Molecular Fingerprinting of Carbohydrate Structure Phenotypes of Three Porifera Proteoglycan-like Glyconectins*

Yann Guerardel‡, Xavier Czeszak‡, Lazar T. Sumanovski‡, Yannis Karamanos§, Octavian Popescu¶, Gerard Strecker†, and Gradimir N. Misevic**‡‡

From the **Laboratoire des Processus Intégratifs Cellulaires, UMR 6037 CNRS, Faculté des Sciences et Techniques de Rouen, 76821 Mont Aignan Cedex, France, †Unité de Glycobiologie Structurale et Fonctionnelle, Université des Sciences et Technologies de Lille, UMR 8576 CNRS, 59655 Villeneuve D’Ascq, France, §Department of Research, University Hospital of Basel, CH-4038 Basel, Switzerland, ¶Laboratoire de Biochimie Moléculaire et Cellulaire, Université d’Artois, Faculté J. Perrin, rue J. Sourraz, SP18, 62307 Lens, France, and **Molecular Biology Center and Institute for Interdisciplinary Experimental Research, Babes-Bolyai University, 400066 Cluj-Napoca, Romania

Glyconectins (GNs) represent a new class of proteoglycan-like cell adhesion and recognition molecules found in several Porifera species. Physico-chemical properties of GN carbohydrate moieties, such as size, composition, and resistance to most glycosaminoglycan-degrading enzymes, distinguish them from any other type of known glycoproteins. The molecular mechanism of GN-mediated self/non-self discrimination function is based on highly species-specific and Ca\(^{2+}\)-dependent GN to GN associations that approach the selectivity of the evolutionarily advanced immunoglobulin superfamily. Carbohydrates of glyconectins 1, 2, and 3 are essential for species-specific auto-aggregation properties in three respective Porifera species. To obtain a structural insight into the molecular mechanisms, we performed carbohydrate structural analyses of glyconectins isolated from the three sponge model systems, Microciona prolifera (GN1), Halichondria panicea (GN2), and Cliona celata (GN3). The glycan content of all three GNs ranged between 40 and 60% of their total mass. Our approach using sequential and selective chemical degradation of GN glycans and subsequent mass spectrometric and NMR analyses revealed that each glyconectin presents novel and highly species-specific carbohydrate sequences. All three GNs include distinct acid-resistant and acid-labile carbohydrate domains, the latter composed of novel repetitive units. We have sequenced four short sulfated and one pyruvilated unit in GN1, eight larger and branched pyruvilated oligosaccharides in GN2, which represent a heterogenous but related family of structures, and four sulfated units in GN3.

Porifera glyconectins (GNs)\(^{1}\) are emerging as a new class of proteoglycan-like cell adhesion and recognition molecules. In the preceding article (22) the molecular mechanism of GN-mediated self/non-self discrimination was shown to be based on highly species-specific and Ca\(^{2+}\)-dependent glyconectin associations operating via carbohydrates. The selectivity of these macromolecular interactions approached that of the immunoglobulin superfamily. Based on these findings xeno-selectivity of GN to GN recognition in Porifera, the closest existing relatives to the primordial multicellular organisms, was proposed to be a new self-assembly paradigm in the evolution of multicellular forms of life and the divergence of species.

Early work on purified carbohydrates of Microciona prolifera glyconectin 1, published in 1987 and 1993 (1, 2) and followed by experiments on synthetic oligosaccharide epitopes in 2001 (3), showed that GN1 functions in cell aggregation via glycan to glycan interactions. Although the carbohydrates of GN2 and GN3, purified from Halichondria panicea and Cliona celata, respectively, are functional in cell adhesion and recognition, it remains to be demonstrated whether they operate as GN1 through glycan to glycan associations. An analogous molecular mechanism of carbohydrate-carbohydrate binding involvement in cell adhesion and recognition was described somewhat later, in 1989, in mammalian cell culture systems (4–6) and was extended to different types of tumor cells more recently (7, 8). Carbohydrate self-association is complementary to well studied lectin-glycan interactions where the role of carbohydrates in cellular interactions has been well documented for more than two decades (9, 10).

Porifera glyconectin carbohydrates account for 40–60% of their total mass. However, their precise structure is still largely unknown. Only four short oligosaccharides, generated by mild hydrolysis of GN1, have been characterized as a new type of pyruvilated and sulfated oligosaccharide sequences, but their arrangement within the glycan moiety is unknown (11, 12). These initial sequence analyses of GN1 glycans, represented mainly by two macromolecular species with molar masses of 200 and 6 kDa (1, 2), revealed that they represent a novel class of acidic and fucose-containing polysaccharides different from those of classical glycosaminoglycans. Purified carbohydrates from GN2 and GN3 showed unconventional monosaccharide composition, which, together with their resistance to known glycosaminoglycans degrading enzymes, indicated a type of structure yet to be described (see preceding article (22)). Polycrylamide gel electrophoresis showed that GN2 contains one major glycan of molar mass 180 × 10\(^3\) and that GN3 contains one of 110 × 10\(^3\) (see preceding article (22)). The absence of information on the details of the structural diversity of glyconectin carbohydrates has greatly impaired investiga-
Structural Fingerprinting of Glyconectin Carbohydrates

Isolation of Glyconectin Proteoglycans—Glyconectins were extracted from fresh cuts of sponges with artificial Ca^{2+} and Mg^{2+}-free seawater (ACMFSW; 462 mM NaCl, 10.7 mM KCl, 7 mM Na_{2}SO_{4} and 2.1 mM CaCl_{2}) precipitated under mild conditions (0.1 M trifluoroacetic acid at 20 mM CaCl_{2} precipitation, and ultracentrifugation (1, 2). Partial hydrolyses (data not shown). The ethanol-soluble fraction was essentially composed of unhydrolyzed, denatured glycoprotein. In GN1, a minor water-soluble fraction (6% of the total monosaccharide content and minute amounts of Py(4,6)Gal. GN2 had lower proportion of fucose and a molar ratio of Py(4,6)Gal three times higher than GN1. GN3 had high quantities of fucose but no Py(4,6)Gal at all and was the only glyconectin to contain arabinose (10% mol/mol).

