Enzymatic synthesis of branched polylactosamines

Heidi Salminen

Institute of Biotechnogy
and
Department of Biosciences, Division of Biochemistry
and
Viikki Graduate School in Biosciences
Faculty of Science
University of Helsinki
Finland

Academic dissertation

To be presented for public criticism, with the permission of the Faculty of Science, University of Helsinki, in the auditorium 2041 at Viikki Biocenter, Viikinkaari 5, Helsinki, on December 17th, 2003, at 12 o’clock noon

Helsinki 2003
To the memory of my father
## CONTENTS

**ORIGINAL PUBLICATIONS**................................................................. 1

**ABBREVIATIONS**............................................................................. 2

**SUMMARY**......................................................................................... 3

### 1. REVIEW OF THE LITERATURE........................................................ 5

1.1 Introduction to polylactosaminoglycan structures............................... 5
1.2 Expression of branched polylactosaminoglycans................................ 7
   1.2.1 Developmentally regulated expression in human erythrocytes........... 7
   1.2.2 Cell type specific expression...................................................... 8
   1.2.3 Expression in embryonic cells.................................................... 9
   1.2.4 Branched polylactosamines from other sources............................. 9
1.3 Expression of polylactosaminoglycans with terminal type 1 unit............. 10
1.4 Biosynthesis of polylactosaminoglycans.......................................... 12
   1.4.1 β1,3-N-acetylglucosaminyltransferases...................................... 13
   1.4.2 β1,4-galactosyltransferases....................................................... 14
   1.4.3 β1,3-galactosyltransferases....................................................... 15
   1.4.4 β1,6-N-acetylglucosaminyltransferases...................................... 15
     1.4.4.1 Distally acting β1,6-N-acetylglucosaminyltransferases............. 16
     1.4.4.2 Centrally acting β1,6-N-acetylglucosaminyltransferases........... 17
1.5 Structural analysis of polylactosamines by NMR spectroscopy............... 19
   1.5.1 One-dimensional 1H NMR spectroscopy...................................... 19
   1.5.2 Two-dimensional NMR spectroscopy.......................................... 21
1.6 Biological functions that involve polylactosaminoglycans.................... 21
   1.6.1 Selectin mediated cell adhesion.............................................. 22
   1.6.2 Galectin mediated cell adhesion.............................................. 24
   1.6.3 Fertilization............................................................................. 25
   1.6.4 Microbe adhesion..................................................................... 26

### 2. AIMS OF THE STUDY...................................................................... 27

### 3. MATERIALS AND METHODS.......................................................... 28

3.1 Acceptor saccharides...................................................................... 28
3.2 Expression and purification of GST-IGnT1....................................... 28
3.3 Enzyme catalyzed transferase reactions............................................ 28
3.4 Degradative reactions ................................................................. 29
3.5 Chromatographic methods ......................................................... 29
3.6 NMR spectroscopy ................................................................. 29
3.7 Mass spectrometry ................................................................. 29

4. RESULTS .................................................................................. 30

4.1 Synthesis of multiply branched polylactosamines in vitro (I) ................. 30
4.2 Enzymatic synthesis of precursors for a L-selectin oligosaccharide
antagonist (I,II) ........................................................................... 32
4.3 The acceptor specificity of the centrally acting β1,6-N-acetylglucosaminyl-
transferase cloned from human embryonal carcinoma cell line PA1 (III) ............ 33
4.4 Enzymatic synthesis of linear and branched polylactosamines containing
terminal type 1 units (IV) ............................................................ 33
4.5 Structural analysis of the products by NMR spectroscopy (I-IV) ................. 34

5. DISCUSSION ............................................................................. 37

5.1 Centrally acting β1,6-N-acetylglucosaminyltransferases synthesize
multiple branches on linear polylactosamines in vitro .................................. 37
5.2 Branched polylactosamines as precursors for biologically active
oligosaccharides ........................................................................ 39
5.3 Biosynthetic pathways to polylactosamines with terminal type 1 unit ............ 39
5.4 Structural determination of polylactosamines by NMR spectroscopy ............. 40

6. ACKNOWLEDGEMENTS ............................................................. 42

7. REFERENCES ............................................................................ 44
This thesis is based on the following original publications, which are referred to in the text by their roman numerals:


IV  Heidi Salminen, Jari Natunen, Hannu Maaheimo, Risto Renkonen, and Ossi Renkonen. Enzymatic synthesis and NMR spectroscopic analysis of linear and branched polylactosamines containing terminal type 1 epitope. Manuscript in preparation

---

*a* Reprinted from *Biochemistry* with permission of American Chemical Society.

*b* Reprinted from *FEBS letters* with permission of Elsevier.

*c* Reprinted from *Journal of Biological Chemistry* with permission of American Society for Biochemistry and Molecular Biology.
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1D</td>
<td>one-dimensional</td>
</tr>
<tr>
<td>2D</td>
<td>two-dimensional</td>
</tr>
<tr>
<td>β3Gal-T</td>
<td>β1,3-galactosyltransferase</td>
</tr>
<tr>
<td>β4Gal-T</td>
<td>β1,4-galactosyltransferase</td>
</tr>
<tr>
<td>β3Gn-T</td>
<td>β1,3-N-acetylglucosaminyltransferase</td>
</tr>
<tr>
<td>Cer</td>
<td>ceramide</td>
</tr>
<tr>
<td>C2GnT</td>
<td>core 2 β1,6-N-acetylglucosaminyltransferase</td>
</tr>
<tr>
<td>cIGnT</td>
<td>centrally acting IGnT</td>
</tr>
<tr>
<td>dIGnT</td>
<td>distally acting IGnT</td>
</tr>
<tr>
<td>DQFCOSY</td>
<td>double quantum filtered correlated spectroscopy</td>
</tr>
<tr>
<td>Fuc</td>
<td>L-fucose</td>
</tr>
<tr>
<td>Gal</td>
<td>D-galactose</td>
</tr>
<tr>
<td>GalNAc</td>
<td>N-acetyl-D-galactosamine</td>
</tr>
<tr>
<td>Glc</td>
<td>D-glucose</td>
</tr>
<tr>
<td>GlcNAc</td>
<td>N-acetyl-D-glucosamine</td>
</tr>
<tr>
<td>HEV</td>
<td>high endothelial venules</td>
</tr>
<tr>
<td>Hex</td>
<td>hexose</td>
</tr>
<tr>
<td>HexNAc</td>
<td>N-acetylhexosamine</td>
</tr>
<tr>
<td>HMQC</td>
<td>heteronuclear multiple quantum coherence</td>
</tr>
<tr>
<td>HPAE</td>
<td>high-pH anion exchange</td>
</tr>
<tr>
<td>IGnT</td>
<td>blood group I β1,6-N-acetylglucosaminyltransferase</td>
</tr>
<tr>
<td>LacNAc</td>
<td>Galβ1-4GlcNAc</td>
</tr>
<tr>
<td>Lc3Cer</td>
<td>lactotriosylceramide, GlcNAcβ1-3Galβ1-4Glcβ1-Cer</td>
</tr>
<tr>
<td>Lea</td>
<td>Lewis a, Galβ1-3(Fucα1-4)GlcNAc</td>
</tr>
<tr>
<td>Leb</td>
<td>Lewis b, Fucα1-2Galβ1-3(Fucα1-4)GlcNAc</td>
</tr>
<tr>
<td>Lec</td>
<td>Lewis c, Galβ1-3GlcNAc</td>
</tr>
<tr>
<td>Lex</td>
<td>Lewis x, Galβ1-4(Fucα1-3)GlcNAc</td>
</tr>
<tr>
<td>MALDI-TOF MS</td>
<td>matrix-assisted laser desorption/ionization time-of-flight mass spectrometry</td>
</tr>
<tr>
<td>Neu5Ac</td>
<td>N-acetylneuraminic acid</td>
</tr>
<tr>
<td>nLc5Cer</td>
<td>neolactopentaosylceramide, GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4Glcβ1-Cer</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PGC</td>
<td>polyglycosylceramide</td>
</tr>
<tr>
<td>PLN</td>
<td>polylactosamine</td>
</tr>
<tr>
<td>sLex</td>
<td>sialyl Lewis x, Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc</td>
</tr>
<tr>
<td>TOCSY</td>
<td>total correlation spectroscopy</td>
</tr>
</tbody>
</table>
SUMMARY

Carbohydrates that cover cells have diverse biological functions. They function in cell differentiation, development, and aggregation of cells to form organs. Carbohydrate-binding proteins called lectins recognize specific carbohydrate structures and mediate cell-cell adhesion. For example, selectins guide leukocyte traffic to the sites of inflammation and their homing to lymph nodes. Processes such as fertilization and infection of cells by bacteria and viruses involve carbohydrates as well. In addition, many common diseases have turned out to be linked to specific glycosylation.

Protein- and lipid-bound glycans commonly contain polylactosamine (PLN) chains. PLNs are composed of type 2 N-acetyllactosamine (LacNAc, Galβ1-4GlcNAc) units, which are coupled via β1-3 linkages to form a linear structure. Some of the Gal residues in the linear chain may be substituted with LacNAc units at the C-6 hydroxyl group forming a branched PLN. The linear and branched PLNs are known as blood group i and I antigen structures, respectively. PLNs form backbones for bioactive terminal epitopes, such as fucosylated and sialylated structures. Branching is biologically important since each branch may carry these epitopes, and the presentation of multiple epitopes enhances biological activity as compared to glycans with a single epitope.

In this thesis, the conversion of linear PLNs to branched ones was studied. Blood group I β1,6-N-acetylglucosaminyltransferase activity of rat serum was shown to synthesize multiple GlcNAcβ1-6 branches to the midchain galactoses of linear PLNs. This type of branching activity has been termed centrally acting β1,6-N-acetylglucosaminyltransferase (cIGnT) activity as opposed to distally acting β1,6-N-acetylglucosaminyltransferase (dIGnT) activity, which transfers GlcNAc to the distal GlcNAcβ1-3LacNAc-R sequence. The cIGnT(s) converted LacNAcβ1-3LacNAcβ1-3LacNAc into a doubly branched glycan and successful β1,4-galactosylation of the GlcNAc-branches gave LacNAcβ1-3(LacNAcβ1-6)LacNAcβ1-3(LacNAcβ1-6)LacNAc, a prototype of naturally occurring multiply branched polylactosamines. The triply branched LacNAcβ1-3(LacNAcβ1-6)LacNAcβ1-3(LacNAcβ1-6)LacNAc structure was synthesized in a similar way. The triply branched glycan is a precursor for a tetravalent sialyl Lewis x (sLex, Neu5Acα2-3Galβ1-4(Fucα1-3)GlcNAc) glycan that has been shown to inhibit L-selectin-mediated adhesion of lymphocytes to activated endothelium.

In another part of the work, the IGnT from human embryonal carcinoma cell line PA1 was functionally expressed and purified. The acceptor specificity study of IGnT showed that the enzyme possesses cIGnT type activity and is capable of forming multiple branches in a PLN acceptor.

In certain tissues, the type 1 disaccharide unit (Lec, Galβ1-3GlcNAc) is conjugated to the nonreducing end of PLNs. The β1,3-galactosyltransferase (β3Gal-T) activity present in human colon adenocarcinoma cell line Colo 205 was shown to β1,3-galactosylate different PLN acceptors in a fashion that corresponds to the naturally occurring type 1 epitope.
structures. It was shown that branched type 1 epitopes, Lecβ1-3(GlcNAcβ1-6)LacNAc-R, may be synthesized by two alternative pathways; β3Gal-T may react with the prebranched acceptor or cIgGnT (rat serum) may branch Lecβ1-3LacNAc-R.

NMR spectroscopy was used in the structural analysis of the products together with enzymatic methods.
1. REVIEW OF THE LITERATURE

1.1 Introduction to polylactosaminoglycan structures

Poly-N-acetyllactosamines (PLNs, referred here as polylactosamines) form the backbone of many animal and some viral glycans, reviewed in the thesis of A. Leppänen [1], and bacterial glycans [2-4]. Linear polylactosamines consist of type 2 lactosamine (LacNAc, Galβ1-4GlcNAc) units linked together by β1,3-linkages or more rarely by β1,6-linkages [5]. In branched polylactosamines, one or several galactose units are substituted with β1,6-linked GlcNAc or LacNAc units. Branches may be further extended with additional LacNAc units in linear or branched form. The type 1 lactosamine (Lec, Galβ1-3GlcNAc) unit may occur in the nonreducing terminus or internal positions of PLN chains. The expression of type 1 structures is restricted to epithelial tissues in humans (see section 1.3). N-acetylactosdiamine (LacdiNAc, GalNAcβ1-4GlcNAc) unit occur at terminal ends of PLNs on vertebrate and invertebrate glycoproteins [6]. Examples of PLN structures are presented in Figure 1.

Polylactosamines can be attached to N-glycans and O-glycans of glycoproteins and glycolipids, and occur as free oligosaccharides in secretions. Complex type bi-, tri-, and tetraantennary N-glycans and four major O-glycan core types carry PLN chains (Figure 2). PLNs occur favorably on the Manα1-6 rather than the Manα1-3 arm of N-glycans [7]. In O-glycans, PLNs usually occur on the β1,6-branch of core 2 or core 4 structures [8]. O-glycans usually carry PLNs that are composed of 1-3 LacNAc units [9-11], whereas N-glycans contain PLNs of three or more LacNAc units [12-14]. Proteoglycans carry keratan sulfates, sulfated linear polylactosamines that are N- or O-linked to the protein [15]. In glycolipids, polylactosamines are formed on lactosylceramides (LacCer, Galβ1-4Glcβ1-Cer). LacCer is
elongated with LacNAc or Lec units and the corresponding core structures are called neolacto (nLC) and lacto (LC) series, respectively [16]. Glycolipids that have large glycans (about 20-50 monosaccharides) are called polyglycosylceramides (PGCs) [17, 18]. Human milk [19] and seminal plasma [20] contain significant amounts of free oligosaccharides. PLN chains in milk are attached to lactose (Lac, Galβ1-4Glc) and those in seminal plasma to GlcNAcβ1-3/4Glc.

The biological activity and heterogeneity of polylactosamines is provided by the degree of polylactosamine elongation and branching, in addition to different mosaccharide units linked to internal and terminal positions. Linear chains are frequently terminated by functional, often cell-type specific structures, such as the blood group ABO antigens and Lewis antigens (Figure 3). Internal GlcNAc units may be α1,3-fucosylated. The branched chains carry terminal groups on the branches as well.

