNAc-negative T68Q/F59Q double mutant probe). The observation that *O*-linked chains on TFF2 do not express LacdiNAc (Fig. 7) appears to be of considerable importance as it demonstrates that the effect exerted by the hydrophobic patch on activation of β 4-GalNAc-T3/T4 is dependent on steric aspects.

Biological Relevance of Protein-specific LacdiNAc Expression—The restricted and apparently protein-specific expression of LacdiNAc points to important biological functions. LacdiNAc-positive glycans play vital roles in regulating the circulatory half-lives of pituitary glycoprotein hormones (4–6) and other glycoproteins (31) including tenascin-R produced by oligodendrocytes and small interneurons in the hippocampus and cerebellum (12). Other glycoproteins containing LacdiNAc-terminating glycans include human glycodefin, a human glycoprotein with potent immunosuppressive and contraceptive activities (7), and zona pellucida glycoproteins from murine eggs (14). Although still speculative, LacdiNAc expression on ZP3 has been claimed to form the molecular basis of initial sperm-egg binding (14, 32). Many human pathogens, among them *Schistosoma mansoni*, synthesize LacdiNAc (LDN) and fucosylated LacdiNAc glycans, such as GalNAc β 1–4-(Fuc α 1–3)GlcNAc-R, termed LDNF. Both LacdiNAc and fuco-

³ D. Bonar and F.-G. Hanisch, unpublished information.

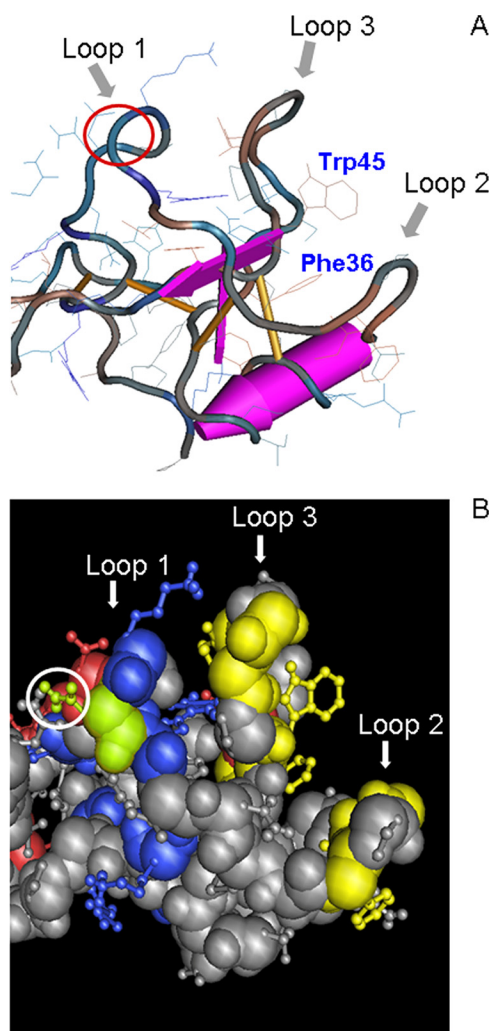


FIGURE 10. Three-dimensional view of porcine spasmodic protein N-terminal domain P1. *A*, backbone conformation of pSP depicted using a worm diagram of domain I with toggled amino acid side chains (24). The three loops of the domain are indicated by arrows. The putative binding pocket presumably involved in protein-protein interaction is formed by loops 2 and 3 and characterized by highly conserved Phe-36 and Trp-45 (corresponding to Phe-59 and Trp-68). The amino acid position p15 in loop 1 corresponding to the *N*-glycosylation site in TFF2 is marked by a red circle. Amino acid numbering refers to the expressed protein. *B*, space-fill model of pSP (amino acid side chains as a ball-and-stick model) with marked charged residues (basic, blue; acidic, red), and green-highlighted Asn-15 (Asn-38) together with yellow-highlighted residues Phe-21, Gly-23, Ile-24, Phe-30, Phe-36, Val-40, Gly-42, Val-43, Trp-45 (corresponding to Phe-44, Gly-46, Ile-47, Phe-53, Phe-59, Val-63, Gly-65, Val-66, Trp-68), forming a hydrophobic patch. Amino acid numbering refers to the expressed protein (numbering in brackets to the SwissProt database entries).