Generation of Oligosaccharides—To compare their carbohydrate moieties, the three glyconectins were fragmented by partial acid hydrolysis. GNs were hydrolyzed repetitively under mild conditions using 0.1 M trifluoroacetic acid at 80 °C for 1 h. After each hydrolysis step, four volumes of ethanol were added, and the liberated oligosaccharides were collected in the ethanol fraction. Release of oligosaccharides from glyconectins was assessed by TLC analysis on silica gel using butanol/acetic acid/H_{2}O (40/30/20) as solvent. Typically, four hydrolysis steps were sufficient to liberate all hydrolysable material. Oligosaccharides were further purified by gel filtration on a Bio-Gel P-2 column (Bio-Rad). GN1 oligosaccharides were reduced by NaBH_{4} and GN2 and GN3 oligosaccharides by NaBD_{4} in aqueous NH_{4}OH 1% solution. Eventually purification was done according to Ciucanu and Kerek (13).

Matrix-assisted Laser Desorption Ionization Time-of-flight (MALDI-TOF) MS—A Vision 2000 time-of-flight instrument (Finnigan Mat) MALDI-MS equipped with a 337-nm UV laser was used for molecular mass measurements of oligosaccharides. 1 µl of sample at a concentration of 100 pmol/µl was mixed with an equal volume of matrix and allowed to crystallize. 2,5-dihydroxybenzoic acid and 3-aminonaphtalene was used for neutral and acidic oligosaccharides, respectively.

Electrospray Mass Spectrometry—Triple quadrupole instrument (Micromass Ltd., Altrincham, UK) fitted with an atmospheric pressure ionization electrospray source was used for MS measurements in positive and negative ion mode. A mixture of glycopropylene glycol water solvent was used to calibrate the quadrupole mass spectrometer. Underviolated oligosaccharides were dissolven in methanol/water (50/50) and permethylated oligosaccharides in acetonitrile, in both cases at concentration of 10 pmol/µl. Nanoflow probe at 50 nM/min was used for infusion. Quadrupole was scanned from 200 to 2000 Da with a scan duration of 3 s and a scan delay of 0.1 s. The samples were sprayed using 1.4 kV needle voltage, and the declustering (cone) was typically set at 70 V. For collision-induced dissociation (CID) experiments, the pressure of argon was set at 4 × 10^{-3} mbar and the collision energy was set to values ranging from 25 to 75 eV.

NMR Spectroscopy—Bruker® ASX400 and DMX600 spectrometers both with a 5-mm HETCQ mixed probe-head operating in the pulse Fourier transform mode controlled by an Aspect 3000 computer, were used for NMR experiments. Oligosaccharides were dissolved in 400 µl of H_{2}O after three exchanges with H_{2}O (99.97% atom H, Eurisop-top, CEA group, Gif-sur-Yvette France) and intermediate lyophilizations. NMR analyses were performed at 300 K. The chemical shifts (δ) were referenced to internal acetone (δH = 2.225 and δD = 3.515 ppm under the conditions used). Two-dimensional homonuclear (COSY90, TOCSY) and heteronuclear (HMQC) experiments were performed using standard Bruker® pulse programs. The main pulses and variable delays were optimized for each pulse program and sample.

Gas Chromatography—Monosaccharides were analyzed by GC/MS as per(trimethylsilyl) (14) and perheptafluorobutyryl (15) derivatives. Methylation of oligosaccharides was done as described by Ciucanu and Kerek (13) before analysis on GC/MS.

Other Analytical Methods—Oligosaccharide size was measured by gel filtration on a Bio-Gel P-2 or P-4 column (Bio-Rad) equilibrated in acetic acid 0.5% and calibrated with hydrolyzed dextran in combination with thin layer chromatography. Amino acid composition was analyzed by Pico-Tag high pressure liquid chromatography, and SO_{4}^{2−} content was analyzed by Dionex ion-exchange HPAEC, both after 6 h HCl hydrolysis at 100 °C for 12 h.

RESULTS

Compositional Analysis—As demonstrated previously, the three glyconectins GN1, GN2, and GN3, showed species-specific carbohydrate compositions (see preceding article (22)). GN1 was characterized by a fucose content of one-third of the total monosaccharide content and minute amounts of Py(4,6)Gal. GN2 had lower proportion of fucose and a molar ratio of Py(4,6)Gal three times higher than GN1. GN3 had high quantities of fucose but no Py(4,6)Gal at all and was the only glyconectin to contain arabinose (10% mol/mol).

Structural Fingerprinting of Glyconectin Carbohydrates—Glyconectins were hydrolyzed repetitively under mild conditions (0.1 M trifluoroacetic acid at 80 °C for 1 h). After each hydrolysis step, the liberated oligosaccharides were extracted by ethanol as described under “Experimental Procedures.” These hydrolysis conditions were devised to preserve most of the hexose (Hex) and N-acetyltalosamin (HexNAc) linkages, as well as the pyruvate and sulfate groups, and to cleave the more fragile deoxyhexose (DiHex) and pentose (Pent) linkages. In addition, hydrolysis products were removed repeatedly in an attempt to preserve some of the labile sequences containing Fuc or Ara residues.

The stepwise release of oligosaccharides from glyconectins was monitored by TLC analysis of the ethanol-soluble fractions. Liberated oligosaccharides showed a constant composition from the first to the last hydrolysis step, suggesting that identical material was sequentially cleaved off of the glyconectins by partial hydrolyses (data not shown). The ethanol-soluble fractions of each species were pooled for analysis. Typically, four hydrolysis steps were sufficient to liberate all hydrolysable material.

After hydrolyses were completed, ethanol-insoluble fractions were extracted by water to separate the possible ethanol-insoluble oligosaccharides from the remaining-insoluble glycoproteins. In GN1, a minor water-soluble fraction (6% of the total initial sugar content) was extracted this way, whereas in GN2 and GN3 no water-soluble components were observed as revealed by TLC and confirmed by sugar analyses (data not shown). This suggested that the vast majority of hydrolyzed sugars were ethanol-soluble and that the ethanol-insoluble fraction was essentially composed of unhydrolyzed, denatured glycoprotein.