**Figure 2.** The structures of N-glycan and O-glycan cores that carry polylactosamine chains. For N-linked glycan the possible side chains are marked with arrowheads. A tetraantennary N-glycan contains all four side chains, a triantennary one may contain R-GlcNAcβ1-2 and R-GlcNAcβ1-6 side chains and is called 2,6-branched, or R-GlcNAcβ1-2 and R-GlcNAcβ1-4 chains when it is called 2,4-branched. A biantennary N-glycan has R-GlcNAcβ1-2 side chains.
1.2 Expression of branched polylactosaminoglycans

The human blood group i and I antigens are characterized as linear and branched polylactosamines LacNAcβ1-3LacNAc-R and LacNAcβ1-3(LacNAcβ1-6)LacNAc-R, respectively [21-23]. Ii antigens are present in erythrocytes, various tissues and body fluids, and they are recognized as histo-blood group antigens [24-26]. The expression of Ii antigens is developmentally regulated in human erythrocytes and other tissues including epithelia, and during mouse embryogenesis [26, 27].

1.2.1 Developmentally regulated expression in human erythrocytes

Fetal and neonatal erythrocytes contain predominantly i antigens. The expression of i antigens decreases after birth, while the expression of I antigens increases until adult I antigen level is reached at 18 months [28, 29].

High molecular weight glycopeptides from adult erythrocytes were found to contain large and multiply branched N-linked PLNs that carry blood group epitopes [30-32]. At the same time, the anion transport (band 3) and glucose transport (band 4.5) glycoproteins, and glycolipids were shown to be carriers of Ii and blood group antigens on erythrocytes, band 3 being the major carrier [33-35]. One of the major lactosaminoglycans on fetal band 3 is a bi-antennary N-glycan carrying a linear polylactosamine of six LacNAc units on the Manα1,6-arm and a shorter chain on the Manα1,3-arm. The longer chain may be capped with Fucα1,2 or Neu5Acα2,3 residue, whereas the shorter chain contains Neu5Acα2,6 residue [36]. Adult band 3 polylactosamine chains are a few LacNAc units longer and uniformly branched with LacNAc units. Typically, one or more unbranched LacNAc units are located between the branched ones. Fucosylation and sialylation take place preferentially at the terminal region.
The number of terminal substitutions changes during erythrocyte maturation: α2,3-sialylation decreases and α1,2-fucosylation increases [7]. The polylactosamine chains of adult band 4.5 are similar to those of band 3, except that the number of branches and blood group determinants is smaller [37]. Another major glycoprotein on the erythrocyte membrane, glycoporphin A, is differently glycosylated from band 3 and 4.5. Glycoporphin A lacks polylactosamines, but instead contains single α2,6-sialylated LacNAc units linked to a biantennary N-glycan core and α2,3- and/or α2,6-sialylated core 1 O-glycans [38, 39].

The polylactosamines of glycolipids also change during erythrocyte maturation. Both neutral and acidic glycolipids of newborn and fetal erythrocytes contain smaller quantities of long PLN chains, and chains are more sialylated and less branched than in adult erythrocytes [40-42]. The branched glycolipid structures of adult erythrocytes have been characterized as (LacNAcβ1-3)1-2(LacNAcβ1-6)LacNAcβ1-3(LacNAcβ1-3)0-1Lacβ1-Cer. The terminal Gal residues are capped with Fucα1,2 and Galα1,3 or GalNAcα1,3 residues depending of the blood group [43-45]. Larger, branched polylactosamines with blood group epitopes are present on polyglycosylceramides (PGCs) that comprise up to 60 saccharide residues per ceramide [17]. Lex and Ley epitopes on PGCs have been detected in minor amounts [18].

The binding of anti-ABO antibodies to infant erythrocytes, which are deficient of branched ABH blood group active polylactosamines, has been shown to be relative weak as compared to the binding to adult erythrocytes, which present bivalent ABH determinants. Thus, it has been suggested that the lack of branched bivalent ABH determinants in the fetus may have a protective effect in an ABO-incompatible pregnancy [46].

### 1.2.2 Cell type specific expression

Erythrocytes and granulocytes directly differentiate from the same precursor stem cells. In contrast to adult erythrocytes, granulocytes express only linear polylactosamines on tetraantennary N-glycans [13, 47], core 2 based O-glycans [9], and glycolipids [48, 49]. PLN chains bound to N-glycans and glycolipids may be polyfucosylated, capped with Neu5Acα2,3 or Neu5Acα2,6 residues, and form Lex and sLex determinants [13, 47-49]. O-glycans, which are mainly derived from leukosialin (CD43), are elongated at the core β1,6-branch predominantly with one LacNAc unit and have Neu5Acα2,3 residues at the terminal ends. Minor O-glycans contain 2-3 repeating LacNAc units on the core β1,6-branch, and some O-glycans carry the sLex epitope [9]. Among the blood cells, only granulocytes and monocytes are enriched with Lex and sLex structures, whereas the ABO blood group antigens are restricted to erythroid cells.
1.2.3 Expression in embryonic cells

Embryonal carcinoma (EC) cells are undifferentiated stem cells that resemble the cells in the early embryo, and have been used as a model in studies of early embryogenesis [50]. Glycopeptides containing unusual large oligosaccharides composed of LacNAc-repeats have been identified from human teratocarcinoma-derived EC cells, PA1 [51, 52]. Detailed structural analysis of the glycans has shown that they are branched polylactosamines attached to tri- and tetraantennary N-glycan cores. The PLN chains contain 10-18 LacNAc units in linear form and 25-40% of the galactose units are branched with LacNAc units. Their nonreducing ends are substituted with Neu5Acα2,3/α2,6 residues or with the disialyl structure Neu5Acα2-9Neu5Acα2-3/6. Notably, no fucosylation has been detected [14]. Glycolipids characterized from PA1 cells contain short, unbranched glycans, for example Galβ1-3/GlcNAcβ1-3Lac-Cer capped with Neu5Acα2,3/α2,6 residues [53].

Mouse EC cells and early embryonic cells have protein-bound large glycans that contain highly branched and terminally fucosylated PLNs [54, 55]. The protein-carbohydrate linkage appears to be N-glycosidic [55, 56]. During the differentiation of EC cells, the large glycopeptides disappear almost completely [54]. Consistently, undifferentiated EC cells express mainly I antigen and differentiated cells i antigen [57, 58]. In the course of mouse embryogenesis, I antigen is expressed in the early embryonic cells throughout the preimplantation period, whereas i antigen is first detected in the differentiated cells of the 5-day embryo [57, 59]. The high-molecular weight glycoproteins of EC cells carry the I/i antigens and the stage-specific embryonic antigen-1, SSEA-1 [58]. The antigenic determinant of SSEA-1 is Lex [60]. SSEA-1 is expressed in undifferentiated EC cells and appears in mouse eight-cell embryos, but disappears from differentiated EC cells and embryonic cells during embryogenesis [58, 61]. Oligosaccharides with Lex determinant(s) have been shown to inhibit mouse embryo compaction in vitro, suggesting that SSEA-1 might participate in the adhesive events of preimplantation embryogenesis [62, 63].

1.2.4 Branched polylactosamines from other sources

The increased expression of branched PLNs in N-glycans has been detected in some cells after malignant transformation. Highly metastatic human colon carcinoma cells express more PLN side chains with branched galactose residues than cells with low metastatic potential [64]. Human thyroglobulin from malignant thyroid tissues carries high-molecular mass oligosaccharides, which are likely to contain highly branched PLNs, whereas normal thyroid tissues lack analogous oligosaccharides [65].

Branched PLNs play roles in cellular interactions. Laminin is a high molecular weight glycoprotein that occurs in the basement membranes of a variety of tissues and has been shown to interact with the cell adhesion proteins galectin-1 and galectin-3 [66, 67]. Laminin
from the murine Engelbreth-Holm-Swarm (EHS) tumor has been shown to contain both linear and branched PLNs linked to biantennary N-glycans [68]. Mouse uterine epithelial cells express N-linked PLNs that are involved in cell adhesion processes [69]. Estrogen stimulates PLN synthesis in uteri, and the resulting PLNs may be highly branched [70]. The surface of the parasitic prozoan *Trypanosoma brucei* is covered with a single species of the variant surface glycoprotein (VSG), which protects the organism against lysis by host serum components. Type II VSG expresses branched PLNs on a biantennary N-glycan [71].

O-glycans of human gastric [72, 73], colonic [74, 75], and bronchial mucins [76], and milk secretory IgA [77] contain singly branched PLNs. In addition, O-glycans of rat colonic [78] and swine tracheal mucins [79, 80], and pig stomach cell linings [81] bear branched PLNs. Glycolipids that carry singly branched PLNs are present in human myelogenous leukemia cells (HL-60) as minor components [82], hog gastric mucosa [83], and epithelial cells of rat small intestine [84, 85]. A structure where the branch is branched again has been characterized from hog gastric mucosa [86]. The branched PLNs linked to O-glycans and glycolipids in mucin producing tissues contain variably type 1 or type 2 units next to the branched galactose residue, Galβ1-3/4GlcNAcβ1-3(GlcNAcβ1-6)Gal-R. Human milk glycans contain also terminal type 1 units. Human PLNs with terminal type 1 units are discussed in section 1.3.

In addition to human erythrocytes, branched PLNs occur on rabbit and bovine erythrocytes. The glycolipids of rabbit erythrocytes contain 1-7 repeated branched LacNAc units [87-89], but a glycolipid structure with only one branched unit has been characterized from bovine erythrocytes [22]. In both species, the LacNAc-branches are capped with Galα1,3 residues and in rabbit, the terminal end is capped with Galα1,3 residue [87-89] and in bovine with Neu5Accα2,3 residue [22]. In contrast to human glycoporphin A, which lacks PLN, the O-glycans of bovine glycoporphin contain 1-2 LacNAc units in a linear form and branched LacNAcβ1-3(LacNAcβ1-6)LacNAc structures. The PLNs of bovine glycoporphin are also capped with Galα1,3 or Neu5Accα2,3 residues [90, 91].

Polyglycosylceramides with highly branched PLNs, in addition to those present in erythrocytes, have been found in human placenta [92], rabbit small intestine [93], and the pancreatic carcinoma cell line PANC-1 [94].

### 1.3 Expression of polylactosaminoglycans with terminal type 1 unit

Expression of the Galβ1-3GlcNAc (Lec) disaccharide, also called type 1 unit, at the nonreducing terminus of PLNs is restricted to the epithelia of the gastrointestinal tract, other mucin-producing tissues, and human milk. Usually type 1 units are substituted and present ABH, Lea, or Leb blood group antigens. Table 1 presents type 1 epitopes derived from human polylactosaminoglycans.
Table 1. The structures, origin, and carriers of type 1 epitopes characterized from human.

<table>
<thead>
<tr>
<th>Type 1 epitope structures</th>
<th>Origin</th>
<th>Carriers</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Gastric mucin&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Core 1 and 2 based O-glycans</td>
<td>[72, 73]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Gastric mucosa and epithelium</td>
<td>Glycolipids</td>
<td>[108, 109]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Colonic mucin</td>
<td>Core 3 based O-glycans</td>
<td>[74, 75]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Colonic epithelial cells</td>
<td>Mono- to tetra-antennary N-glycans (NFA-2)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>[97]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Small intestinal epithelial cells</td>
<td>Glycolipids</td>
<td>[98]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Foetal gastrointestinal (meconium) -mucins</td>
<td>Core 2 and 3 based O-glycans</td>
<td>[74, 75]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Foetal gastrointestinal (meconium) -mucosal cells</td>
<td>Glycolipids</td>
<td>[97]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Foetal gastrointestinal (meconium) - epithelial cells</td>
<td>Bi- to tetra-antennary N-glycans (NCA-2)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>[96]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Ovarian cyst mucins</td>
<td>Core 1, 2 and 6&lt;sup&gt;d&lt;/sup&gt; based O-glycans</td>
<td>[101, 102]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Amniotic mucins</td>
<td>Core 2 based O-glycans</td>
<td>[110]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Bronchial mucin</td>
<td>Core 2 and 4 based O-glycans</td>
<td>[111]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Milk</td>
<td>Core 1 and 2 based O-glycans of β4Gal-T</td>
<td>[112]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Milk</td>
<td>Core 2 based O-glycans of IgA</td>
<td>[76, 103]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Milk</td>
<td>Core 2 based O-glycans of IgA</td>
<td>[77]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Milk</td>
<td>Core 2 based O-glycans of IgA</td>
<td>[77]</td>
</tr>
<tr>
<td>Galβ1-GlcNAcβ1Galβ1-R</td>
<td>Milk</td>
<td>Core 1 based O-glycans of β4Gal-T</td>
<td>[5]</td>
</tr>
</tbody>
</table>

<sup>a</sup> The terminal units have been characterized as Galβ1-3/4GlcNAc, type 1 unit being predominant
<sup>b</sup> Normal faecal antigen-2
<sup>c</sup> Nonspecific cross-reacting antigen-2
<sup>d</sup> GlcNAcβ1-6GalNAc
Structural characterization of O-glycans from gastric mucins has indicated that the glycans are mainly branched and type 1 units are predominant at the distal ends [72, 73]. The O-glycans from colonic mucins are linear or branched derivatives of the linear structures. Type 1 and type 2 units are variably present at terminal ends [74, 75]. The largest O-glycans in gastric and colonic mucins contain four and three Lec/LacNAc units, respectively [72-75]. N-linked glycopeptides representing a major part of the glycans in small intestinal epithelial cells have been isolated and shown to bear Lec, LacNAc, or LacNAcβ1-3LacNAc sequences. Type 2 units are predominant, and on the average, one type 1 unit is present in one N-glycan. However, glycolipids from the same cells contain mainly type 1 determinants [95]. In adults, colon epithelial cells express normal faecal antigen-2 (NFA-2) and in fetuses nonspecific cross-reacting antigen (NCA-2), which are counterparts of carcinoembryogenic antigen (CEA) produced by colon adenocarcinomas. About 75% of the sugar chains of NFA-2 and 34% of NCA-2 contain type 1 units in their outer chain moieties, but only trace amounts are present in CEA. The side chains of NCA-2 and NFA-2 also contain repeating Lec/LacNAc units [96, 97]. Lactotetraosylceramide (Lecβ1-3Lacβ1-Cer) is one of the major glycolipid components in colon and small intestine [95, 98, 99], whereas the isomeric LacNAcβ1-3Lacβ1-Cer is present only in trace amounts [98, 99]. Larger glycolipids, composed of 9-10 monosaccharide units, have been found from small intestine. They are branched, fucosylated, and carry terminal type 1 units [100].