sylated LacdiNAc as well as other modified LDN-glycans represent important recognition determinants in adaptive immune responses and are recognized by various carbohydrate-binding proteins within the innate immune system, such as DC-SIGN (33–38). Another role of LacdiNAc expression in innate immune defense can be seen in the entrapment of *Helicobacter pylori* in the visco-elastic mucin gel overlaying the gastric surface epithelium. This mucin layer is characterized by the preponderant expression of LacdiNAc-positive MUC5AC, which was recently claimed to play a role in *H. pylori* colonization (39).

There is a growing body of evidence that β 4-GalNAc-T3 and LacdiNAc expression is involved in the modulation of tumor

cell growth and signaling. β 4-GalNAc-T3 was shown to predict a favorable prognosis for neuroblastoma and to suppress the malignant phenotype via decreasing β 1-integrin signaling (40). In the same context the enzyme was demonstrated to regulate cancer stemness and the invasive properties of colon cancer cells through modifying EGFR glycosylation and signaling (41).

Recent work has demonstrated the involvement of LacdiNAc in the self-renewal of mouse embryonic stem cells by regulating LIF/STAT3 signaling (42). Undifferentiated state mouse embryonic stem cells expressed LacdiNAc at higher levels than differentiated state cells, and expression of the glycan on the LIF receptor (gp130) was shown to be required for the stable colocalization of the receptor with lipid raft/caveolar proteins, such as caveolin-1, and for the transduction of a sufficiently strong LIF/STAT3 signal.

The functional implication of LacdiNAc-modified glycans in important biological processes, as outlined above, raises the question of how the specificity of enzymatic sugar transfer is achieved. These considerations have driven our attempts to define the minimal essential elements of a strong determinant as found on TFF domains. Future work will have to show whether the hypothesized role of hydrophobic patches can be generalized to explain protein-specific LacdiNAc formation.

Acknowledgments—We thank Prof. Werner Hoffmann (University of Magdeburg, Germany) for valuable discussions of our data and for kindly providing plasmid DNA as the template for TFF2 constructs.

REFERENCES

- Breloy, I., Schwientek, T., Gries, B., Razawi, H., Macht, M., Albers, C., and Hanisch, F.-G. (2008) Initiation of mammalian *O*-mannosylation *in vivo* is independent of a consensus sequence and controlled by peptide regions within and upstream of the α -dystroglycan mucin domain. *J. Biol. Chem.* **283**, 18832–18840
- Miller, E., Fiete, D., Blake, N. M., Beranek, M., Oates, E. L., Mi, Y., Roseman, D. S., and Baenziger, J. U. (2008) A necessary and sufficient determinant for protein-selective glycosylation *in vivo*. *J. Biol. Chem.* **283**, 1985–1991
- Hanisch, F.-G., and Breloy, I. (2009) Protein-specific glycosylation: signal patches and cis-controlling peptidic elements. *Biol. Chem.* **390**, 619–626
- Manzella, S. M., Hooper, L. V., and Baenziger, J. U. (1996) Oligosaccharides containing β 1,4-linked *N*-acetylgalactosamine, a paradigm for protein-specific glycosylation. *J. Biol. Chem.* **271**, 12117–12120
- Lowe, J. B., and Marth, J. D. (2003) A genetic approach to mammalian glycan function. *Annu. Rev. Biochem.* **72**, 643–691
- Green, E. D., van Halbeek, H., Boime, I., and Baenziger, J. U. (1985) Structural elucidation of the disulfated oligosaccharide from bovine lutropin. *J. Biol. Chem.* **260**, 15623–15630
- Dell, A., Morris, H. R., Easton, R. L., Panico, M., Patankar, M., Oehninger, S., Koistinen, R., Koistinen, H., Seppala, M., and Clark, G. F. (1995) Structural analysis of the oligosaccharides derived from glycodefin, a human glycoprotein with potent immunosuppressive and contraceptive activities. *J. Biol. Chem.* **270**, 24116–24126
- Manzella, S. M., Dharmesh, S. M., Cohick, C. B., Soares, M. J., and Baenziger, J. U. (1997) Developmental regulation of a pregnancy-specific oligosaccharide structure, NeuAc α 2,6 GalNAc β 1,4GlcNAc on selected members of the rat placental prolactin family. *J. Biol. Chem.* **272**, 4775–4782
- Skelton, T. P., Kumar, S., Smith, P. L., Beranek, M. C., and Baenziger, J. U. (1992) Pro-opiomelanocortin synthesized by corticotrophs bears asparagine-linked oligosaccharides terminating with SO₄-4GalNAc β 1,4GlcNAc β 1,2Man α . *J. Biol. Chem.* **267**, 12998–13006
- Fiete, D., Mi, Y., Oats, E. L., Beranek, M. C., and Baenziger, J. U. (2007)