In similar conditions, the three glyconectins exhibited distinct susceptibility to acid hydrolysis as reflected by the proportions of carbohydrate in the ethanol-soluble fraction after hydrolysis: 68% of the total carbohydrate content for GN1, 56% for GN2, and 41% for GN3. Analysis of each fraction (Table I) revealed that the hydrolyzed materials had very distinct sugar compositions compared with unhydrolyzed materials. In particular, cleaved oligosaccharides were systematically enriched with...
in fucose and arabinose compared with unhydrolyzed fractions. This reflects the preferential cleavage by partial hydrolysis of domains of the molecule rich in acid-labile linkages. On the contrary, glucuronic acid was retrieved predominantly in hydrolysis-resistant fractions. All of these data strongly suggested that the carbohydrate moieties of glyconectins are not constituted by homogeneous polymers but rather by structurally unrelated domains, an acid-labile, presumably external, domain and an acid-resistant polysaccharidic core. This work concentrated on the comparison of the acid-labile domains of glyconectins from the three species by sequencing the hydrolyzed oligosaccharides.

**Sequencing of GN1**—TLC analysis of GN1 hydrolysate showed that the apparent size of the released oligosaccharides ranged from 2 to 6 degrees of polymerization (Fig. 1). These oligosaccharides almost exclusively contained Fuc, Gal, and GlcNAc with traces of Man, GlcA, and Py(4,6)Gal. Considering that Gal represents about 95% of the total number of hexoses detected, all hexose residues within the described sequences could be attributed to Gal. For the same reasons, dHex and HexNAc residues could be assigned to Fuc and GlcNAc monosaccharides, respectively. The fraction of released oligosaccharides was analyzed by ES/MS-MS in negative and positive modes before and after reduction with NaBH4 in order to observe the sulfate group to the GlcNAc residue in these three compounds.

**Glycan fingerprinting**—The structural fingerprinting of glyconectins from the three species by sequencing the hydrolyzed oligosaccharides. The smallest oligosaccharide, B, showed a very simple fragmentation pattern characterized by the simultaneous presence of the Y-type ions \([\text{M-HexNAc}]^+\) at \(m/z\) 282 in their fragmentation spectra permitted the localization of the sulfate group to the GlcNAc residue in these three compounds. The smallest oligosaccharide, B, showed a very simple fragmentation pattern characterized by a parent ion at \(m/z\) 446 (native) or 448 (reduced) and a B-type ion \([\text{M-dHex}]^-\) or \([\text{M-dHex-ol}]^-\) at \(m/z\) 282. This demonstrated unequivocally that the Fuc residue was in the terminal reducing position, which characterized compound B as the dimer \((\text{SO}_3^-)\text{GlcNAc-Fuc}\) or \((\text{SO}_3^-)\text{GlcNAc-\{Fuc\}Fuc}\). The fragmentation pattern of compound G was characterized by the simultaneous presence of the Y-type ions \([\text{M-HexNAc}]^+\) at \(m/z\) 592/594 and the B-type ion \([\text{M-dHex/dHex-ol}]^-\) at \(m/z\) 629. On this basis, compound G was described as an extension of the linear isomer of compound C with an additional GlcNAc residue in the terminal nonreducing position. In this case, and contrary to compound C, the absence of \([\text{M-dHex}]^-\) fragment excluded the possibility of a branched structure with a terminal nonreducing Fuc residue and demonstrated the presence of a single linear isomer.

Compounds D, E, and F differed from the previous compounds by the sulfation position on a hexose residue as indicated by the ubiquitous presence of the \((\text{SO}_3^-)\text{Hex-H}_2\text{O}]^+\) ion at \(m/z\) 241. The sequence of compound D was established as GlcNAc\((\text{SO}_3^-)\text{Gal-Fuc}\) due to the presence of \([\text{M-dHexol}]^-\) B ion and \([\text{M-HexNAc}]^-\) Y ion at \(m/z\) 444 and 407, respectively. The presence of branched isomers could be excluded because this sequence would not be able to generate the characteristic fragment with mass of 444. Similarly, for compound E, \([\text{M-HexNAc-ol}]^-\) B and C ions at \(m/z\) 403 and 421 demonstrated the
presence of a GlcNAc residue in the terminal reducing position. The occurrence of the Y ion [M-Hex] at m/z 464 established the sequence as Gal-(SO₃)Gal-GlcNAc. Finally, compound F appears to be an extension of compound E with a Fuc residue in the terminal reducing position, as demonstrated by the [M-dHexol] C ion at m/z 606. The simultaneous presence of [M-Hex], [M-dHexol-HexNAc] and [M-dHexol-Hex] at m/z 610, 403, and 444, respectively, confirmed the proposed sequence. As in the case of structures D and E, structure F cannot exist as the alternative branched isomer. In addition, the absence of (M-HexNAc) and (M-HexNAc-Hex) fragments confirmed the internal positioning of the HexNAc residue within the proposed sequence.

These results are in agreement with previously published data on the structure of GN1 glyconectin (12). In particular, two oligosaccharides with sequences SO₃(3)GlcNAc(β1-3)Fuc and Gal(α1-2)[SO₃(3)]Gal(β1-4)GlcNAc(β1-3)Fuc matched perfectly with compounds B and F. Importantly, three new motifs were discovered, namely the sequences GlcNAc-(SO₃)Gal-Fuc, (SO₃)GlcNAc-[Fuc]Fuc, and GlcNAc-Fuc-(SO₃)
GlcNAc-Fuc. The latter originated from the incomplete hydrolysis of Fuc-labile linkages.

Electrospray analysis in positive mode of the same fraction showed only a few components that could not be sequenced in ES/MS because of low quantities. In particular, a minor ion at m/z 622 was observed. Its calculated composition (1 PyGal, 1 HexNAc, and 1 dHex) matches the acid-labile trisaccharide Py(4,6)Gal-GlcNAc-Fuc previously described in GN1 (11). The low amounts of Py(4,6)Gal detected in the total oligosaccharide fraction confirmed that this oligosaccharide was a very minor component.

These data suggest that the GN1 acid-labile domain is not composed of a single type of repetitive and short units, as observed in glycosaminoglycans, but rather of more complex and heterogeneous polymeric sequences in which Fuc residues may be substituted by either GlcNAc, (SO$_3$)GlcNAc, or (SO$_3$)Gal and in which (SO$_3$)Gal may be substituted by either Gal or GlcNAc residues. The GN1 polysaccharide therefore presents a novel and highly heterogeneous structure.