O-glycans from ovarian cyst mucins have terminal type 1 units and repeating type 1 units, Galβ1-3GlcNAcβ1-3Galβ1-3GlcNAcβ1-R [101, 102], which may be branched from Gal (underlined) [101]. O-glycans of respiratory mucins have been isolated and characterized from a patient suffering from bronchiectasis. Four common elements exist in these O-glycans: Lec/LacNAcβ1-3LacNAcβ1-6(GlcNAcβ1-3)GalNAc and Lec/LacNAcβ1-3(LacNAcβ1-6)Galβ1-3GalNAc [76, 103].

Free oligosaccharides in human milk contain linear and branched PLNs that are composed of type 2 units and have terminal type 1 or 2 units [104, 105]. The largest characterized structure is LacNAcβ1-6(Lecβ1-3)LacNAcβ1-6(Lecβ1-3)Lac where the first branch linked to the lactose is branched again [106]. Larger oligosaccharides exist in milk, but they have not been characterized in detail [107].

1.4 Biosynthesis of polylactosaminoglycans

The biosynthesis of oligosaccharides is mainly carried out by enzymes called glycosyltransferases in the Golgi apparatus. Golgi glycosyltransferases are type II transmembrane proteins. They have a short amino-terminal domain placed in the cytosol, a membrane-spanning domain, and stem- and carboxy-terminal catalytic domains in the Golgi lumen. Soluble glycosyltransferases in secretions and body fluids are derived from their membrane-associated forms by proteolytic cleavage at the stem region.
Glycosyltransferases transfer sugar moieties from activated donor sugars (e.g. UDP-GlcNAc, UDP-Gal) to specific acceptor molecules. Linear polyglycosamines are synthesized by the alternating action of β1,3-N-acetylglucosaminyltransferases (β3Gn-Ts) and β1,4-galactosyltransferases (β4Gal-Ts). The terminal galactose may alternatively be transferred by a β1,3-galactosyltransferase (β3Gal-T), forming a β1-3 linkage to the terminal GlcNAc of PLN. Blood group I β1,6-N-acetylglucosaminyltransferases (IGnTs) transfer GlcNAc in β1-6 linkage to linear PLNs forming branched PLNs.

During recent years, a number of new glycosyltransferase genes have been identified from human genomic sequence databases. β3Gn-T, β4Gal-T, β3Gal-T, and IGnT families have been shown to include several members. Individual members may exhibit differences in substrate specificity, kinetic parameters, and expression patterns, and their genes may be differently regulated. For references of the glycosyltransferase families, see [113, 114].

### 1.4.1 β1,3-N-acetylglucosaminyltransferases

β1,3-N-acetylglucosaminyltransferase (β3Gn-T) activities are found in several cells, tissues, and body fluids of humans [115-120] and other animals [121-124]. The β3Gn-T activities have shown to elongate free Lac(NAc) [115-117, 124] and different types of oligosaccharide acceptors [115, 116, 123, 124], N-linked glycoproteins [116, 121], and glycolipids [118-120, 122] that contain terminal Lac(NAc) unit. Human serum contains activities for all these substrates, and in addition, for desialylated and defucosylated O-glycans of mucins and keratan sulfate [116].

The partially purified β3Gn-T from Novikoff tumor cell ascites fluid elongates LacNAc and different oligosaccharide acceptors that present LacNAc-Man sequences. Optimal activity is encountered with the Galβ1-4GlcNAcβ1-2(Galβ1-4GlcNAcβ1-6)Man pentasaccharide acceptor. Asialo α1-acid glycoprotein, containing the pentasaccharide structure on N-linked glycans, is a much better acceptor than asialofetuin and asialotransferrin, which are N-glycoproteins deficient of that structure [125]. The β3Gn-T purified from calf serum has similar specificity for N-glycoproteins [126]. The partially purified human serum β3Gn-Ts transfer GlcNAc more efficiently to LacNAc and asialo α1-acid glycoprotein than to Lac [127].

Some β3Gn-T activities [117, 118, 124] and the β3Gn-Ts purified from human serum [127] transfer GlcNAc to acceptors that have terminal type 1 unit, Galβ1-3GlcNAc, but the transfer rate is much lower than the transfer rate to Galβ1-4GlcNAc. The β3Gn-T from calf serum has no activity to type 1 acceptor [126]. α1,2-fucosylation of terminal Gal or α1,3-fucosylation of subterminal GlcNAc blocks the β3Gn-T reaction [116-118, 124, 126].

Seven human β3Gn-Ts (iGnT, β3Gn-T2-T7) have been identified to date and six of them (iGnT, β3Gn-T2-T6) have been cloned and characterized [128-132]. The i-extension enzyme (iGnT) elongates polyglycosamines. The iGnT transcript is ubiquitously expressed in
various adult tissues, and especially highly in adult brain and in fetal brain and kidney [128]. β3Gn-T2-T7 have been found to be structurally similar to the members of the β1,3-galactosyltransferase family, but differ from iGnT [129-132]. The β3Gn-T2 and β3Gn-T5 transcripts are ubiquitously expressed in tissues and cells, whereas the expression of β3Gn-T3 is restricted to colon, jejunum, stomach, esophagus, placenta, and trachea, and β3Gn-T4 is mainly expressed in the brain [129, 130]. iGnT, β3Gn-T2, and -T4 prefer LacNAcβ1-3Lac to Lecβ1-3Lac, whereas β3Gn-T3 utilizes both substrates at a comparable rate [128, 129]. Transfection of Namalwa KJM-1 cells with β3Gn-T2, -T3, or -T4 cDNAs has indicated that these enzymes are able to initiate and elongate polylactosamine synthesis in vivo [129]. β3Gn-T5 has been identified as Lc3Cer synthase as it transfers GlcNAc to lactosylceramide. β3Gn-T2, -T3, and -T4 do not react with LacCer [130]. The β3Gn-T6 transcript is mainly expressed in stomach, colon, and small intestine. β3Gn-T6 transfers GlcNAc to GalNAc forming the core 3 structure of O-glycans [131].

1.4.2 β1,4-galactosyltransferases

Early studies have shown that bovine milk β1,4-galactosyltransferase (β4Gal-T) transfers galactose to GlcNAc, but in the presence of α-lactalbumin, the transfer to GlcNAc is inhibited and the enzyme preferentially acts as lactose synthase [133, 134]. Acceptor specificity studies have shown that bovine milk β4Gal-T reacts preferentially to the GlcNAcβ1,6-branch of the acceptors GlcNAcβ1-3(GlcNAcβ1-6)Gal-R, where R is H, GlcNAc, Glc, Glc-OMe, or GlcNAcβ1-6(Galβ1-3)GalNAc-ol [135-138].

β4Gal-T1 is abundant in bovine and human milk in a soluble form, and is the first galactosyltransferase for which the corresponding cDNA has been isolated [139]. To date, seven members of the human β4Gal-T family (β4Gal-T1-T7) have been identified, reviewed in [140]. β4Gal-T2 is affected by α-lactalbumin in a similar manner to bovine milk β4Gal-T (β4Gal-T1) [141]. β4Gal-T1 has been shown to elongate more efficiently linear PLN chains and the GlcNAcβ1,6-branch than β4Gal-T2-T5 [142]. β4Gal-T1 is also most efficient in adding galactose to GlcNAcβ1-6/2Manα1-6Manβ1-octyl, and in addition, it prefers β1,6-linked GlcNAc [143, 144]. β4Gal-T1 galactosylates most efficiently O-glycan core 4 branch but poorly core 2 branch [143, 145]. Core 2 branch is utilized by β4Gal-T4 [143]. O-glycan core 2 and core 6 (GlcNAcβ1-6GalNAc) have been shown to be the best substrates for β4Gal-T5 [146]. β4Gal-T3 and -T4 catalyze the transfer of galactose to lactoseries glycolipids (Lc3Cer, nLc5Cer). Both enzymes have shown a strong preference for Lc3Cer over nLc5Cer [141, 147]. β4Gal-T6 synthesizes lactosylceramide [148], and β4Gal-T7 galactosylates xylose attached to proteoglycan core proteins [149, 150].

The expression of human β4Gal-T1-T6 mRNAs has been comparatively analyzed [151]. β4Gal-T1 and β4Gal-T3 are widely expressed, but only β4Gal-T3 is expressed in high levels in the brain. β4Gal-T4 and -T5 are also widely expressed but in lower levels than
β4Gal-T1 and -T3. Expression of β4Gal-T2 is restricted to fetal brain and adult heart, muscle, and pancreas. β4Gal-T6 is expressed only in adult brain. The fact that only the β4Gal-T1 gene is expressed in murine lactating mammary glands indicates that β4Gal-T1 but not β4Gal-T2 is responsible for lactose synthesis in lactating mammary glands.

1.4.3 β1,3-galactosyltransferases

A β1,3-galactosyltransferase (β3Gal-T) has been first detected from pig trachea [152]. The purified enzyme β1,3-galactosylated GlcNAc, GlcNAcβ1-3Lac, O-glycan core 3, and lactotriosylceramide (Lc3Cer) [153]. The β3Gal-T activity present in Colo 205 cells has also been shown to β1,3-galactosylate GlcNAc [154] and Lc3Cer [155].

A family of six human β1,3-galactosyltransferases (β3Gal-T1-T6) have been cloned to date [156-159]. Three of them, β3Gal-T1, β3Gal-T2, and β3Gal-T5, catalyze the synthesis of the type 1 structure [156-158]. β3Gal-T5, originally cloned from Colo 205 cells, has activity toward GlcNAc, Lc3Cer, O-glycan core 3, and also to the terminal GalNAc of globoside (Gb4, GalNAcβ1-3Galα1-4Galβ1-4Glcβ1-Cer) [160-162]. β3Gal-T1 and β3Gal-T2 react with GlcNAc, Lc3Cer, and nLc3Cer, but not with Gb4 [156, 157]. All three β3Gal-Ts are active in some extent on GlcNAcβ1-2/4/6Man substrates [157, 160, 162], but at least β3Gal-T5 prefers the GlcNAcβ1-3Galβ1-β (R=OMe or Glcβ1-Cer) acceptor sequence [162].

Studies of β3Gal-Ts tissue distribution have shown that β3Gal-T1 and β3Gal-T2 are expressed in brain and at least in some extent in colon, and in addition β3Gal-T2 is expressed in heart [156, 157, 162]. Substantial β3Gal-T5 expression has been detected in colon, stomach, jejunum, and pancreas [158].

1.4.4 β1,6-N-acetylglucosaminyltransferases

β1,6-N-acetylglucosaminyltransferases (β1,6GnTs) are classified according to their substrate specificities. Blood group I β1,6-N-acetylglucosaminyltransferases (IGnTs) generate GlcNAcβ1,6-branches on linear poly lactosamines. O-glycan core branching enzymes convert core 1 to core 2 structure [Galβ1-3GalNAc → Galβ1-3(GlcNAcβ1-6)GalNAc] and core 3 to core 4 structure [GlcNAcβ1-3GalNAc → GlcNAcβ1-3(GlcNAcβ1-6)GalNAc] and are called core 2 and core 4 β1,6-N-acetylglucosaminyltransferases (C2GnT, C4GnT), respectively. β1,6-N-acetylglucosaminyltransferase V (GnTV) synthesizes a GlcNAcβ1,6-branch to GlcNAcβ1-2Manβ1-6Man in N-glycan core. None of the β1,6GnTs have requirements for Mn²⁺ ions, whereas β3Gn-Ts require divalent metal ions for full activity.

Two types of IGnTs with different acceptor site specificities have been characterized. The distally acting IGnT (dIGnT) acts on the subterminal Gal residue of the PLN chain. The centrally acting IGnT (cIGnT) transfers GlcNAc to the midchain Gal residues of completed or
growing PLN chains (Figure 4). A $\beta_{1,6}$GnT activity that transfers GlcNAc to terminal galactose (tIGnT) has been reported [121, 122, 163, 164], but this type of activity has not been detected in any of the purified or cloned enzymes.

1.4.4.1 Distally acting $\beta_{1,6}$-N-acetylglucosaminyltransferases

Distally acting IGnT activity has been detected in hog gastric mucosa [165, 166], in Novikoff ascites tumor cells [164], and in a partially purified rat intestine enzyme [123]. All characterized dIGnTs transfer GlcNAc to terminal GlcNAc$\beta_1$-3Gal units of oligosaccharides, and the substitution of the terminal GlcNAc by galactose blocks the transfer reaction [123, 165]. Rat intestine and hog gastric dIGnTs have been shown to branch only the subterminal Gal (underlined) in GlcNAc$\beta_1$-3Gal$\beta_1$-4GlcNAc$\beta_1$-3Gal$\beta_1$-4Glc-PA/GlcNAc acceptors, indicating that dIGnTs do not branch midchain galactoses [123, 167]. Enzymes purified to apparent homogeneity from bovine trachea [168] and rat small intestine [169] possess three different but related activities: core 2 and 4 branching activities, and dIGnT activity (C2GnT/C4GnT/dIGnT activities).

Three human C2GnT enzymes have been cloned: C2GnT-leukocyte type (C2GnT-L, C2GnT1) [170], C2GnT-mucin type (C2GnT-M, C2GnT2) [171, 172], and C2GnT3 [173]. C2GnT1 and C2GnT3 have only C2GnT activity [170, 171, 173], whereas C2GnT2 has C2GnT/C4GnT/dIGnT activities [171, 172]. C2GnT2 transcripts are heavily expressed in mucin producing tissues [171]. An ortholog of C2GnT2 that also has C2GnT/C4GnT/dIGnT activities has been identified in bovine herpes virus type 4 (pBORFF3-4, v-C2GnT2) [174]. The cDNA corresponding to the purified rat small intestine dIGnT has been cloned, and the rat dIGnT has been indicated to be an ortholog of human and viral C2GnT2s [169]. Human and viral C2GnT2s exhibit the strongest activities as C2GnTs, moderate activities as C4GnTs, and weakest activities as dIGnTs [171, 174]. In contrast, rat and bovine dIGnTs contain comparable amounts of C2GnT and C4GnT activities and lower but significant amounts of dIGnT activity [168, 169]. A weak dIGnT activity has been detected in human cIGnT (IGnT1) [142, 171] and C2GnT3 [173], and a very weak cIGnT activity has been observed in human and viral C2GnT2s [171, 174].