TFF Domain Determines LacdiNAc Formation

- N-Linked oligosaccharides on the low-density lipoprotein receptor homolog SorLA/LR11 are modified with terminal GalNAc-4-SO₄ in kidney and brain. *J. Biol. Chem.* **282**, 1873–1881
- Martinez-Pomares, L., Crocker, P. R., Da Silva, R., Holmes, N., Colominas, C., Rudd, P., Dwek, R., and Gordon, S. (1999) Cell-specific glycoforms of sialoadhesin and CD45 are counter-receptors for the cysteine-rich domain of the mannose receptor. *J. Biol. Chem.* **274**, 35211–35218
 - Woodworth, A., Fiete, D., and Baenziger, J. U. (2001) Tenascin-R in the cerebellum is modified with N-linked oligosaccharides terminating with GalNAc-4-SO₄. *Glycobiology* **11**, (Abstr. 45)
 - Hooper, L. V., Beranek, M. C., Manzella, S. M., and Baenziger, J. U. (1995) Differential expression of GalNAc-4-sulfotransferase and GalNAc-transferase results in distinct glycoforms of carbonic anhydrase VI in parotid and submaxillary glands. *J. Biol. Chem.* **270**, 5985–5993
 - Dell, A., Chalabi, S., Easton, R. L., Haslam, S. M., Sutton-Smith, M., Patankar, M. S., Lattanzio, F., Panico, M., Morris, H. R., and Clark, G. F. (2003) Murine and human zona pellucida 3 derived from mouse eggs express identical O-glycans. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15631–15636
 - Siciliano, R. A., Morris, H. R., Bennett, H. P., and Dell, A. (1994) O-glycosylation mimics N-glycosylation in the 16-kDa fragment of bovine proopiomelanocortin: the major O-glycan attached to Thr-45 carries SO₄-4GalNAcβ1-4GlcNAcβ-1-, which is the archetypal nonreducing epitope in the N-glycans of pituitary glyco hormones. *J. Biol. Chem.* **269**, 910–920
 - Breloy, I., Pacharra, S., Ottis, P., Bonar, D., Grahn, A., and Hanisch, F.-G. (2012) O-Linked N,N'-diacetylglucosamine (LacdiNAc)-modified glycans in extracellular matrix glycoproteins are specifically phosphorylated at subterminal N-acetylglucosamine. *J. Biol. Chem.* **287**, 18275–18286
 - Sato, T., Gotoh, M., Kiyohara, K., Kameyama, A., Kubota, T., Kikuchi, N., Ishizuka, Y., Iwasaki, H., Togayachi, A., Kudo, T., Ohkura T., Nakanishi, H., and Narimatsu H. (2003) Molecular cloning and characterization of a novel human β1-4-N-acetylgalactosaminyltransferase, β4GalNAc-T3, responsible for the synthesis of N,N'-diacetylglucosamine, GalNAcβ1-4-GlcNAc. *J. Biol. Chem.* **278**, 47534–47544
 - Gotoh, M., Sato, T., Kiyohara, K., Kameyama, A., Kikuchi, N., Kwon, Y. D., Ishizuka, Y., Iwai, T., Nakanishi, H., and Narimatsu, H. (2004) Molecular cloning and characterization of β1,4-N-acetylgalactosaminyltransferase IV synthesizing N,N'-diacetylglucosamine. *FEBS-Lett.* **562**, 134–140
 - Hanisch, F.-G., Ragge, H., Kalinski, T., Meyer, F., Kalbacher, H., and Hoffmann, W. (2013) Human gastric TFF2 peptide contains an N-linked fucosylated N,N'-diacetylglucosamine (LacdiNAc) oligosaccharide. *Glycobiology* **23**, 2–11