**GN2 Sequencing**—Unlike GN1, partial hydrolysis of GN2 by 0.1 M trifluoroacetic acid did not release any small oligosaccharides but a material of greater size, which did not migrate in TLC (Fig. 1). This cleaved fraction contained large quantities of Gal (>40% of the total monosaccharides) and Py(4,6)Gal (>20%), as well as Man, GlcNAc, Fuc, and traces of GalNAc. The fucose content (7%) was much lower than observed in the GN1 hydrolysate (50%). Gel filtration analysis revealed that the released oligosaccharides ranged in size from 15 to 20 Glc degrees of polymerization (not shown). Such large sizes prevented us from determining their exact mass by MS analyses in MALDI-TOF or ES/MS. However, relevant structural data were inferred from other analytical techniques.

The nature of the reducing terminal monosaccharides was determined by KaB$_4$Na labeling and hydrolysis. TLC and HPAEC analyses revealed that only galactose and fucose were converted to polyols (Fig. 4). Tritiated Gal-ol was three times as abundant as Fuc-ol, suggesting that more oligosaccharides originated from the cleavage of galactose linkages than fucose linkages. This was in agreement with the low Fuc and very high Gal content.

The size of oligosaccharides released from GN2 and their high proportion of Py(4,6)Gal suggested that each molecule contained more than one of these residues. Thus, the oligosaccharides are either linear, with internal Py(4,6)Gal residues, or branched, with Py(4,6)Gal residues in terminal positions. To discriminate between these possibilities we performed multiple NMR analyses. On the HOHAHA spectrum, we observed a β anomic signal at δ 4.817 ppm, which correlated with a broad multiplet at δ 3.656–3.696 ppm and with a pseudo singlet at δ 4.155 ppm, which could be attributed, respectively, to H-2/H-3 and H-4 and H-6/6′ protons. Furthermore, the observation of an intense singlet at δ 1.465/26.32 ppm (1H/13C) was in accordance with the occurrence of a pyruvate group in the oligosaccharide mixture (data not shown). These 1H NMR parameters matched those previously described for terminal β-β-Py(4,6)Gal (16).

**Table II**

| Compound | m/z Native | m/z Reduced | Calculated composition | Proposed sequence | ES/MS-MS fragments [M−H]$^+$ | Inferred structure |
|----------|------------|-------------|------------------------|------------------|---------------------|------------------|
| A        | 300$^a$    |             | HexNAc (SO$_3$)HexNAc  | 97               | (SO$_3$)GlcNAc      |
| B        | 446        | 448$^a$     | HexNAc (SO$_3$)HexNAc  | 97; B, 282       | (SO$_3$)GlcNAc-Fuc  |
| C        | 592        | 594$^a$     | HexNAc (SO$_3$)HexNAc  | 97; Y 448; sec. frag., 282 | Fuc-(SO$_3$)GlcNAc-Fuc |
| D        | 608        | 610$^a$     | HexNAc (SO$_3$)HexNAc  | 97; B, 444; sec. frag., 241 | Gal-(SO$_3$)Gal-GlcNAc |
| E        | 624        | 626$^a$     | HexNAc (SO$_3$)HexNAc  | 97; B, 403; sec. frag., 241 | (SO$_3$)GlcNAc-Fuc  |
| F        | 770        | 772$^a$     | HexNAc (SO$_3$)HexNAc  | 97; B, 403; sec. frag., 241 | Gal-(SO$_3$)Gal-GlcNAc |
| G        | 793$^a$    | 795         | HexNAc (SO$_3$)HexNAc  | 97; B, 629; C, 647; Y, 592; sec. frag., 241 | (SO$_3$)GlcNAc-Fuc  |

$^a$ Fragment ions originate from either native or reduced material.
tion spectrum of this residue matched perfectly published data on the *Bacillus circulans* exopolysaccharide (20) and confirmed its identification as terminal Py(4,6)Gal. In accordance with NMR data, no substituted Py(4,6)Gal residue was observed in the methylation analysis. Results from composition, linkage, and NMR analyses demonstrated that the large oligosaccha-

![ES/MS-MS fragmentation patterns in negative mode.](image1)

**Fig. 3.** ES/MS-MS fragmentation patterns in negative mode. Native (A) and NaBH₄-reduced oligosaccharide C (B) from GN1 0.1 M trifluoroacetic acid hydrolysis at m/z 591.8 and 594.0, respectively, are shown. 281.9 (in A) and 282 (in B) are secondary fragments.

![Analysis of reducing monosaccharides from 0.1 M trifluoroacetic acid hydrolysis of GN2.](image2)

**Fig. 4.** Analysis of reducing monosaccharides from 0.1 M trifluoroacetic acid hydrolysis of GN2. The sample was reduced by KB₃H₄ and hydrolyzed by 4 M trifluoroacetic acid at 100 °C for 4 h. The resulting monosaccharides were analyzed by HPAEC (A) and TLC (B). Each fraction eluted from HPAEC was collected, and radioactivity was counted. TLC was autoradiographed.
rides released from glyconectins of GN2 by mild hydrolysis are branched structures, capped in equal amounts by terminal Py(4,6)Gal residues and Gal residues. Moreover, the fact that no trace of internal Py(4,6)Gal could be detected in the material released under mild hydrolysis conditions strongly suggests that Py(4,6)Gal residues are always found in terminal nonreducing position on the native glycans isolated from GN2.