Figure 4. Site-specificity of distally and centrally acting blood group I $\beta_{1,6}$-N-acetylglucosaminyltransferases.
1.4.4.2 Centrally acting β1,6-N-acetylglucosaminyltransferases

Centrally acting IGnT activity has been detected for the first time in human serum [175], in rat tissues and serum [123, 176], and in a partially purified rat intestine enzyme [123]. These cIGnTs have been demonstrated to form GlcNAcβ1,6-branch to the internal galactose residue in Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc/Glc acceptors. The cIGnT activity present in rat serum and several other mammalian sera has also been shown to form multiple branches to linear PLNs (publication I, this thesis). A cIGnT enzyme, which also generates multiple branches to PLNs, has been purified from hog small intestine [177].

The cDNA encoding IGnT has been first cloned from human embryonal carcinoma cells, PA1 [178]. A fusion protein with glutathionesulotransferase (GST)-IGnT has been functionally expressed and purified (publication III, this thesis). The analysis of the substrate specificity of the GST-IGnT using free polylactosamines as acceptors has indicated that the enzyme possesses cIGnT type activity, but not dIGnT type activity (publication III, this thesis), which is in accordance with the activity present in PA1 cells [179]. However, some dIGnT activity has been observed in analogous recombinant protein A-IGnT when the polylactosamine acceptor is linked to a Manα1-6Man derivative [142, 171]. Also very weak C2GnT and C4GnT activities have been detected in protein A-IGnT [171].

The first IGnT has been designated IGnT1 (IGnT6 in publication III) after demonstration that the human I locus codes also two other IGnT forms, designated IGnT2 and IGnT3 [180]. The IGnT1-T3 forms have also been called IGnTB, IGnTA, and IGnTC, respectively [181]. IGnT1 and IGnT2 transcripts are expressed in many adult and fetal tissues ubiquitously, but IGnT1 is expressed at lower levels than IGnT2. IGnT3 is strongly expressed in bone marrow, heart, stomach, and small intestine. IGnT3 has been indicated to be the most feasible candidate for the expression of the blood group I antigen in erythrocytes, because the expression of the IGnT3 transcript is markedly increased during erythrocyte differentiation, and IGnT3 is the most strongly expressed IGnT in erythrocytes [180]. In mammary gland, the highest transcript level has been detected for IGnT2, and this IGnT may be largely responsible for the synthesis of soluble milk I-antigens [180]. All three IGnT forms consist of three exons and possess a different exon 1 but identical exon 2 and 3 coding regions. Mutations found in exons 2 and 3 abolish the branching activity of all three IGnTs and cause the i phenotype in erythrocytes [180-182].

The mouse I locus codes also three transcript forms, designated IGnTA, IGnTB, and IGnTC, which have three exons and share the second and third exons [183-185]. The mouse and human IGnT proteins have homologous amino acid sequences, and the pairs mIGnTB and hIGnTA (IGnT2), mIGnTA and hIGnTB (IGnT1), and mIGnTC and hIGnTC (IGnT3), show the highest identities [185].

Table 2 shows the relative reactivities of different acceptors with the comprehensively characterized cIGnTs. The relative reactivities with rat serum cIGnT are shown in Table 4, section 4.1. Generally, a complete LacNAc residue bound to the acceptor galactose residue is
required for cIGnT activity. The purified hog small intestine cIGnT and the cloned cIGnTs are capable to synthesize multiply branched PLNs. The partially purified rat intestine cIGnT differs from the other cIGnTs by reacting poorly with the reducing end galactose (underlined) in GlcNAcβ1-3LacNAcβ1-3Gal-β-R and LacNAcβ1-3LacNAcβ1-3Gal-β-R acceptors.

Table 2. Relative reactivities of different acceptor oligosaccharides with purified and cloned cIGnTs.

<table>
<thead>
<tr>
<th>Acceptor</th>
<th>cIGnT purified from rat intestine</th>
<th>cIGnT purified from hog small intestine</th>
<th>IGnT1 cloned from human PA1 cells</th>
<th>IGnT1 cloned from the mouse</th>
<th>IGnTA cloned from the mouse</th>
<th>IGnTB cloned from the mouse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galβ1-4R</td>
<td>-</td>
<td>-</td>
<td>(&lt;1)</td>
<td>(&lt;1)</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>GlcNAcβ1-3Galβ1-4R</td>
<td>7 (100)</td>
<td>6 (92)</td>
<td>6 (100)</td>
<td>100 (100)</td>
<td>2 (92)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>LNβ1-3Galβ1-4R</td>
<td>100 (92)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>LNBβ1-3Galβ1-4R</td>
<td>- (-)</td>
<td>- (-)</td>
<td>4 (6)</td>
<td>6 (6)</td>
<td>10 (12)</td>
<td>20 (20)</td>
</tr>
<tr>
<td>GlcNAcβ1-3LNβ1-3Galβ1-4R</td>
<td>9 (50)</td>
<td>50 (59)</td>
<td>4 (6)</td>
<td>6 (6)</td>
<td>10 (12)</td>
<td>20 (20)</td>
</tr>
<tr>
<td>GlcNAcβ1-3LNβ1-3Galβ1-4R</td>
<td>11 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>GlcNAcβ1-3LNβ1-3Galβ1-4R</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>LNβ1-3LNβ1-3Galβ1-4R</td>
<td>6 (21)</td>
<td>21 (18)</td>
<td>6 (10)</td>
<td>10 (10)</td>
<td>18 (18)</td>
<td>18 (18)</td>
</tr>
<tr>
<td>LNβ1-3LNβ1-3Galβ1-4R</td>
<td>121 (194)</td>
<td>12 (18)</td>
<td>46 (7)</td>
<td>170 (194)</td>
<td>181 (164)</td>
<td>181 (164)</td>
</tr>
<tr>
<td>LNβ1-3LNβ1-3Galβ1-4R</td>
<td>7 (10)</td>
<td>10 (10)</td>
<td>6 (10)</td>
<td>10 (10)</td>
<td>18 (18)</td>
<td>20 (20)</td>
</tr>
<tr>
<td>LNBβ1-3LNβ1-3Galβ1-4R</td>
<td>256 (33)</td>
<td>148 (148)</td>
<td>48 (33)</td>
<td>148 (148)</td>
<td>33 (33)</td>
<td>33 (33)</td>
</tr>
<tr>
<td>Neu5Acα2-6LNβ1-3Galβ1-4R</td>
<td>-</td>
<td>-</td>
<td>48 (33)</td>
<td>148 (148)</td>
<td>33 (33)</td>
<td>33 (33)</td>
</tr>
<tr>
<td>Neu5Acα2-3LNβ1-3Galβ1-4R</td>
<td>3 (2)</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>2 (2)</td>
<td>2 (2)</td>
</tr>
<tr>
<td>Fucα1-3LNβ1-3Galβ1-4R</td>
<td>35 (123)</td>
<td>123 (123)</td>
<td>35 (123)</td>
<td>123 (123)</td>
<td>123 (123)</td>
<td>123 (123)</td>
</tr>
<tr>
<td>GalNAc-pNP</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Galβ1-3GalNAcβ1-4Galβ1-4R</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>GlcNAcβ1-3/6GalNAcα-pNP</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
<tr>
<td>GlcNAcβ1-3(Galβ1-6)GalNAcα-pNP</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
<td>&lt;1</td>
</tr>
</tbody>
</table>

Reference [123] [177] [142] [184] [184] [184]

a)The abbreviations used are: LN=Galβ1-4GlcNAc, LNB=Galβ1-3GlcNAc, pNP=para-nitrophenol. Underlines indicate the reaction site of cIGnT. b)R=Glc-PA (PA, 2-aminopyridine), R=Glc for the values in brackets. c)R=GlcNAc, R=Glc-PA for the values in brackets. d)Calculated from data presented in [177]. e)R=GlcNAcβ1-6Manα1-6Manβ1-octyl. f)R=Glc, R=GlcNAc for the values in brackets. g)Relative amounts of products were not determined exactly, but it was concluded that in major monobranched product the branch occur in middle Gal. -: activity not detected, blank: information not available.

The cIGnT activity of rat serum is inhibited if either of the adjoining GlcNAc residues of the branch accepting Gal is α1,3-fucosylated [186]. Similar specificity is present in human IGnT1 (publication III, this thesis) and mouse IGnTA and IGnTB [184]. Terminal α2,3-linked sialic acid abolishes the activity of the rat intestine enzyme [123] and diminishes the activity of mouse IGnTA and IGnTB [184]. Surprisingly, Neu5Acα2,6-substituted acceptors are efficiently branched by IGnTA and IGnTB [184]. Rat intestine and hog gastric cIGnTs show no reactivity, and IGnTA and IGnTB react weakly with Galβ1-3GlcNAcβ1-3Galβ1-4R.
acceptors that contain a terminal type 1 unit [123, 177, 184]. Branching of internal GalNAcβ1-4GlcNAc determinants in polylactosamines has been demonstrated with rat serum and IGnT1. However, the IGnT1 activity to internal GalNAcβ1-4GlcNAc unit is quite low as compared to LacNAc unit [187].

1.5 Structural analysis of polylactosamines by NMR spectroscopy

Nuclear magnetic resonance (NMR) is a phenomenon that occurs when the nuclei of certain atoms are immersed in a static magnetic field and exposed to radio frequency electromagnetic radiation [188, 189]. Nuclei, such as \(^1\)H and \(^{13}\)C, possess a property called spin that generates a nuclear magnetic moment. In a magnetic field, magnetic moments are aligned with the field (lower energy) or against the field (higher energy). The magnetic moments at the lower energy level are excited into the higher level with electromagnetic radiation. The frequency of radiation needed is determined by the difference in energy between the energy levels. The magnetic field at the nucleus is not equal to the applied magnetic field; electrons around the nucleus shield it from the applied field. Therefore, nuclei that are chemically nonequivalent experience the external field slightly differently and resonate at different frequencies. The chemical shift of a nucleus is the difference between the resonance frequency of the nucleus and a standard, relative to the standard. This quantity is reported in parts per million (ppm) and given the symbol delta, \(\delta\). Nuclei that are connected through chemical bonds, typically less than or equal to three bond lengths, effect to each other's spin states. This effect is called spin-spin coupling or J coupling, and it shows up in the NMR spectrum as splitting of lines when the nuclei are nonequivalent i.e. have different chemical shifts. The distance between two split absorption lines is called the J coupling constant or the spin-spin splitting constant (Hz) and is a measure of the magnetic interaction between two nuclei. The \(^1\)H signal intensities i.e. the integrals of the signal curves in the spectrum give the ratios of the protons in the molecule. For \(^{13}\)C the signal intensities are often not given. This is because the natural abundance and sensitivity of \(^{13}\)C is much lower than \(^1\)H (1.11% and 99.98%, respectively), and detection methods used to improve the sensitivity of \(^{13}\)C give inaccurate signal intensities. The next sections deal mainly with the NMR techniques used in this study.

1.5.1 One-dimensional \(^1\)H NMR spectroscopy

A 1D \(^1\)H NMR spectrum can be recorded from a 1-10 nmol oligosaccharide sample. The oligosaccharides are usually dissolved in D_2O for recording. In D_2O, the OH protons and the NH protons of N-acetyl groups exchange to deuterons and give no signal. The residual HDO in the solution gives water signal that can be suppressed. The WEFT (water eliminated Fourier transform) technique [190] is well suited for water suppression in polylactosamine
NMR spectroscopy. Acetone is usually used as an internal standard, and the signals are referenced to the acetone signal at 2.225 ppm.

Most of the protons of the oligosaccharides resonate between $\delta=3.4-3.9$ ppm. This region of overlapping resonances is called the bulk-region, and these signals usually cannot be assigned. The protons not resonating in this region are referred to as structural reporter groups. In many cases, the assignment of these signals and their comparison to reference compounds is sufficient for the identification of the oligosaccharide structure. A proton in the equatorial position at a certain carbon in a pyranose ring resonates at a $\sim0.5$ ppm lower field than the corresponding axial proton. Therefore a H-1 involved in an $\alpha$-glycosidic linkage resonates at $\sim4.9-5.6$ ppm and in $\beta$-glycosidic linkage at $4.3-4.8$ ppm. The coupling constant $^{3}J_{1,2}$ for axial H-1 and H-2 protons in $\beta$-gluco- and $\beta$-galacto-configuration is $\sim7-8$ Hz. Instead, the $^{3}J_{1,2}$ for equatorial H-1 and axial H-2 protons in $\alpha$-gluco-, $\alpha$-galacto-, and $\beta$-manno-configuration, or for equatorial H-1 and H-2 protons in $\alpha$-manno-configuration, is $3-4$ Hz. Thus $\alpha$- and $\beta$-glycosidic linkages of Glc and Gal, but not for example Man, can also be inferred from $^{3}J_{1,2}$ [191].

The expansion of the 1D $^1$H NMR spectrum of the pentasaccharide Gal$\beta$1-4GlcNAc$\beta$1-3(GlcNAc$\beta$1-6)Gal$\beta$1-4GlcNAc in Figure 5 shows the typical reporter group signals of branched poly lactosamines (published in [176, 192]). The presence of the GlcNAc$\beta$1,6-linkage, not $\beta$1,2-linkage, in this pentasaccharide was firmly established by a series of one- and two-dimensional NMR experiments [192].

![Figure 5](image)

**Figure 5.** Expansion of $^1$H NMR spectrum of the branched pentasaccharide. The two H-1 signals of GlcNAc 3 arise from the $\alpha$- and $\beta$-pyranosic forms of the oligosaccharide. The signals marked by an asterisk arise from the reducing end ManNAc epimer.
Different monosaccharide residues linked to polylactosamines can also be determined by 1D $^1$H NMR spectroscopy. For example, the constituents of sLex, Fucα1,3 and Neu5Acα2,3 residues, have characteristic structural reporter group signals: Fuc H-1, H-5, and H-6, and Neu5Ac H-3 (both axial and equatorial) [193]. The H-1 protons of terminal type 1 unit, Galβ1-3GlcNAc, resonate at the same regions as the H-1’s of Galβ1-4GlcNAc, but the H-1 of Gal resonates at upper field and the H-1 of GlcNAc at lower field in type 1 unit (publication IV, and references therein).