 - Do, K. Y., Do, S. I., and Cummings, R. D. (1997) Differential expression of LacdiNAc sequences (GalNAcβ1-4GlcNAc-R) in glycoproteins synthesized by Chinese hamster ovary and human 293 cells. *Glycobiology* **7**, 183–194
 - Müller, S., and Hanisch, F.-G. (2002) Recombinant MUC1 probe authentically reflects cell-specific O-glycosylation profiles of endogenous breast cancer mucin: high density and prevalent core 2-based glycosylation. *J. Biol. Chem.* **277**, 26103–26112
 - Albert, T. K., Laubinger, W., Müller, S., Hanisch, F. G., Kalinski, T., Meyer, F., and Hoffmann, W. (2010) Human intestinal TFF3 forms disulfide-linked heteromers with the mucus-associated FCGBP protein and is released by hydrogen sulfide. *J. Proteome Res.* **9**, 3108–3117
 - Ceroni, A., Maass, K., Geyer, H., Geyer, R., Dell, A., and Haslam, S. M. (2008) Glycoworkbench: A tool for the computer-assisted annotation of mass spectra of glycans. *J. Proteome Res.* **7**, 1650–1659
 - Gajhede, M., Petersen, T. N., Henriksen, A., Petersen, J. F., Dauter, Z., Wilson, K. S., and Thim, L. (1993) Pancreatic spasmolytic polypeptide: first three-dimensional structure of a member of the mammalian trefoil family of peptides. *Structure* **1**, 253–262
 - De, A., Brown, D. G., Gorman, M. A., Carr, M., Sanderson, M. R., and Freemont, P. S. (1994) Crystal-structure of a disulfide-linked trefoil motif found in a large family of putative growth-factors. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1084–1088
 - Carr, M. D., Bauer, C. J., Gradwell, M. J., and Feeney, J. (1994) Solution structure of a trefoil-motif-containing cell-growth factor, porcine spasmolytic protein. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2206–2210
 - Petersen, T. N., Henriksen, A., and Gajhede, M. (1996) Structure of porcine pancreatic spasmolytic polypeptide at 1.95 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 730–737
 - Bork, P., Downing, A. K., Kieffer, B., and Campbell, I. D. (1996) Structure and distribution of modules in extracellular proteins. *Q. Rev. Biophys.* **29**, 119–167
 - Monné, M., Han, L., Schwend, T., Burendahl, S., and Jovine, L. (2008) Crystal structure of the ZP-N domain of ZP3 reveals the core fold of animal egg coats. *Nature* **456**, 653–657
 - Hopf, M., Göhring, W., Ries, A., Timpl, R., and Hohenester, E. (2001) Crystal structure and mutational analysis of a perlecan-binding fragment of nidogen-1. *Nat. Struct. Biol.* **8**, 634–640
 - Chiu, M. H., Tamura, T., Wadhwa, M. S., and Rice, K. G. (1994) *In vivo* targeting function of N-linked oligosaccharides with terminating galactose and N-acetylglucosamine residues. *J. Biol. Chem.* **269**, 16195–16202
 - Benoff, S. (1997) Carbohydrates and fertilization: an overview. *Mol. Hum. Reprod.* **3**, 599–637