To obtain further sequence information about GN2 glycans, recurrent partial hydrolysis of GN2 was performed in 1 M trifluoroacetic acid. Such hydrolysis conditions should preserve the pyruvate groups but not the possible sulfate groups. TLC analysis revealed that small quantities of oligosaccharide fragments were generated. They were separated from the large saccharidic material by gel filtration on a Bio-Gel P-4 (Bio-Rad). Half of the sample was not treated and the other half was reduced with NaBD₄ and then permethylated. Subsequently, both native and reduced permethylated oligosaccharides were analyzed by ES/MS-MS. Permethylated material was also analyzed in GC/MS. ES/MS experiments of the native material in negative mode did not show any ion, establishing that no sulfated oligosaccharides were present. These results suggested, as expected, that the putative sulfate groups were quantitatively removed by the hydrolysis. In contrast, ES/MS-MS in positive mode permitted sequencing of six oligosaccharides, H to M (Table III). Analysis of permethylated reduced oligosaccharides in positive mode gave the sequences of 15 compounds (Table IV). Only three major compounds (H, I, and J) could be sequenced in GC/MS (Table V) because of the low sensibility of the method. The data compiled from ES/MS-MS and GC/MS allowed the identification of 16 different oligosaccharides from trifluoroacetic acid 1 M hydrolysate of GN2. In ES/MS-MS positive mode, compounds were detected as [M+Na⁺] adducts. All native oligosaccharides showed intense [M-COOH] ions at m/z M-44 and [M-PyHex] Y-type ions at m/z M-232, whereas all permethylated oligosaccharides showed [M-PyHex] Y-type ions at m/z M-274, demonstrating that all compounds were substituted in the terminal position by at least one Py(4,6)Gal residue. Taking also into account the data obtained from analysis of material released under milder conditions, we concluded that oligosaccharides were always released from the extremities of the molecule and never originated from an internal domain.

Most compounds possessed either Hex or HexNAc residues in terminal reducing positions. Compounds I, J, K, and M had unsubstituted terminal-reducing HexNAc residues as shown by ES/MS-MS experiment of native oligosaccharides, because of [M-HexNAc] C ion at m/z M-203, [M-HexNAc-COOH] C ion at m/z M-247, and [HexNAc] Y ion at m/z 244. This was confirmed by observing [M-HexNAcol] C and [HexNAcol] Y ions at m/z M-276 and m/z 317, respectively, after reduction and permethylation of oligosaccharides (Fig. 6). Similarly, unsubstituted terminal Hex residues were characterized in compounds H, P, and W due to the presence of [M-Hex] C ion at m/z M-162, [M-Hex-COOH] C ion at m/z M-206, and [Hex] Y ion at m/z 203.
Mass spectrometry fingerprinting of GN2 native fragment oligosaccharides

Composition and sequences deduced from the ES/MS sequencing in positive mode of 1 M trifluoroacetic acid-hydrolyzed native fragments isolated from GN2. sec. frag., secondary fragments.

| Compound | m/z native | Calculated composition | Proposed sequence | ES/MS-MS fragments [M+Na]+ |
|----------|------------|------------------------|-------------------|-----------------------------|
| H        | 435        | PyHex                  | PyHex-Hex         | [M-44], 391; [C-44], 229; Y, 203 |
| I        | 476        | PyHex                  | PyHex-HexNAc      | [M-44], 432; [C-44], 229; Z, 226; Y, 244 |
| J        | 638        | PyHex                  | PyHex-HexHexNAc   | [M-44], 594; B, 416; [B-44], 373; C, 434; [C-44] 391; Z, 226, 388, Y, 244, 406 |
| K        | 800        | PyHex                  | PyHex-Hex-HexNAc  | [M-44], 756; [B-44], 373, 535; C, 597; [C-44], 391, 553; Z, 226, 388, 550; Y, 244, 406, 568 |
| L        | 800        | PyHex                  | PyHex-[Hex]HexNAc | [M-44], 756; [B-44], 373, C, 203; [C-44], 391; Z, 388, 550, 620; [Z-44], 576; Y, 406, 568, 638; [Y-44], 594 |
| M        | 902        | PyHex                  | PyHex-Hex-Hex-HexNAc | [M-44], 918; [B-44], 525, 697; C, 759; [C-44], 391, 553, 715; Z, 388, 550, 712, Y, 244, 406, 568, 730; sec. frag., 347, 365, 599, 527 |

Mass spectrometry fingerprinting of GN2 permethylated fragment oligosaccharides

Composition and sequences deduced from the ES/MS sequencing in positive mode of 1 M trifluoroacetic acid-hydrolyzed, NaBD₄-reduced, permethylated fragments isolated from GN2.

| Compound | m/z Reduced | Calculated composition | Proposed sequence | ES/MS-MS fragments [M+Na]+ |
|----------|-------------|------------------------|-------------------|-----------------------------|
| N        | 520         | PyHex                  | PyHex-dHex        | Y, 246 |
| H        | 550         | PyHex                  | PyHex-Hex         | Y, 276 |
| I        | 591         | PyHex                  | PyHex-HexNAc      | [M-59], 532; B, 299; C, 315; Y, 317 |
| O        | 724         | PyHex                  | PyHex-Hex-dHex    | B, 501; C, 519; Y, 246, 450 |
| P        | 754         | PyHex                  | PyHex-Hex-Hex     | [M-59], 695; B, 501; C, 519; Y, 276, 480 |
| Q        | 754         | PyHex                  | PyHex-[Hex]Hex    | [M-59], 695, Y, 480, 536 |
| J        | 785         | PyHex                  | PyHex-HexHexNAc   | B, 501, C, 519; Z, 299, 503; Y, 317, 521 |
| R        | 958         | PyHex                  | PyHex-[Hex]Hex    | Y, 480, 684, 740 |
| K        | 999         | PyHex                  | PyHex-HexHexHexNAc| B, 501, 705; C, 519, 723 |
| L        | 999         | PyHex                  | PyHex-[Hex]HexNAc | B, 501; C, 519; Y 781 |
| S        | 1162        | PyHex                  | PyHex-Hex-Hex-[Hex]Hex | Y, 480, 684, 888, 944 |
| T        | 1162        | PyHex                  | PyHex-Hex-[PyHex]Hex-Hex | Y, 944, 740, 888; sec. frag., 465 |
| U        | 1218        | PyHex                  | PyHex-Hex-[PyHex]Hex-Hex | Y, 944, 740; sec. frag., 465 |
| V        | 1366        | PyHex                  | Hex-Hex-Hex-[PyHex]Hex | Y, 1148, 1092, 944, 740, 538 |
| W        | 1366        | PyHex                  | PyHex-Hex-Hex-Hex-Hex | Y, 1092, 888, 684, 276 |