1.5.2 Two-dimensional NMR spectroscopy

The structural identification of an oligosaccharide is not possible by 1D $^1$H NMR spectroscopy in cases where not all structural reporter groups can be assigned, or the assignments do not unambiguously determine the structure. By the use of two-dimensional $^1$H-$^1$H NMR techniques, TOCSY (total correlation spectroscopy) [194] and COSY (correlated spectroscopy) [195, 196] or DQFCOSY (double quantum filtered COSY) [197], almost all proton signals can be assigned. The completely assigned proton spectrum determines the linkage position to a certain ring carbon, as the proton resonance at that position is shifted downfield [191]. However, other protons nearby may also be shifted, and therefore these methods alone do not firmly establish the structure. To identify the structure unambiguously the two-dimensional $^1$H-$^{13}$C NMR techniques are used. The HMQC (heteronuclear multiple quantum coherence) [198, 199] and HSQC (heteronuclear single quantum coherence) [200] spectra indicate carbon-proton linkages i.e. the $^{13}$C signals can be assigned by these methods. The downfield shift (5-10 ppm) of a $^{13}$C resonance indicates a glycosidic linkage to this carbon [191]. HMBC (heteronuclear multiple bond correlation) determines $^1$H-$^{13}$C couplings through two or three bonds [200, 201]. HMBC technique is the best choice when enough sample (~0.5 µmol in the nanoprobe) is available because both the sequence and substitution positions can be determined unambiguously by this method.

1.6 Biological functions that involve polylactosaminoglycans

Selectins, a class of C-type (Ca$^{2+}$-dependent) lectins, and galectins (formerly called S-type lectins) mediate cell adhesion by recognizing carbohydrate ligands. Selectins bind to capping groups of (poly)lactosamines, whereas galectins bind β-galactosides. Sperm-egg binding in fertilization involves carbohydrates. Many microorganisms exploit host cell-surface glycoconjugates as receptors for cell adhesion. Some pathogens have binding activity to (poly)lactosamines and/or their terminal substituents.
1.6.1. Selectin mediated cell adhesion

The reaction cascade that leads to the extravasation of leukocytes from the bloodstream into inflamed tissue and lymphocyte homing to lymph nodes is initiated by selectin-carbohydrate interactions, reviewed in [202]. Inflammatory stimuli induce E- and P-selectins to appear on the surface of the vascular endothelium, and activation of platelets releases the granule-stored P-selectin to the surface of platelets. E- and P-selectins bind to their ligands on leukocytes, whereas L-selectin, which is constitutively expressed on leukocytes, binds to glycans on endothelial cells and on other leukocytes. The selectins mediate the tethering and rolling of leukocytes along the vessel wall, which is followed by firm adhesion, and penetration of the cells through the vascular wall. In a similar fashion, L-selectin mediates the initial attachment of lymphocytes to lymph node high endothelial venules (HEV) during lymphocyte recirculation.

The HEV of lymph nodes express several mucin-like glycoprotein ligands for L-selectin [203]. These ligands need to be sialylated, fucosylated, and sulfated for optimal binding to L-selectin [203-205]. O-glycan structures of mouse glycosylation-dependent cell adhesion molecule-1 (GlyCAM-1), the best characterized L-selectin-ligand, have been analyzed in detail. The simplest monosulfated O-glycans contain 6-sulfo-sLex and 6'-sulfo-sLex epitopes on core 2 (Figure 6). These glycans account for less than 25% of the O-linked chains and the rest of the chains are more complex structures and may present multiple sLex and 6/6'-sulfo-sLex epitopes [206]. Oligosaccharide analysis of MECA-79 positive (Galβ1-4(6-sulfo)GlcNAcβ1-3Galβ1-3GalNAc) GlyCAM-1 has indicated that O-glycans bear twin 6-sulfo-sLex on biantennary core 2 in addition to 6-sulfo-sLex on core 2 (Figure 6) [207]. Lymphocyte homing is markedly reduced in mice lacking a HEV-restricted GlcNAc-6-O-sulfotransferase (designated HEC-GlcNAc6ST or LSST) indicating that GlcNAc-6-sulfation is significant for L-selectin ligand activity [205].

![Figure 6. Three O-glycans on GlyCAM-1.](image-url)
CD34 is another L-selectin glycoprotein ligand on HEV. CD34 on human tonsillar HEV carry a sulfated and fucosylated O-glycan, which putatively has biantennary core 2 structure where 6-sulfo-sLex is linked to core galactose [208]. L-selectin binds to sLex or its sulfated derivatives with a relatively low affinity [203, 209]. High affinity L-selectin ligands may be generated by O-glycans, which are presented on the protein backbone in unique spacing and/or clustered combinations [209, 210].

Inflammation in rejecting solid organ transplants is characterized by heavy infiltration of lymphocytes into the graft [211, 212]. During the acute rejection of rat kidney or cardiac allograft, the graft endothelium begins to express sialyl Lewis x oligosaccharides, which support L-selectin-dependent lymphocyte adhesion [213, 214]. The expression of L-selectin ligands, sLex and sulfo-sLex, is also induced in rejecting human allografts [215, 216] and in various inflammatory diseases [217, 218]. The expression patterns of sLex and sulfo-sLex are different between different inflamed tissues, suggesting that every organ expresses a specific ‘zip code’ which regulates leukocyte traffic to that given organ [218].

Enzymatically synthesized multivalent sLex polylactosaminoglycans at very low concentrations inhibit L-selectin-mediated lymphocyte binding to the endothelium of rejecting organ transplants and to lymph nodes of rats [214, 219-223]. Linear and branched tetravalent sLex glycans (Figure 7) are effective in inhibiting lymphocyte adhesion to activated cardiac epithelium, the IC50-values being around 1 nM in an ex vivo Stamper-Woodruff binding assay [221]. Multivalency enhances the inhibitory effect of sLex glycans; multivalent sLex glycans are much better inhibitors than monovalent sLex tetrasaccharide [214, 220, 222]. A divalent sLexLex glycan (Figure 7) inhibits lymphocyte binding to cardiac allograft with an IC50 value of 5 nM and to lymph node endothelium with a ten times higher IC50 value [223]. This suggests that divalent sLexLex glycan and other L-selectin oligosaccharide antagonists may inhibit lymphocyte traffic to the site of inflammation without significantly altering the normal lymphocyte recirculation via lymph nodes.

![Image of enzymatically synthesized biologically active oligosaccharides]

**Figure 7.** Examples of enzymatically synthesized biologically active oligosaccharides.

The best characterized ligand for P-selectin is P-selectin glycoprotein ligand-1 (PSGL-1), which also interacts with L- and E-selectins. O-glycans on PSGL-1 from the promyelocytic cell line HL-60 contain two species of fucosylated and sialylated structures: one having sLex-Lex-Lex on core 2 branch, and the other having sLex on core 2 branch and Neu5Aca2,3 residue on core galactose [11]. The latter structure at Thr-57 of the protein
backbone and sulfation of three tyrosine residues (Tyr-46, -48, and -51) are necessary for high affinity binding of PSGL-1 to P-selectin [224-226]. The optimal L-selectin binding to PSGL-1 and leukocyte rolling requires sialylated and fucosylated core 2 O-glycans attached to Thr-57 and sulfation of Tyr-46 and -51 [227]. E-selectin binding to PSGL-1 does not require sulfation, but is dependent of sialylated and fucosylated, likely core 2 based O-glycans [228]. In contrast, N-linked glycans with SLex-like structures are essential for E-selectin binding to E-selectin ligand-1 (ESL-1) [229-231].

1.6.2 Galectin mediated cell adhesion

Galectins are widely distributed in all living organisms, and from mammals alone, fourteen galectins have been identified [232, 233]. Some galectins (Gals), such as Gal-1 and Gal-3, are expressed in many tissues and cell types. Others are restricted to specific tissues, such as Gal-2 and Gal-7, which are expressed specifically in gastrointestinal tract and stratified epithelia, respectively [233]. The galectins are found mainly in the cytoplasm, but several are secreted from cells by an unusual mechanism and interact with appropriately glycosylated protein ligands at the cell surface or within the extracellular matrix [234]. For example, Gal-1 and -3 bind to basement membrane laminin and integrins [235]. The galectins bind multivalently and are capable of cross-linking ligands [236]. Many essential functions, where galectins play roles in regulating cell-cell and cell-matrix adhesion, have been implicated. These include cell motility, growth, differentiation, and apoptosis, development and tumor metastasis [235, 237, 238]. In addition, leukocyte turnover that is not associated with apoptosis has been shown to involve Gal-1; the binding of Gal-1 to leukocytes induces their phagocytic recognition [239].

Individual galectins differ significantly in their recognition of galactosyl residues within oligosaccharides. The oligosaccharide specificity of 13 galectins has been studied and the previous data reviewed in [240]. The galectins recognize both type 2 (Galβ1-4GlcNAc) and type 1 (Galβ1-3GlcNAc) disaccharides. Three OH-groups are required for galectin binding, i.e. 4-OH and 6-OH of Gal, and 3-OH or 4-OH of Glc(NAc) depending on the glycosidic linkage [240, 241]. Substitutions at 4-OH and 6-OH of Gal, for example with Neu5Acα2,6, abolish galectin binding. Le-type structures, such as Lex, Lea, and Leb, are not recognized by galectins either. Gal-3 has the highest affinity to type 1 (Lecβ1-3Lac) and type 2 (LacNAcβ1-3Lac) saccharides. Gal-2 and Gal-7 strongly prefer type 1 saccharides (Lecβ1-3Lac) to type 2 (LacNAcβ1-3Lac). Increase in the number of the repeating LacNAc units (1 to 3 units) enhances the affinity of Gal-3, Gal-7, Gal-8, and Gal-9. In contrast, Gal-1 does not show any particular preference for repeated N-acetyllactosamine structures, although it has been previously identified as a binding protein of laminin (see section 1.2.4). Many of the galectins show increasing affinity for N-glycans when the branching number increases from mono- to bi-, tri- and tetraantennary glycans. The ‘glycoside clustering effect’ may be in some extent responsible for the enhanced affinity. LacNAcβ1,6-branches in polylactosamines may
be recognized by galectins. Galectin-3 has been shown to bind glycopeptides derived from adult erythrocytes better than those derived from cord erythrocytes [241].

1.6.3 Fertilization

Mammalian egg cells are surrounded by an extracellular matrix called zona pellucida (ZP). In the fertilization process, sperm initially binds in a species-specific manner to the surface of ZP. The binding induces the sperm acrosome reaction, or cellular exocytosis, which leads to the exposure of the inner acrosomal membrane. Then the acrosome-reacted sperm penetrates through the ZP and fuses with the egg cell, reviewed in [242]. The mammalian sperm-egg interaction is most thoroughly studied in the mouse. Mouse ZP, like all mammalian ZPs, contains three major families of glycoproteins (ZP1, ZP2, and ZP3). The sperm recognizes and binds to glycans on ZP3 and undergoes acrosome reaction. The acrosome-reacted sperm likely binds to ZP2 and remains bound to ZP. Penetration through the ZP is probably achieved by a combination of sperm motility and enzymatic hydrolysis [242].

Mouse ZP2 and ZP3 carry tri- and tetraantennary N-glycans with variable amounts of lactosamine repeats. Considerable amount of the glycans are acidic (95%): most are sialylated (80%) and a residual amount is sulfated. The desialylated N-glycans have terminal GlcNAcβ1-3, Galβ1-4, GalNAcβ1-, and Galα1-3 units, in a ratio of 63:31:4:2. The O-glycans of ZP3 have a trisaccharide structure GlcNAc-Galβ1-3GalNAc, reviewed in [243, 244]. Several studies have proposed the precise glycan sequences on ZP3 required to sperm binding, but the results are controversial. Candidates that participate in binding are glycans terminated either with α1-3 or β1-4 linked Gal, or with β1-3 linked GlcNAc, reviewed in [243, 244]. In recombinant ZP3, two vicinal O-linked oligosaccharides are essential for high affinity sperm binding. On the other hand, removal of N-glycans from intact egg cells results in decreased sperm binding. It is possible that N-linked glycans induce a conformation of ZP3 necessary for the presentation of the essential O-glycans [244].

Various exogenous oligosaccharides are able to inhibit mouse sperm-egg binding. The enzymatically synthesized tetravalent oligosaccharide with terminal LacNAc or Galα1-3 units (corresponds the branched tetravalent sLex structure in Fig. 7, page 23, without Fucα1,3 and Neu5Acα2,3 residues) inhibits binding by 80-90% at a concentration of 4-10 µM [245]. 40-60% inhibition of binding is achieved with LacNAcβ1-3GlcNAc and Galα1-3LacNAc at concentrations 10-70 µM, and with their α1,3-fucosylated counterparts at <1µM concentrations [246]. Notably, fucosylated ZP-glycans have not been characterized from mouse [244].

Human sperm binding to ZP can be inhibited by various fucosylated and/or sialylated oligosaccharides such as sialyl Lewis x and glycoproteins such as glycodelin-A, reviewed in [244]. Glycodelin-A carries biantennary N-glycans whose terminal ends comprise of
LacdiNAc (GalNAcβ1-4GlcNAc) and in lesser extent LacNAc units that can also be α1,3-fucosylated or α2,6-sialylated [247].

1.6.4 Microbe adhesion

Many pathogens bind to specific carbohydrate structures on the host cell surface. *Helicobacter pylori* that causes gastric ulcers and cancer has binding activity to Lecβ1-3Lacβ1-Cer and Lewis b antigen present on human gastric epithelium [109, 248]. In addition, *H. pylori* binds to α2,3-sialylated PGCs from human erythrocytes [249]. Some *H. pylori* strains carry partially fucosylated PLNs in the O-antigen chains of the lipopolysaccharides. The PLNs contain internal Lex units and terminal Lex or Ley units, depending on the strain [2, 250]. By expressing Lewis epitopes that are also expressed in normal gastric tissue [251], *H. pylori* may camouflage from the host and survive in this way in gastric environment [2]. It is also possible that *H. pylori* binds through a Lex-Lex interaction i.e. the interaction of the Lex units in O-antigen chains with Lex units on host cell surface [250].