 - Nyame, A. K., Leppänen, A. M., Bogitsh, B. J., and Cummings, R. D. (2000) Antibody responses to the fucosylated LacdiNAc glycan antigen in *Schistosoma mansoni*-infected mice and expression of the glycan among schistosomes. *Exp. Parasitol.* **96**, 202–212
 - Nyame, A. K., Leppänen, A. M., DeBose-Boyd, R., and Cummings, R. D. (1999) Mice infected with *Schistosoma mansoni* generate antibodies to LacdiNAc (GalNAcβ1→4GlcNAc) determinants. *Glycobiology* **9**, 1029–1035
 - Nyame, A. K., Kwar, Z. S., and Cummings, R. D. (2004) Antigenic glycans in parasitic infections: implications for vaccines and diagnostics. *Arch. Biochem. Biophys.* **426**, 182–200
 - van den Berg, T. K., Honing, H., Franke, N., van Remoortere, A., Schiphorst, W. E., Liu, F. T., Deelder, A. M., Cummings, R. D., Hokke, C. H., and van Die, I. (2004) LacdiNAc-glycans constitute a parasite pattern for galectin-3-mediated immune recognition. *J. Immunol.* **173**, 1902–1907
 - van Die, I., van Vliet, S. J., Nyame, A. K., Cummings, R. D., Bank, C. M., Appelmelk, B., Geijtenbeek, T. B., and van Kooyk, Y. (2003) The dendritic cell-specific C-type lectin DC-SIGN is a receptor for *Schistosoma mansoni* egg antigens and recognizes the glycan antigen Lewis x. *Glycobiology* **13**, 471–478
 - van Remoortere, A., Vermeer, H. J., van Roon, A. M., Langermans, J. A., Thomas, A. W., Wilson, R. A., van die, I., van den Eijnden, D. H., Agoston, K., Kérékyarto, J., Vliegthart, J. F., Kamerling, J. P., van dam, G. J., Hokke, C. H., and Deelder, A. M. (2003) Dominant antibody responses to Fuca1-3GalNAc and Fuca1-2Fuca1-3GlcNAc containing carbohydrate epitopes in Pan troglodytes vaccinated and infected with *Schistosoma mansoni*. *Exp. Parasitol.* **105**, 219–225
 - Kenny, D. T., Skoog, E. C., Lindén, S. K., Struwe, W. B., Rudd, P. M., and Karlsson, N. G. (2012) Presence of terminal N-acetylglucosamineβ1-4N-acetylglucosamine residues on O-linked oligosaccharides from gastric MUC5AC: involvement in *Helicobacter pylori* colonization? *Glycobiology* **22**, 1077–1085
 - Hsu, W. M., Che, M. I., Liao, Y. F., Chang, H. H., Chen, C. H., Huang, Y. M., Jeng, Y. M., Huang, J., Quon, M. J., Lee, H., Huang, H. C., and Huang, M. C. (2011) β4-GalNAc-T3 expression predicts a favourable prognosis and suppresses cell migration and invasion via β1-integrin signaling in neuroblastoma. *Am. J. Pathol.* **179**, 1394–1404
 - Che, M. I., Huang, J., Hung, J. S., Lin, Y. C., Huang, M. J., Lai, H. S., Hsu, W. M., Liang, J. T., and Huang, M. C. (2014) β1-4-N-acetylglucosaminyltransferase 3 modulates cancer stemness through EGFR signaling pathway in colon cancer cells. *Oncotarget* **5**, 3673–3684
 - Sasaki, N., Shinomi, M., Hirano, K., Ui-Tei, K., and Nishihara, S. (2011) LacdiNAc (GalNAcβ1-4GlcNAc) contributes to self-renewal of mouse embryonic stem cells by regulating leukemia inhibitory factor/STAT3 signaling. *Stem Cells* **29**, 641–650