in native oligosaccharides, as well as [Hexol] Y ion at m/z 276 and [M-Hexol] C ion at m/z 235 in permethylated reduced oligosaccharides. In the case of compounds T and U, Hex-ol residues were identified because the secondary fragment at m/z 465 distinguished the substituted dimer Hex-ol. Thus, sequences of the two simplest compounds, H and I, were easily established in ES/MS-MS as PyHex-Hex and PyHex-HexNAc, respectively. In the GC/EI-MS spectrum of compound H, ions deriving from fragmentation of the Hex-ol residue (m/z at 89, 90, 134, 437, and 480) demonstrated that the PyHex residue was substituted for the terminal hexose residue in the C-4 position. Similarly, for compound I, fragmentation of the Hex-NAc-ol residue (ions at m/z 89, 131, 435, 479, 523, and 524) clearly established that it was substituted in the C-3 position. For compound J, the fragmentation pattern established that the HexNAc-ol residue was substituted in the C-3 position (Table V, Fig. 7). Identification of the J₀ fragment at m/z 323 indicated that the internal Hex residue was substituted in the C-3 substitution as shown previously. All other compounds appeared as linear or branched variations of oligosaccharides H and I with additional hexose residues as shown in Table IV. They differed by the number of hexose residues and the position of substitution in the case of branched oligosaccharides. Branched oligosaccharides (L, Q, R, and S) were identified...
because of the simultaneous loss of PyHex and Hex residues. Hex residues in terminal nonreducing positions were distinguished from Hex residues in reducing position after reduction and permethylation through the presence of \([\text{M-Hex}]^-\) ions at \(m/z\) M-218. Branching position was deduced from recurring Y-type and secondary fragmentations of the oligosaccharides. According to the calculated composition, compound U was the only compound to have two PyHex residues. Successive \(m/z\) losses of 274 demonstrated that both residues were in terminal positions. Although branched oligosaccharides were often accompanied by their linear counterparts, both types of isomers could readily be identified according to their different fragmentation patterns.

Two minor compounds (N and O) were observed in ES/MS only as permethylated derivatives (Table IV). Both presented dHex residue in terminal reducing positions, as shown by the Y

| Compound | \([\text{M+H}]^+\) | Calculated composition | Proposed sequence |
|----------|----------------|-----------------------|------------------|
| H        | 528            | PyHex                 | Py(4,6)Gal(1–4)/Hex |
| I        | 569            | PyHex                 | Py(4,6)Gal(1–3)HexNAc |
| J        | 773            | PyHex                 | Py(4,6)Gal(1–3)Hex(1–3)HexNAc |

**Fig. 6.** ES/MS-MS positive mode fragmentation pattern of oligosaccharide J from GN2 after 1 M trifluoroacetic acid hydrolysis. Native oligosaccharide (A) and NaBD₄-reduced and permethylated oligosaccharide (B) are shown.
Furthermore, for compound O, the Y ion at m/z 450 demonstrated the occurrence of the dimer PyHex-dHex in the reducing position, whereas B and C ions at m/z 501 and 519 demonstrated the occurrence of the dimer PyHex-Hex in terminal nonreducing positions. These data established the sequence of N as PyHex-dHex and O as PyHex-Hex-dHex.

The data compiled from structural analyses conducted on GN2 oligosaccharides released in two different hydrolysis conditions strongly suggest that the acid-labile domain of this molecule was constituted by a highly heterogeneous, branched polymer. The saccharidic core is mostly constituted by a polymer of Gal, Man, GlcNAc, and Fuc residues in internal position, ending with Gal and Py(4,6)Gal residues. In conclusion, composition, linkage, and NMR analyses demonstrated that glycan fragments released by mild hydrolysis of GN2 were large, highly branched oligosaccharides, capped equally by terminal Py(4,6)Gal residues and Gal residues.

**GN3 Sequencing**—GN3 hydrolysate differed notably from the two previous samples in regard to the unique presence of arabinose among the fucose residues (Tables I and VI). In addition, GN3 is the only species showing significant amounts of GalNAC but not Py(4,6)Gal. TLC analysis revealed that this material was constituted by a mixture of small oligosaccharides with sizes comparable with those hydrolyzed from GN1. Thus, as described for GN1, this material was analyzed in ES/MS-MS, first as native and second as reduced (NaBD₄) oligosaccharides. Only [M − H]⁻ adducts were produced, because of the ubiquitous presence of sulfate groups.

Eleven oligosaccharides (compounds X to AH; see Table VI), out of which nine were sequenced, were observed in negative mode, whereas no oligosaccharides were detected in positive mode. As in GN1, the fragmentation pattern of each compound in negative mode was characterized by the presence of a very intense ion at m/z 97, indicating that they were all sulfated. Sulfate groups were positioned either on a dHex residue because of the intense Z and Y ions at m/z 225 and 243, respectively, or on a Pent residue because of the intense Z and Y ions at m/z 211 and 229, respectively. Considering both the fragmentation patterns and composition analysis, it was deduced that compounds X, Z, AB, AD, AF, and AG contained sulfated Ara residues, whereas compounds Y, AA, AC, and AE contained sulfated Fuc residues.

Compounds X and Y were the smallest oligosaccharides sequenced from GN3. From their calculated composition and fragmentation patterns, it was deduced that X was composed of one HexNAc and one sulfated Pent residue and that Y was composed of one HexNAc and one sulfated dHex residue. In contrast to all the other sequenced compounds, reduction by NaBD₄ did not affect these two oligosaccharides. They had identical fragmentation patterns before and after reduction. A possible explanation is that sulfation of the monosaccharide in the terminal reductive position prevents reduction of...
that (SO\(_3\))dHex and (SO\(_3\))Pent residues are in reducing static interactions. On the basis of our results, we propose the [M-dHexol-HexNAc] B and C ions at

for either native or reduced oligosaccharide. We observed only

cence of two consecutive HexNAc residues; this was not the case

for the oligosaccharide by NaBD\(_4\), presumably through electro -

case of [M-dHexol] B ion at
dHex to the reducing position was done for compound AA,

the occurrence of [M-HexNAc] Z ions. Direct attribution of

nonreducing position was demonstrated for both compounds by

were unclear, but the presence of HexNAc in the terminal

position for AB and AC. The fragmentation spectra of Z and AA

or an additional HexNAc residue in the reducing position.

cause of the presence of intense [M-HexNAc] Y ions at

m/z 225 and 243.