*Escherichia coli* heat-labile enterotoxin is able to bind terminal LacNAc units of LacNAcβ1-3Lacβ1-Cer, PGCs, and glycoproteins [252]. A certain *E. coli* strain binds to terminal GlcNAc units of PLNs on erythrocytes [253]. *Plasmodium falciparum* malaria parasites invade human erythrocytes by binding to sialylated, likely Neu5Acα2,3-linked, glycans [254, 255]. *Mycoplasma pneumoniae* and *Streptococcus suis* have binding activity to α2,3-sialylated PLNs on erythrocytes [256, 257]. *Pseudomonas aeruginosa* is able to bind both type 1 and type 2 disaccharide determinants [258].

Fucosylated oligosaccharides of human milk inhibit the binding of several pathogens and bacterial toxins such as *Campylobacter jejuni* and heat stable enterotoxin of *E. coli* to their host cells [259]. Thus, human milk oligosaccharides may block infection in infants by interfering with adhesion and binding events for bacterial colonization and infection.
2. AIMS OF THE STUDY

The aims of the present study were:
1. To study the substrate specificity of the centrally acting β1,6-N-acetylgalcosaminyltransferase activity present in rat serum.
2. To enzymatically synthesize multiply branched polylactosamines to serve as precursors for a potential L-selectin oligosaccharide antagonist.
3. To study the acceptor specificity of a β1,6-N-acetylgalcosaminyltransferase cloned from the human embryonal carcinoma cell line PA1.
4. To study the acceptor specificity of the β1,3-galactosyltransferase activity present in the human colon adenocarcinoma cell line.
5. To study the capability of the centrally acting β1,6-N-acetylgalcosaminyltransferase activity to branch polylactosamines that contain a terminal type 1 unit.
3. MATERIALS AND METHODS

3.1 Acceptor saccharides

For the numbered structures (except 20) see Tables 4 and 5 and Figure 8 in part 4.

The radiolabeled acceptors used in publication I were obtained as follows: acceptors 1 and 2 were enzymatically synthesized as described in [175] and (I), acceptors 5-7 were obtained from metabolically labeled embryonal carcinoma cells [260, 261], acceptor 3 was obtained by incubating glycan 6 with UDP-Gal and β4Gal-T, and acceptor 4 by incubating glycan 5 with UDP-GlcNAc and Β3Gn-T [175]. The unlabeled acceptors 1, 2, and 3 were synthesized from LacNAc (Sigma) by stepwise Β3Gn-T and Β4Gal-T reactions. Acceptor 20 LacNAcβ1-3Lexβ1-3LacNAc (Lex is Galβ1-4(Fucα1-3)GlcNAc) used in publication II was synthesized as follows: GlcNAcβ1-3Lexβ1-3LacNAc was synthesized enzymatically from glycan 2 [262] and converted to radiolabeled glycan 20 by incubation with UDP-[3H]Gal and Β4Gal-T. Acceptors used in publication III were synthesized as described: acceptor 6 [263], acceptor 3 [264], acceptor 21 (II), and LacNAcβ1-3Lex [265]. In publication IV, acceptor 24 was synthesized from GlcNAcβ1-3Galβ1-OMe (Sigma) by stepwise Β4Gal-T and Β3Gn-T reactions, radiolabeled acceptors 25 and 27 were synthesized as described in [266] and [263]. Unlabeled acceptors 2 and 6 were synthesized as above and acceptor 26 was purchased from Sigma.

UDP-Gal and UDP-GlcNAc were from Sigma (MO, USA). UDP-[3H]Gal and UDP-[14C]Gal were from Amersham (UK).

3.2 Expression and purification of GST-IGnT1

The β1,6-N-acetylgalcosaminyltransferase (IGNt1, EC 2.4.1.150) from the human cell line PA1 was previously cloned and sequenced [178]. Construction, expression, and purification of a functional recombinant glutathione sulfotransferase (GST)-IGNT fusion protein that represents the stem and Golgi luminal regions (amino acids 26-400) of native IGNT1 is described in publication III.

3.3 Enzyme catalyzed transferase reactions

Human serum β1,3-N-acetylgalcosaminyltransferase (β3Gn-T) and bovine milk β1,4-galactosyltransferase (β4Gal-T, EC 2.4.1.90, Sigma) reactions were carried out essentially as described in [263] and [133]. β1,6-N-acetylgalcosaminyltransferase (IGNt) reactions are described in publications as follows: rat serum IGNt (I, IV), human serum IGNt (I) and
recombinant GST-IGnT1 (III). Human colon adenocarcinoma cells (Colo 205) were grown and lyzed, and used in β1,3-galactosyltransferase (β3Gal-T) reactions as described in (IV).

3.4 Degradative reactions

Enzymatic hydrolysis with jack bean (exo)-β-galactosidase (EC 3.2.1.23, Sigma) and endo-β-galactosidase from Bacteroides fragilis (EC 3.2.1.103, Boehringer Mannheim, Germany) were performed as described in [261] and [175]. Diplococcus pneumoniae β-galactosidase (EC 3.2.1.23, Boehringer Mannheim) reactions were carried out as described in [267]. Partial acid hydrolysis and mild acid hydrolysis were carried out as described in [267] and [268].

3.5 Chromatographic methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Described in publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desalting of reaction mixtures in a mixed bed of Dowex AG 1 (AcO⁻) and Dowex AG 50 (H⁺)</td>
<td>I, III, IV</td>
</tr>
<tr>
<td>Paper chromatography</td>
<td>I, III, IV</td>
</tr>
<tr>
<td>Size-exclusion chromatography on a Biogel P-4 column</td>
<td>I</td>
</tr>
<tr>
<td>Size-exclusion chromatography on a Superdex 75 HR 10/30 or Superdex peptide HR 10/30 column</td>
<td>I&lt;sup&gt;a&lt;/sup&gt;, II&lt;sup&gt;b&lt;/sup&gt;, IV&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>High-pH anion-exchange chromatography on a CarboPac PA-1 column</td>
<td>I, IV</td>
</tr>
</tbody>
</table>

<sup>a</sup> Referenced to [138], <sup>b</sup> referenced to [221]

3.6 NMR spectroscopy

NMR spectroscopy was performed in D₂O at 500 MHz by Varian unity 500 spectrometer (Varian Inc., CA, USA). One-dimensional <sup>1</sup>H NMR spectroscopy was carried out as described in publication I. Two-dimensional NMR experiments are described in IV.

3.7 Mass spectrometry

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was performed with a Vision 2000 (Finnigan MAT Ltd., UK) or BIFLEX<sup>TM</sup> (Bruker-Franzen Analytic, Germany) mass spectrometer as described in publication I.
4. RESULTS

4.1 Synthesis of multiply branched polylactosamines in vitro (I)

The substrate specificity of the centrally acting β1,6-N-acetylglucosaminyltransferase(s) (cIGnTs) present in rat serum was studied by incubating polylactosamine acceptors with UDP-GlcNAc and rat serum. The structures of the acceptors and the products of the transferase reactions are shown in Table 4.

The branching reaction of saccharide 1 (438 nmol) gave the doubly branched octasaccharide 10 (254 nmol) and the singly branched heptasaccharides 8 and 9 (78 nmol) almost in equal amounts. The products were isolated by size-exclusion chromatography on a Biogel P-4 column followed by separation of the saccharides 8 and 9 by HPAE chromatography. Reincubation of the heptasaccharides 8 and 9 with UDP-GlcNAc and rat serum gave octasaccharide 10. This established that octasaccharide 10 was generated via two different pathways: 1 → 8 → 10 and 1 → 9 → 10. Saccharide 10 was completely β1,4-galactosylated by bovine milk β1,4-galactosyltransferase into the decasaccharide LacNAcβ1-3(LacNAcβ1-6)LacNAcβ1-3(LacNAcβ1-6)LacNAc (11), the smallest possible prototype of a naturally occurring multiply branched polylactosaminoglycan backbone. A preparative scale branching reaction of saccharide 2 gave the singly branched hexasaccharide 12 and a small amount of the doubly branched heptasaccharide 14, but not the singly branched hexasaccharide 13. The branch in hexasaccharide 12 induced some reactivity of the subterminal Gal; reincubation of hexasaccharide 12 converted it to heptasaccharide 14. All reaction products were identified by MALDI-TOF MS and 1D 1H NMR spectroscopy.

A panel of radiolabeled acceptors (1-7), obtained by enzymatic synthesis and from metabolically labeled embryonal carcinoma cells [260, 261], were incubated with UDP-GlcNAc and rat serum to define the minimal optimal sequence for the cIGnT reaction. The unreacted acceptors and the products were separated by paper chromatography. The position of the branch(es) in products 8-10, 12, 14, and 15 were established by subjecting them to (exo)-β-galactosidase (Jack bean) and endo-β-galactosidase (Bacteroides fragilis) digestion and subsequently identifying the cleavage products by paper chromatography. Partial acid hydrolysis of saccharide 10 further confirmed that both midchain galactoses were branched. The relative reactivities of the acceptors, shown in Table 4, indicate that the LacNAcβ1-3LacNAc sequence is the binding epitope for cIGnT(s). The poor reactivity of the subterminal galactoses of acceptors 2, 4, 6, and 7 indicate that rat serum does not contain significant dIGnT activity. Activity to terminal galactoses, tIGnT activity, was not detected either.

The sera of several mammalian species was shown to contain cIGnT activity with similar acceptor specificity to the one in rat serum cIGnT(s).
Table 4. Structures of the acceptors and the products obtained in reactions catalyzed by the rat serum cIgNt activity. The relative reactivities of the individual acceptor sites are also shown.

<table>
<thead>
<tr>
<th>Acceptor (acceptor site is underlined)</th>
<th>Relative reactivity&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Product(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>120&lt;sup&gt;b&lt;/sup&gt;</td>
<td>GlcNAcβ1 6 Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1 (8)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>150&lt;sup&gt;b&lt;/sup&gt;</td>
<td>GlcNAcβ1 6 Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1 (9)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>c</td>
<td>GlcNAcβ1 6 Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1 (10)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>100</td>
<td>GlcNAcβ1 6 Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1 (12)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>0</td>
<td>GlcNAcβ1 6 Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1 (13)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>0</td>
<td>GlcNAcβ1 6 Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1 (14)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>200</td>
<td>GlcNAcβ1 6 Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1 (15)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>12&lt;sup&gt;d&lt;/sup&gt;</td>
<td>GlcNAcβ1 5 Gal (16)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>20</td>
<td>GlcNAcβ1 5 Gal (17)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>5</td>
<td>GlcNAcβ1 5 Gal (18)</td>
</tr>
<tr>
<td>Galβ1-4GlcNAcβ1 Galβ1-4GlcNAcβ1</td>
<td>0</td>
<td>GlcNAcβ1 5 Gal (19)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Relative reactivities of the acceptors were compared in parallel experiments with equimolar samples of the reference pentasaccharide 2. <sup>b</sup> The data were calculated by taking into account that octasaccharide 10 carried branches in both midchain galactoses, while in 8 and 9 the branch was present either in reducing LacNAc (43%) or middle LacNAc (57%) unit. <sup>c</sup> The relative reactivities of the midchain galactoses are included in the data of glycans 8 and 9. <sup>d</sup> The location of the branch was not established.
4.2 Enzymatic synthesis of precursors for a L-selectin oligosaccharide antagonist (I,II)

The successful synthesis of glycan 10 opened up an enzymatic synthesis route to a linear tetravalent sLex saccharide (Figure 7, page 23) that has been shown to be a nanomolar inhibitor of L-selectin-mediated adhesion of lymphocytes to activated endothelium [221]. Glycan 10 was converted by sequential β3Gn-T, distally acting IGnT, and β4Gal-T reactions into glycan 23 (Synthesis route A, Figure 8). The obtained backbone was further α2,3-sialylated and α1,3-fucosylated into the tetravalent sLex glycan [221].

To simplify the synthesis route to glycan 23, the conversion of the linear octasaccharide LacNAcβ1-3LacNAcβ1-3LacNAcβ1-3LacNAc (21) by cIGnT into triply branched glycan 22 was considered (Synthesis route B, Figure 8). Acceptor 21 is available by chemical synthesis [269, 270], and this in turn would make the synthesis of glycan 22 upscalable. However, it was not known whether cIGnTs are able to generate triply branched polylactosamines. To study this, the octasaccharide acceptor 21 was enzymatically synthesized from the heptasaccharide LacNAcβ1-3Lexβ1-3LacNAc (Lex is Galβ1-4(Fucα1-3)GlcNAc) (20) by sequential human serum β3Gn-T and bovine milk β4Gal-T reactions and finally removing Fuc by mild acid hydrolysis. The fucosylated acceptor was used to prevent the action of cIGnT that is also present in human serum [175]. After each reaction step, the products were purified by size-exclusion chromatography followed by paper or HPAE chromatographies, and their structures were confirmed by MALDI-TOF MS and 1D 1H NMR spectroscopy. Incubation of glycan 21 with UDP-GlcNAc and the cIGnT activity present in rat serum gave several products that were separated by HPAE chromatography. The MALDI-TOF mass spectrum of the principal product showed a major signal that was assigned to (M+Na)+ of Gal4GlcNAc7. The 1H NMR spectrum established the structure of triply branched undecasaccharide 22. Enzymatic β1,4-galactosylation of glycan 22 gave glycan 23, the precursor of tetravalent sLex glycan.

**Synthesis route A**

**Synthesis route B**

*Figure 8.* The original synthesis route A and the novel, simplified synthesis route B to glycan 23. Abbreviations used are: GN, GlcNAc and LN, Galβ1-4GlcNAc.
4.3 The acceptor specificity of the centrally acting $\beta$1,6-N-acetylgalcosaminyltransferase cloned from human embryonal carcinoma cell line PA1 (III)

The cDNA coding for the $\beta$1,6-N-acetylgalcosaminyltransferase (IGN1) responsible for biosynthesis of branched polylactosamines in human PA1 cells was previously isolated, but it was not known whether it codes for a cIGN1 or a dIGN1 enzyme [178]. To determine the specificity of IGN1, a fusion protein representing the catalytic ectodomain of the IGN1 and glutathionesulphotransferase (GST) was expressed in Baculovirus-infected Sf9 insect cells. The functionality of the recombinant GST-IGN1 was shown by incubating separately the acceptors LacNAc$\beta$1-3LacNAc (3) and GlcNAc$\beta$1-3LacNAc (6) with UDP-GlcNAc and Sf9 cell lysate. Paper chromatography of the reaction products showed that tetrasaccharide 3 had been converted in significant amounts into putative LacNAc$\beta$1-3(GlcNAc$\beta$1-6)LacNAc (15), whereas acceptor 6 remained unchanged. The purified recombinant GST-IGN1 also converted glycan 3 to glycan 15. MALDI-TOF MS confirmed the formation of the pentasaccharide product Gal$_2$GlcNAc$_3$. The structure of glycan 15 was initially characterized by (exo)$\beta$-galactosidase (Jack bean) and endo-$\beta$-galactosidase (Bacteroides fragilis) digestions. The experiments showed that the new GlcNAc was not transferred to the terminal Gal, but to the internal Gal or GlcNAc residues. The $^1$H NMR spectrum of glycan 15 complemented the data, confirming the structure of glycan 15. The GST-IGN1 reaction with glycan 21 generated probably isomeric mono- and dibranched products according to MALDI-TOF MS analysis of the reaction mixture. The fucosylated LacNAc$\beta$1-3Lex acceptor was not branched in the transferase reaction. The acceptor specificity of GST-IGN1 showed that it represents cIGN1 type enzyme and resembles rat serum cIGN1 in substrate specificity.

4.4 Enzymatic synthesis of linear and branched polylactosamines containing terminal type 1 units (IV)

A number of oligosaccharide acceptors were incubated with UDP-Gal and lysate of Colo 205 cells known to contain $\beta$3Gal-T and $\beta$4Gal-T activities [155, 158, 271] (Panel A, Table 5). The $\beta$4Gal-T activities were inhibited by adding $\alpha$-lactalbumin to the reaction mixture, and the residual $\beta$1,4-linked Gal residues were removed from the products with Diplococcus pneumoniae $\beta$-galactosidase. The products were purified by combining paper and HPAE chromatographies with size-exclusion chromatography. The purity of the $\beta$1,3-galactosylated products was established by MALDI-TOF MS, and their structures were identified by NMR spectroscopy. The linear acceptors 2, 24, and 6 were most efficiently $\beta$1,3-galactosylated, whereas acceptors 25 and 26 showed weak reactivity. Branch-specific activity was detected with acceptor 27, where the $\beta$1,3-linked GlcNAc was preferred over the GlcNAc$\beta$1,6-branch.
Table 5. Panel A. Structures of the acceptors and the products obtained in reactions catalyzed by the β3Gal-T activity of Colo 205 cells. Panel B. Structures of the products obtained from acceptor 28 in cIGnT reaction of rat serum.

<table>
<thead>
<tr>
<th>Panel A</th>
<th>Acceptors</th>
<th>β1,3-galactosylated products</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image1" alt="" /></td>
<td><img src="image2" alt="" /></td>
<td></td>
</tr>
<tr>
<td><img src="image3" alt="" /></td>
<td><img src="image4" alt="" /></td>
<td></td>
</tr>
<tr>
<td><img src="image5" alt="" /></td>
<td><img src="image6" alt="" /></td>
<td></td>
</tr>
<tr>
<td><img src="image7" alt="" /></td>
<td><img src="image8" alt="" /></td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image9" alt="" /></td>
<td><img src="image10" alt="" /></td>
</tr>
</tbody>
</table>

Panel B

<table>
<thead>
<tr>
<th>GlcNAcβ1,6-branched products</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image11" alt="" /></td>
</tr>
<tr>
<td><img src="image14" alt="" /></td>
</tr>
<tr>
<td><img src="image17" alt="" /></td>
</tr>
<tr>
<td><img src="image20" alt="" /></td>
</tr>
</tbody>
</table>

An alternative synthesis route to branched polylactosamines with terminal type 1 unit was revealed. Incubation of Lecβ1-3LacNAcβ1-LacNAc (28) with UDP-GlcNAc and cIGnT activity of rat serum gave the doubly branched octasaccharide 34 (24%) and a mixture of singly branched heptasaccharides 35+36 (42%) (Panel B, Table 5) that were separated by paper chromatography. MALDI-TOF mass spectra of glycans 34 and 35+36 confirmed the molecular compositions of Hex3HexNAc5 and Hex3HexNAc4, respectively. 1D 1H NMR spectroscopy confirmed the structure of glycan 34 and glycans 35+36, and showed the presence of both heptasaccharides in almost equal amounts. Lecβ1-3LacNAc (30) and LacNAcβ1-3LacNAc (3) as a control were subjected in parallel to branching reactions. MALDI-TOF mass spectra revealed that 55% of glycan 30 was converted into a pentasaccharide Gal2GlcNAc3, whereas glycan 3 was totally branched to a pentasaccharide. The results suggest that at least the cIGnT type enzymes can branch Lecβ1-3LacNAcβ1-OR type glycans.

4.5 Structural analysis of the products by NMR spectroscopy (I-IV)

The 500 MHz 1H NMR spectra of the products 8-12 and 14 fully confirmed the structures predicted from the degradative tracer experiments (I). The spectrum of glycan 8 showed the H-1 signal of the β1,6-linked GlcNAc at 4.585 ppm, which was identical to the corresponding signal in LacNAcβ1-3(GlcNAcβ1-6)LacNAc (15) reported in [176, 192]. The spectrum of glycan 9 showed the H-1 signal of the β1,6-linked GlcNAc at 4.593 ppm, which corresponded
to the H-1 signal of the analogous residue in GlcNAcβ1-3(GlcNAcβ1-6)LacNAcβ1-3LacNAc [167]. The spectrum of glycan 10 showed two H-1 signals of β1,6-linked GlcNAc residues at 4.585 ppm and at 4.593 ppm, revealing the presence of GlcNAcβ1,6-branches at both midchain galactoses. Branching caused marked changes in the structural reporter group signals of the main chain. The H-1 and H-4 signals of the galactoses in acceptor 1 were consistently shifted upfield when branched. A similar upfield shift has also been reported for GlcNAcβ1-3(GlcNAcβ1-6)LacNAc sequence [164]. GlcNAcβ1,6-branching of the reducing end LacNAc unit regularly shifts the H-1 resonance of the reducing end GlcNAc downfield [164, 192]. The H-1 signal of the terminal Gal was not affected by branch formation. The β1,4-galactosylation of glycan 10 gave glycan 11. The spectrum of glycan 11 revealed a new doublet of two protons at 4.465 ppm, which were assigned to the H-1’s of the branch-galactoses in accordance with the analogous H-1 resonance of the branch-galactose in LacNAcβ1-3(LacNAcβ1-6)LacNAc [272]. The spectra of glycans 12 and 14 agreed in relevant parts with the spectra of glycans 8 and 10. In addition, the spectrum of glycan 14 was practically identical to the spectrum of the analogous glycan where the distal branch was synthesized by dIgnt activity of hog gastric microsomes and the central one by cIgnt activity of rat serum [167].

All structures along the synthesis route from LacNAcβ1-3Lexβ1-3LacNAc (20) to LacNAcβ1-3LacNAcβ1-3LacNAcβ1-3LacNAc (21) were established by 1D 1H NMR spectroscopy (II). Glycan 21 was converted into the triply branched glycan 22 by the cIgnt activity of rat serum. The spectrum of glycan 22 revealed three new protons, one at 4.585 ppm and two at 4.593 ppm. The first one originates from the GlcNAcβ1-6 linked to the reducing end LacNAc unit, and the latter two from the GlcNAcβ1-6 residues linked to the midchain LacNAc units. The reporter group signals of glycan 22 were practically identical with their counterparts found in the spectra of the doubly branched glycan 10. The spectrum of the β1,4-galactosylated glycan 23 showed three H-1 signals of the newly transferred branch galactoses at 4.464 ppm, which corresponded to the signals of glycan 11.

LacNAcβ1-3LacNAc was converted into LacNAcβ1-3(GlcNAcβ1-6)LacNAc (15) with the recombinant GST-Ignt1 enzyme (III). The 1H NMR spectrum of pentasaccharide 15 showed practically identical structural reporter group signals with the authentic pentasaccharides synthesized by the cIgnt of rat serum [192] and Pa1 cell lysate [179]. The similarity of the H-1 signal of the new GlcNAc residue and the H-1 and H-4 signals of the midchain Gal residue of saccharide 15 with the corresponding signals of authentic saccharides indicated that the new GlcNAc residue was β1,6-bonded to the midchain galactose.

1D 1H NMR spectra were recorded for the β1,3-galactosylated products (IV). However, the resonances of the putative Lex units differed from the corresponding signals reported for Lecβ1-3Lac [273] and when compared to the 1H chemical shifts of LacNAcβ1-3Lac [273] the substitution with either Galβ1-3 or Galβ1-4 could not be confirmed. Therefore, the type 1 linkage (underlined) on Galβ1-3GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-OMe (29) was unambiguously identified by the assignment of the 1H signals from
DQFCOSY and TOCSY spectra, and subsequently of the $^{13}$C signals from a HMOC spectrum. The C-3 resonance of the subterminal GlcNAc (83.35 ppm) was shifted downfield as compared to the resonance of unsubstituted GlcNAc (~70 ppm) indicating glycosidic linkage of type 1. The H-1 doublet of Galβ1-3 residue with the $^3J_{1,2}$ coupling constant of 7.8 Hz confirmed the β anomic configuration. The unambiguous identification of glycan 29 established the H-1 resonances of type 1 unit GlcNAc at 4.725 ppm and Gal at 4.441 ppm. The linear glycans 28 and 30 were identified from 1D $^1$H NMR spectra according to that data. The $^1$H NMR spectra of Lecβ1-6LacNAc/Lac, the linear glycans 31 and 32, showed the H-1 signals of the reducing part (Rβ1-6LacNAc/Lac) that corresponded the signals recorded for their acceptor saccharides. The H-1 signals of the GlcNAcβ1,6 residues were similarly shifted downfield (~0.050 ppm) as the signals of the β1,3-galactosylated GlcNAcβ1,3 residues in glycans 28-30. Accordingly, the signals at 4.437 ppm (glycan 31) and 4.426 ppm (glycan 32) resonating upfield from 4.441 ppm were assigned as the H-1 signals of the β1,3-linked galactoses. The $^1$H NMR spectrum of Lecβ1-3(GlcNAcβ1-6)LacNAc, glycan 33, established that β1,3-linked GlcNAc in the acceptor 27 was β1,3-galactosylated. The H-1 resonances of type 1 unit were similar to those in glycans 28-30 and the resonances of the reducing part Rβ1-3(GlcNAcβ1-6)LacNAc were in accordance to those reported for LacNAcβ1-3(GlcNAcβ1-6)LacNAc (15) [192].

The branching reaction of glycan 28 with rat serum cIGnT gave octa- and heptasaccharide products (IV). The 1D $^1$H spectrum of the octasaccharide 34 revealed the characteristic H-1 signal of Galβ1-3 at 4.440 ppm establishing the terminal type 1 linkage and excluding GlcNAc substitution of the nonreducing end Gal. The two H-1 signals of the β1,6-linked GlcNAcs resonated exactly in the same positions as the corresponding signals of glycan 10 (I), confirming the structure of glycan 34. The spectrum of the heptasaccharide products showed that this material represented a mixture of isomeric glycans 35+36. Two doublets of nearly equal size in the area of H-1 signals of β1,6-linked GlcNAcs were similar to those originating from glycan 34, revealing the presence of almost equal amounts of products branched at midchain or reducing end LacNAc units.
DISCUSSION

5.1 Centrally acting $\beta 1,6$-N-acetylglucosaminyltransferases synthesize multiple branches on linear polylactosamines \textit{in vitro}

The cIGnT activity present in rat, human, and other mammalian sera was shown to convert hexasaccharide 1 into the doubly branched octasaccharide 10 via two distinct pathways: $1 \rightarrow 8 \rightarrow 10$ and $1 \rightarrow 9 \rightarrow 10$ (I). The first GlcNAc branch transferred to acceptor 1 did not prevent the secondary branching reaction, which led to the octasaccharide 10. In addition, it was shown that the LacNAc$\beta 1$-3LacNAc sequence is required for efficient activity of cIGnT(s). The results suggest that cIGnT(s) interact with hexasaccharide 1 in the bimodal manner shown in Figure 9. The transferase reaction of rat serum cIGnT(s) with octasaccharide LacNAc$\beta 1$-3LacNAc$\beta 1$-3LacNAc$\beta 1$-3LacNAc (21) gave the triply branched 22 as a major product and minor products, which probably represented the isomeric singly and doubly branched products (II).

Figure 9. Postulated interaction of the hexasaccharide acceptor 1 with the cIGnT present in mammalian serum. Interaction with site A leads to the formation of glycan 8 and interaction with site B leads to the formation of glycan 9. The secondary branching reactions lead to glycan 10.

The fusion protein GST-IGnT1, representing the catalytic ectodomain of IGnT expressed in human PA1 cells, was functionally expressed, purified, and shown to have cIGnT activity (III). The lack of the cytoplasmic and membrane binding segment in the recombinant IGnT1 probably did not cause major changes in the substrate specificity, because PA1 cell lysates have been shown to contain cIGnT rather than dIGnT activity as well [179]. Recombinant IGnT1 converted octasaccharide 21 into mono- and dibranched products. The triply branched product was not formed as with rat serum, which was likely due to the small amount of the enzyme available.
At present, similar substrate specificity to mammalian serum cIGnT and GST-IGnT1 has been reported for several other cIGnT enzymes. These include protein A-IGnT1, which is an analogue of GST-IGnT1 [142], the cloned mouse IgNTA and IgNTB [184], and cIGnT purified from hog small intestine [177] (Table 2, section 1.4.4.2). Protein A-IGnT1 has been shown to convert LacNAcβ1-3LacNAcβ1-3LacNAcβ1-6Manα1-6Manβ1-octyl into singly and doubly branched products, and branching was more favored at middle LacNAc unit. Similarly, the cloned mouse IgNTA and IgNTB also branch the middle LacNAc unit of LacNAcβ1-3LacNAcβ1-3Lac acceptor more efficiently than the reducing end Lac unit. In contrast, rat serum cIGnT (I) and cIGnT purified from hog small intestine [177] branch both midchain and reducing end LacNAc units in acceptor 1 equally well. It is not clear whether the different site specificities are due to the specificities of the enzymes or the acceptors used. N-acetyl group in the reducing end of the acceptor has been shown to enhance the human serum cIGnT activity; the relative reactivities of GlcNAcβ1-3LacNAcβ1-3LacNAc (2) and GlcNAcβ1-3LacNAcβ1-3Lac of 100 and 41, respectively, have been reported (the reacting units are underlined) [175]. On the other hand, for example α1,3-fucosyltransferases have different site specificities on linear poly lactosamines [274].