Composition and sequences are deduced from the ES/MS sequencing in negative mode of 0.1 M trifluoroacetic acid-hydrolyzed native and NaBD\(_4\)-reduced oligosaccharides isolated from GN3. sec. frag., secondary fragments.

| Compound | m/z Native | m/z Reduced | Calculated composition | Proposed sequence | ES/MS-MS fragments [M-H] |
|----------|------------|-------------|------------------------|-------------------|-------------------------|
| X        | 432        | 432         | HexNAc Pent SO\(_3\) | HexNAc-(SO\(_3\))Pent | 97; Z, 211; Y, 229 |
| Y        | 446        | 446         | HexNAc dHex SO\(_3\) | HexNAc-(SO\(_3\))dHex | 97; Z, 225; Y, 243 |
| Z        | 578        | 578         | HexNAc dHex, Pent SO\(_3\) | HexNAc-(SO\(_3\))Pent-dHex | 97; Z, 357; sec. frag., 211 |
| AA       | 592        | 595         | HexNAc 2xHexSO\(_3\) | HexNAc-(SO\(_3\))dHex-dHex | 97; B, 428; Z, 375; sec. frag., 225 |
| AB       | 635        | 638         | 2xHexNAc Pent SO\(_3\) | HexNAc-(SO\(_3\))Pent-HexNAc | 97; Y, 435; sec. frag., 211 |
| AC       | 649        | 652         | 2xHexNAc dHex SO\(_3\) | HexNAc-(SO\(_3\))dHex-HexNAc | 97; Y, 449; sec. frag., 225 |
| AD       | 781        | 784         | 2xHexNAc Pent, dHex SO\(_3\) | HexNAc-(SO\(_3\))Pent-HexNAc-dHex | 97; B, 414, 617; C, 432; Y, 581 sec. frag., 211 |
| AE       | 795        | 798         | 2xHex SO\(_3\) | HexNAc-(SO\(_3\))dHex-HexNAc-dHex | 97; B, 428, 631; C, 446; Y, 595; sec. frag., 225 |
| AF       | 943        | 946         | 2xHexNAc Hex, dHex, Pent SO\(_3\) | Hex-HexNAc-(SO\(_3\))Pent-HexNAc-dHex | 97; B, 779, 576; Y, 784, 581 sec. frag., 211, 414 |
| AG       | 957        | 960         | 2xHexNAc Hex, 2xHexSO\(_3\) | Hex-HexNAc-(SO\(_3\))Pent-HexNAc | 97 |
| AH       | 984        | 987         | 3xHexNAc Pent, dHex SO\(_3\) | [SO\(_3\)Pent + 2xHexNAc + dHex]-HexNAc | 97; B, 763; sec. frag., 211, 617, 560 |

DISCUSSION

This work presents the first comparative data on the glycan moieties of glycoconnectins isolated from the three model sponge species M. prolifera (GN1), H. panicea (GN2), and C. celata (GN3). We focused on the analyses of oligosaccharides from the acid-sensitive domains of these highly heterogeneous glycoconjugates using a combination of gas chromatography, electros-
pray mass spectrometry, and NMR methods. The results clearly established that the glycan moieties of the three glyconectins have strictly species-specific structures. They differed in the following ways: (i) in monosaccharide composition and, in particular, in the presence of Fuc, Ara, and Py(4,6)Gal residues; (ii) in susceptibility to mild acid hydrolysis as reflected by the quantities of liberated material, ranging from 41% (mol/mol) to 68% of the total glycan moieties; (iii) in the size of the liberated oligosaccharides under mildly acidic conditions, 1–4 monosaccharides for GN1, superior to 10 for GN2, and between 1 and 5 for GN3; (iv) in the nature of the negative charge (essentially sulfate groups for GN1 and GN3 and carboxyl groups for GN2) and the position of the sulfate group (Gal and GlcNAc for GN1, Ara and Fuc for GN3); and (v) in the sequence of liberated oligosaccharides. Taking into account only the longest identified oligosaccharides, we summarized the structures of the acid-labile glycan motifs present on each glyconectin in Table VII. As previously observed for GN1, the carbohydrate moieties of the three glyconectins analyzed are constituted by a repetition of heterogeneous glycan motifs distinct from the disaccharidic units found in glycosaminoglycans. In addition to their intrinsic heterogeneity, glyconectins exhibit a large variability according to the species, whereas most glycosaminoglycan structures are highly conserved throughout evolution. The present analysis of GN1 glycan moieties is in agreement with previous work describing four oligosaccharides (14, 15). According to their reactivity toward the cell adhesion blocking antibodies Block 1 and Block 2, the pyruvilated triasaccharide Py(4,6)Gal-GlcNAc-Fuc and the sulfated disascharide SO_3-3GlcNAcβ1–3-Fuc were previously shown to be involved directly in the aggregating properties of GN1 (1, 2). The synthetic form of the sulfated unit was revealed to self-associate in plasmon resonance kinetic binding experiments (3). The further identification of the oligosaccharides Fuc-[SO_3]GlcNAc-Fuc and GlcNAc-Fuc-[SO_3]GlcNAc-Fuc established that the SO_3-3GlcNAcβ1–3-Fuc motif may be part of a larger repetitive unit. However, the extreme susceptibility of the fucose bond prevented observation of the complete repetition unit. This result strongly suggests that the glyconectin-glyconectin interaction in GN1 not only may take place through the terminal nonreducing end of their glycan moieties but also may involve many internal epitopes. Sequencing of released oligosaccharides did not reveal the existence of the internal Py(4,6)Gal-GlcNAc-Fuc motif. In accordance with this finding, NMR analysis conducted on total glyconectins or released intact glycan showed that external Py(4,6)Gal was present in minute amounts, whereas no trace of internal Py(4,6)Gal was detected (data not shown). Furthermore, we observed the novel motifs GlcNAc-[SO_3]GlcNAc-Fuc and GlcNAc-Fuc-[SO_3]GlcNAc-Fuc; their potential involvement in the auto-aggregation properties of GN1 remains to be determined.

Although the overall sequences of the acid-labile carbohydrate moieties isolated from M. prolifera (GN1) and C. celata (GN3) were distinct from one another, a few fragments such as the G compound in GN1 [GlcNAc-Fuc-[SO_3]GlcNAc-Fuc] and the AE compound in GN3 [HexNAc-[SO_3]Fuc-HexNAc-Fuc] did show some structural similarities. The carbohydrate moiety from GN3 appeared to be the most homogeneous of the three species analyzed, although only two possible repetitive tri- and tetrasaccharidic motifs were found, and in these sulfated Ara and sulfated Fuc residues are integrated in an identical fashion.