Rat serum did not contain significant amounts of dIGnT activity as there were no branching at the subterminal Gal of acceptor 2 and only minor amount of acceptor 6 was branched. However, the pre-branched glycan 12 was distally branched, yielding glycan 14, when it was reincubated with UDP-GlcNAc and rat serum. Protein A-IGnT1 also contains some dIGnT activity: it branches very weakly the subterminal Gal residue of the GlcNAcβ1-3LacNAcβ1-6Manα1-6Manβ1-octyl acceptor [142] (see Table 2). Thus, cIGnTs may show minor dIGnT activity with certain acceptors.

Fucosylation of polylactosamines has been shown to prevent the action of cIGnT of rat serum [186] and the cloned mouse IgNTA and IgNTB [184]. The recombinant GST-IGnT1 could not branch the fucosylated acceptor LacNAcβ1-3Lex either (III). Hence, fucosylation and GlcNAcβ1,6-branching seem to be reciprocal in LacNAc units on polylactosamines.

The doubly branched glycan 10 was successfully converted with bovine milk β4Gal-T into LacNAcβ1-3(LacNAcβ1-6)LacNAcβ1-3(LacNAcβ1-6)LacNAc (11). Glycan 11 presents the smallest possible multiply branched sequence found in natural polylactosaminoglycans. Another biosynthesis route to multiply branched polylactosaminoglycans has been proposed by Ujita et al. In their model β1,6-N-acetylglucosaminyltransferase (IGnT1), β1,3-N-acetylglucosaminyltransferase (iGnT), and β1,4-galactosyltransferase (β4Gal-T1) act sequentially, elongating and branching the growing chain in turn [142].
5.2 Branched polylactosamines as precursors for biologically active oligosaccharides

The triply branched glycan 23, the precursor of the linear tetravalent sLex glycan (shown in Figure 7, page 23), has been initially synthesized from the hexasaccharide 10 by three enzymatic steps [221]. The improved synthesis route to glycan 23 involved the conversion of the linear octasaccharide 21 into the triply branched glycan 22 in a single-step transformation catalyzed by the cGnT activity of rat serum and β1,4-galactosylation of glycan 22 into glycan 23 (II) (Figure 8, page 32). Hence, the novel synthesis route was devoid of the laborious β3Gn-T and distally acting IGnT reactions. The major advantage in the use of the linear glycan 21 as a primer is that it is also accessible in considerable amounts by chemical synthesis in solution [269, 270]. The chemical synthesis of branched glycans related to glycan 23 have also been presented [275, 276] suggesting that the chemical synthesis of glycan 23 may also become available. Expression of the functional human IGnT1 (III) allows the branching reaction by a recombinant enzyme. At present, a repertoire of glycosyltransferases has been cloned, which should allow the whole synthesis of glycan 23 and its conversion into the linear tetravalent sLex glycan with recombinant enzymes. Thus, the availability of chemical, enzymatic, and hybridized chemo-enzymatic approaches should make the large scale synthesis of the tetravalent sLex glycan possible. This in turn, will lead to better assessment of the anti-inflammatory potential of the tetravalent sLex glycan in different in vivo inflammation models. Also its putative roles as antagonists of E- and P-selectins and other sugar-binding proteins will also merit a study.

5.3 Biosynthetic pathways to polylactosamines with terminal type 1 unit

The β3Gal-T activity of Colo 205 cells was shown to use both linear and branched polylactosamines as acceptors (IV). The β3Gal-T activity of Colo 205 cells is most likely due to β3Gal-T5 [158], but β3Gal-T1 and β3Gal-T2 may also be responsible in some part for the activity or activities [271]. In the branched acceptor GlcNAcβ1-3(GlcNAcβ1-6)LacNAc, the GlcNAcβ1,3-branch was strongly favoured over the GlcNAcβ1,6-branch. In contrast, β4Gal-T of bovine milk has been shown to preferentially react at the GlcNAcβ1,6-branch in the corresponding acceptors [135-138]. The complementary branch specificities of β3Gal-T(s) and β4Gal-T are in accordance with the natural type 1-type 2 epitope structures, Lecβ1-3(LacNAcβ1-6)LacNAc (Table 1, section 1.3).

An alternative route to branched polylactosamines with terminal type 1 disaccharides was shown by branching linear Lec[β1-3LacNAc]$_1$ or 2 oligosaccharides with the cGnT activity of rat serum (Scheme 1). It has been reported previously that the cGnTs purified from rat and hog intestine do not branch Lecβ1-3Lac-PA under their reaction conditions [123, 177], but the cloned mouse IGNTA and IGNTB have some reactivity to the Lac unit in Lecβ1-3(LacNAcβ1-3)$_0$ or 1Lac acceptors, although the products were not characterized [184].
Human serum contains similar cIGnT activity to that of rat serum (I), suggesting that also human cIGnT(s) synthesize branches on type 1 polylactosamines.

**Scheme 1.** Two proposed biosynthetic pathways leading to branched type 1 polylactosamines.

![Scheme 1](image1.png)

### 5.4 Structural determination of polylactosamines by NMR spectroscopy

The linear type 2 polylactosamines were readily characterized by 1D $^1$H NMR spectra. The $^1$H structural reporter group signals of glycans ranging in size from two LacNAc units to four LacNAc units followed a similar pattern. Typical signals were obtained for reducing end GlcNAc and nonreducing end Gal residues. The H-1 signals of the internal GlcNAc residues and the H-1 and H-4 signals of internal Gal residues were practically identical in all LacNAc units.

The newly formed type 1 linkage on linear polylactosamine was unambiguously identified by a series of 1D and 2D NMR experiments, and the $^1$H structural reporter group signals of the terminal type 1 unit were determined (IV). Comparison of the $^1$H structural reporter group signals of Lecβ1-3Lac [273] and Lecβ1-3LacNAc tetrasaccharides (IV) showed that Glc and GlcNAc at the reducing ends have different effects on the $^1$H resonances of the terminal type 1 units present in these tetrasaccharides. In contrast, the $^1$H structural reporter group signals of the distal LacNAc units in the type 2 tetrasaccharides LacNAcβ1-3Lac [273] and LacNAcβ1-3LacNAc [192] are consistent with each other. This shows the importance of the production and unambiguous NMR analysis of each natural epitope as references for later 1D NMR works.

In branched polylactosamines, ranging from singly to triply branched glycans, the H-1 signal of the GlcNAcβ1,6 residue bound to the reducing end LacNAc unit resonated at distinct position from those bound to the middle LacNAc unit(s). When branch GlcNAcs were β1,4-galactosylated, all the H-1 signals of the branch Gals resonated at the same position, and the signals of GlcNAcβ1,6 residues were shifted downfield maintaining different resonances for reducing end and midchain LacNAc bound residues. The distinct chemical shifts have yet
been observed when distal galactoses carry $\alpha$2,3-bonded Neu5Ac residues [221]. These features provide possibilities to monitor the action of glycosyltransferases and glycosidases at distinct sites of branched, galactosylated, and sialylated glycans.

In branched polylactosamines with terminal type 1 unit, ranging from singly to doubly branched glycans, the H-1 signal of the GlcNAc$\beta$1,6 residues were in accordance with analogous type 2 polylactosamines. Thus, the only differences in the $^1$H structural reporter group signals were the signals of Gal$\beta$1-3GlcNAc unit. It is worth noting that in Gal$\beta$1-3/4(GlcNAc$\beta$1-6)GlcNAc-R structures the H-1 of GlcNAc$\beta$1-6 residue resonate at the same position independently of the terminal linkage, Gal$\beta$1-3 or Gal$\beta$1-4.
6. ACKNOWLEDGEMENTS

This work was carried out in the Laboratory of Carbohydrate Chemistry at the Institute of Biotechnology and the Division of Biochemistry, Department of Biosciences, University of Helsinki under the supervision of Prof. Emer. Ossi Renkonen. I am deeply grateful to Ossi for introducing me the fascinating word of glycobiology and for his support, encouragement, and enthusiasm during these years.

I am grateful to Prof. Mart Saarma, the Director of the Institute of Biotechnology, and Prof. Carl G. Gahmberg, the Head of the Division of Biochemistry, for providing me the excellent facilities to carry out this work.

I want to warmly thank Prof. Marja Makarow, the Head of the Viikki Graduate School in Biosciences, for good advice and the help that I needed to get this work finished. Docent Esa Kuismanen is acknowledged for constructive discussions at the follow up group meetings. Dr. Eeva Sievi, the coordinator of the School, is acknowledged for kind help in many practical matters.

I am grateful to Prof. Marja Makarow and Docent Jaakko Parkkinen for critically reviewing this thesis and for their valuable comments. Dr. Jarkko Räbinä is acknowledged for carefully reading the thesis and for proposals for improvements.

Warmest thanks are due to my friends and colleagues at the Laboratory of Carbohydrate Chemistry in the past years: Anja, Anne, Antti, Hanna, Hannu, Jari, Jarkko, Juha, Juho, Katja, Krisse, Leena, Maria, Marie, Marjo, Mika, Minna, Natu, Olli, Pinja, Reko, Ritva, Suvi, Tanja, and Tero. I thank you all for help, support, and discussions, and for the enjoyable atmosphere in the lab. Especially I want to thank my long-term colleagues for making this thesis possible by their groundwork, by teaching me the laboratory practice, and by co-operation. Special thanks are due to Hannu for patiently teaching me NMR spectroscopy. I am deeply grateful to Hannu and Natu for their invaluable help in getting manuscript IV written, and helpful comments on this thesis. I thank warmly Suvi for encouragement and for solving my problems with the English language. Big thanks to Hanna for friendship and her humour that cheered me up in hard days. I wish to thank Leena and Lotta for the time we shared at Biomedicum.

I express my gratitude to Docent Risto Renkonen, Dr. Pirkko Mattila, Dr. Laura Hirvas and Jaana Niittymäki, and all the other members of Risto’s group for fruitful collaboration. I was lucky to get a chance to work with you at Biomedicum last year. Thank you for your friendliness and support at that time.
I thank Drs. Catherine E. Costello, Jari Helin, and Ritva Niemelä for their co-operation in mass spectrometry.

I thank prof. Ilkka Kilpeläinen and all the members of the NMR group for help and kindliness.

I thank my dear friends Ansku, Minna H., Minna L., Minna P., and Ninni for sharing the ups and downs of life.

I express my warmest thanks to my mother Marja, Anu and Teemu, Mikko and Johanna, and my grandmother Rauha for always being ready to help me and for optimistic attitude.

Finally, Jouni, Viljo, and Veeti: your love, care, and patience carried me through the hard times to this goal, thank you.

This study was financially supported by the Jenny and Antti Wihuri Foundation, the Finnish Cultural Foundation, the University of Helsinki, and the Viikki Graduate School in Biosciences.

Helsinki, November 2003

Heidi Salminen
7. REFERENCES


89. Dabrowski, J., Dabrowski, U., Bermel, W., Kordowicz, M., and Hanfland, P. (1988) Structure elucidation of the blood group B like and blood group I active octaantenmery ceramide tetracontasaccharide from rabbit erythrocyte membranes by two-dimensional ¹H NMR spectroscopy at 600 MHz *Biochemistry* 27, 5149-5155


145. Ujita, M., Misra, A. K., McAuliffe, J., Hindsgaul, O., and Fukuda, M. (2000) Poly-N-acetyllactosamine extension in N-glycans and core 2- and core 4-branched O-glycans is differentially controlled by i-extension enzyme and different members of the β1,4-galactosyltransferase gene family J. Biol. Chem. 275, 15868-15875


162. Salvini, R., Bardoni, A., Valli, M., and Trinchera, M. (2001) \( \beta_1,3 \)-Galactosyltransferase \( \beta_3 \text{Gal-T5} \) acts on the GlcNAc\( \beta_1-3 \text{Gal}\beta_1-4\text{GlcNAc}\beta_1-\text{R} \) sugar chains of carcinoembryonic antigen and other N-linked glycoproteins and is down-regulated in colon adenocarcinomas. *J. Biol. Chem.* 276, 3564-3573


166. Brockhausen, I., Matta, K. L., Orr, J., Schachtner, H., Koenderman, A. H., and van den Eijnden, D. H. (1986) Mucin synthesis. Conversion of \( R_1-\beta_1-3\text{Gal-R}_2 \) to \( R_1-\beta_1-3\text{(GlcNAc to Gal)}-6\text{Gal-R}_2 \) and of \( R_1-\beta_1-3\text{GalNAc-R}_2 \) to \( R_1-\beta_1-3\text{(GlcNAc to Gal)}-6\text{GalNAc-R}_2 \) by a \( \beta_6 \)-N-acetylglucosaminyltransferase in pig gastric mucosa. *Eur. J. Biochem.* 157, 463-474


169. Korekane, H., Taguchi, T., Sakamoto, Y., Honke, K., Dohmae, N., Salminen, H., Toivonen, S., Helin, J., Takio, K., Renkonen, O., and Taniguchi, N. (2003) Purification and cDNA cloning of UDP-GlcNAc:GlcNAc\( \beta_1-3 \text{Gal-R}_1 \)-\( \beta_1-4\text{Glc(NAc-R)} \) \( \beta_1,6 \)-N-acetylglucosaminyltransferase from rat small intestine: a major carrier of dIGnT activity in rat small intestine *Glycobiology* 13, 387-400


176. Niemelä, R., Räbinä, J., Leppänen, A., Maaheimo, H., Costello, C. E., and Renkonen, O. (1995) Site-directed enzymatic \( \alpha-(1-3)-L \)-fucosylation of the tetrasaccharide Gal[\( \beta(1-4)\text{GlcNAc}\beta(1-3)\text{Gal} \)\( \beta(1-4)\text{GlcNAc} \) at the distal N-acetyllactosamine unit *Carbohydr. Res.* 279, 331-338


191. van Halbeek, H. (1993) NMR spectroscopy of carbohydrates. A tutorial, American Chemical Society, 205th National meeting, Denver, CO, USA


58