Acid hydrolysis of glyconectin glycans into two fractions with distinct monosaccharide compositions showed that the carbohydrate moieties of GNs are constituted by at least two unrelated domains that are either acid-resistant or acid-labile. For GN1 and GN3, the increase in GlcA and decrease in Fuc/Ara of the acid-resistant fraction may in part account for their distinctive properties. However, the overall organization of both domains is still unknown. The observation that g6 and g200 glycans isolated from M. prolifera glyconectin contain similar ratios of GlcA (2, 21) suggests that both g6 and g200 are heterogeneous molecules having both types of domains. Based on the available data, we propose two possible models of organization for GN1 and GN3 carbohydrate moieties within the glycan chain. The first model represents a high molecular weight, linear, acid-sensitive polysaccharide connected to an acid-resistant domain. This polysaccharide may be composed of either heterogeneous short repetitive units or a large homogeneous repetitive unit comprising acid-labile glycosidic bonds (Fuc/Ara). The actual size of a putative homogeneous repetitive unit is difficult to assess because most Fuc- and Ara-glycosidic linkages are cleaved under mildly acidic conditions. The second model represents a mixed, ramified polysaccharide composed of an acid-resistant core connected through Fuc/Ara residues to small oligosaccharides that are released by mild acidic hydrolysis.

In contrast with GN1 and GN3, mild hydrolysis of GN2 released large oligosaccharides that were further fragmented in smaller units using stronger acidic conditions. Analysis of both fractions revealed that the acid-labile carbohydrate moiety of GN2 comprised a highly ramified polysaccharide backbone. It is constituted by an extremely heterogeneous mixture of hexose (mannose and galactose) oligomers all terminated by Py(4,6)Gal residues and randomly interrupted by Fuc and GlcNAc residues. The observed heterogeneity of released oligosaccharides did not permit us to draw definitive conclusions about the ultrastructural organization of repetitive motifs.

Taking into account the previous results on GN1, -2, and -3 carbohydrate-mediated adhesion and the fact that GN1 functions via carbohydrate to carbohydrate associations, we suggest that GN2 and GN3 may also associate through a similar molecular mechanism involving carbohydrate-carbohydrate interactions. In this proposition, the structural differences that we observed between the three glyconectins may account for their species-specific self-recognition. This hypothesis would further imply that the specific carbohydrate sequences of GNs...
Structural Fingerprinting of Glyconectin Carbohydrates

regulate the cell recognition and adhesion functions. Based on the present structural study, the validity of this assumption will be assessed by testing the involvement of isolated oligosaccharides in the aggregation process.

Acknowledgments—We thank E. Maes for NMR analysis and V. Norris for helpful discussions.

REFERENCES
1. Misevic, G. N., Finne, J., and Burger, M. M. (1987) J. Biol. Chem. 262, 5870–5877
2. Misevic, G. N., and Burger, M. M. (1993) J. Biol. Chem. 268, 4922–4929
3. Haseley, S. R., Vermeer, H. J., Kamerling, J. P., and Vliegenthart, J. F. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 9419–9424
4. Eggens, I., Fenderson, B. A., Toyokuni, T., and Hakomori, S. (1989) Biochem. Biophys. Res. Commun. 158, 913–920
5. Eggens, I., Fenderson, B., Toyokuni, T., Dean, B., Stroud, M., and Hakomori, S. (1989) J. Biol. Chem. 264, 9476–9484
6. Kojima, N., and Hakomori, S. (1989) J. Biol. Chem. 264, 20159–20162
7. Kojima, N., Shiota, M., Sadahira, Y., Handa, K., and Hakomori, S. (1992) J. Biol. Chem. 267, 17264–17270
8. Kojima, N., Fenderson, B. A., Stroud, M. R., Goldberg, R. I., Habermann, R., Toyokuni, T., and Hakomori, S. (1994) Glycoconjug. J. 11, 228–248
9. Rabinovich, G. A., Rubinstein, N., and Toscano, M. A. (2002) Biochim. Biophys. Acta 1572, 274–284
10. Lowe, J. B. (2002) Immunol. Rev. 186, 19–36
11. Spillmann, D., Hard, K., Thomas-Oates, J., Vliegenthart, J. F., Misevic, G., Burger, M. M., and Fazal, J. (1993) J. Biol. Chem. 268, 13378–13387
12. Spillmann, D., Thomas-Oates, J. E., van Kuik, J. A., Vliegenthart, J. F., Misevic, G., Burger, M. M., and Finne, J. (1995) J. Biol. Chem. 270, 5089–5097
13. Cuculau, I., and Kerek, F. (1984) Carbohydr. Res. 131, 209–217
14. Kamerling, J. P., Gerwig, G. J., Vliegenthart, J. F., and Clamp, J. R. (1975) Biochem. J. 151, 491–495
15. Zanetta, J. P., Timmerman, P., and Leroy, Y. (1999) Glycoconjugate J. 9, 255–266
16. Yang, B. Y., Brand, J., and Montgomery, R. (2001) Carbohydr. Res. 331, 59–67
17. Bock, K., and Pedersen, C. (1983) Adv. Carbohydr. Chem. Biochem. 41, 27–66
18. Cerantonio, S., Marti, N., and Montrozier, H. (1996) Carbohydr. Res. 285, 59–67
19. Van Calsteren, M. R., Pau-Roblot, C., Begin, A., and Roy, D. (2002) Biochem. J. 363, 7–17
20. Fontaine, T., Wieruszeski, J. M., Talmont, F., Saniez, M. H., Duflot, P., Leleu, J. B., and Fournet, B. (1991) Eur. J. Biochem. 196, 107–113
21. Misevic, G. N., and Burger, M. M. (1990) J. Biol. Chem. 265, 20577–20584
22. Misevic, G. N., Guerrardel, Y., Sumanovski, L. T., Slomianny, M.-C., Demarty, M., Ripoll, C., Karamanos, Y., Maes, E., Popescu, O., and Strecker, G. (2004) J. Biol. Chem. 279, 15579–15590
