The nematode *Caenorhabditis elegans* synthesizes unusual O-linked glycans: identification of glucose-substituted mucin-type O-glycans and short chondroitin-like oligosaccharides

Yann GUÉRARDEL^{*1}, Luis BALANZINO[†], Emmanuel MAES^{*}, Yves LEROY^{*}, Bernadette CODDEVILLE^{*}, Rafael ORIOL[†] and Gérard STRECKER^{*}

^{*}Laboratoire de Chimie Biologique et Unité Mixte de Recherche du CNRS 8576, Université des Sciences et Technologies de Lille, F-59655 Villeneuve d'Ascq, France, and [†]INSERM U 504/Université Paris Sud XI, Villejuif, France

The free-living nematode *Caenorhabditis elegans* is a relevant model for studies on the role of glycoconjugates during development of multicellular organisms. Several genes coding for glycosyltransferases involved in the synthesis of N- and O-linked glycans have already been isolated, but, apart from repetitive dimers of glycosaminoglycans, no detailed structure of either type of component has been published so far. This study aimed to establish the structures of the major O-glycans synthesized by *C. elegans* to give an insight into the endogenous glycosyltransferase activities expressed in this organism. By the use of NMR and MS, we have resolved the sequence of seven of these components that present very unusual features. Most of them

were characterized by the type-1 core substituted on Gal and/or GalNAc by (β 1-4)Glc and (β 1-6)Glc residues. Another compound exhibited the GalNAc(β 1-4)*N*-acetylglucosaminitol sequence in the terminal position, to which was attached a tetramer of β -Gal substituted by both Fuc and 2-*O*-methyl-fucose residues. Our experimental procedure led also to the isolation of glycosaminoglycan-like components and oligomannosyl-type N-glycans. In particular, the data confirmed that *C. elegans* synthesizes the ubiquitous linker sequence GlcA(β 1-3)Gal(β 1-3)Gal(β 1-4)Xyl.

Key words: glycosaminoglycan, mass spectrometry, NMR.

INTRODUCTION

Until now, most studies of N- and O-glycosylation in nematodes concentrated on the identification of immunogenic glycoconjugates of parasitic nematodes. Indeed, it has been suggested that most parasite antigens are either excreted-secreted (ES antigens) or surface glycans with unique structures [1]. Systematic sequencing of glycans isolated from parasitic nematodes demonstrated that these organisms synthesize a wide variety of N-glycans presenting unusual substitutions. They include phosphocholine- and chito-oligomer-substituted glycans in filarial nematodes [2], D-tyvelose capped glycans in *Trichinella spiralis* [3] or highly fucosylated chitobiose core structures in *Haemonchus contortus* [4].

The only O-linked mucin-type glycans described in nematodes so far were isolated from the *Toxocara* excretory-secretory (TES) product released by the infective larval stage of *Toxocara canis* and *T. cati* [5]. The two major structures were trisaccharidic O-glycans characterized by the presence of 2-*O*-methyl-fucose (2-*O*-Me-Fuc) and 4-*O*-methyl-galactose.

The free-living nematode *Caenorhabditis elegans* has been recognized as a good model for the study of the role of glycans in eukaryotes as well as a prototypic model for parasitic nematodes because of the detailed information available on its genome, development and physiology. So far, only glycolipid-

[6,7] and glycosaminoglycan (GAG)-type structures have received proper attention. In particular, the presence of large amounts of unsulphated chondroitin and minute amounts of heparan sulphate-type GAGs have been identified in *C. elegans* [8,9].

No detailed structures of either O- or N-linked oligosaccharides isolated from *C. elegans* have been published. However, the identification of three genes (Gly-12, -13 and -14) homologous to mammalian N-acetylglucosamine: α -3-D-mannoside β -1,2-N-acetylglucosaminyltransferase (GnT1) suggests the presence of N-linked oligosaccharides in *C. elegans* [10]. Gly-12 and Gly-14 were shown to encode active glycosyltransferases, but only mutation of Gly-13 induced defects in embryogenesis, whereas Gly-12- and Gly-14-null mutants displayed wild-type phenotypes [11]. Similarly, *C. elegans* expresses a family of functional polypeptide N-acetylgalactosaminyltransferases (ppGaNTases), involved in the initiation of mucin-type O-glycosylation [12]. Inhibition of these ppGaNTases by antisense RNAs resulted in severe defects in embryonic development [13]. Thus, these data not only suggest the occurrence of O-linked glycans but also suggest that this type of glycosylation plays a critical role in the embryogenesis of *C. elegans*.

Also, in agreement with abundant N- and O-glycosylation in *C. elegans*, 30 different putative fucosyltransferase genes, including 24 α -2, five α -3 and one α -6 fucosyltransferase gene, have

Abbreviations used: 2-*O*-Me-Fuc, 2-*O*-methyl-fucose; GalNAc-ol, *N*-acetylgalactosaminitol; GlcNAc-ol, *N*-acetylglucosaminitol; Xyl-ol, xylitol; Δ HexA, 4-deoxy- α -L-threo-hex-4-enopyranosyluronic acid; ESI-MS, electrospray ionization MS; TES, *Toxocara* excretory-secretory; GAG, glycosaminoglycan; MALDI-TOF, matrix-assisted laser-desorption ionization-time-of-flight; ROESY, rotating-frame Overhauser enhancement spectroscopy; NOE, nuclear Overhauser effect; HMQC, heteronuclear multiple quantum coherence; β -GlcP, β -glucopyranose; β -GalP, β -galactopyranose; Hex, hexose; HexA, hexuronic acid; HexNAc-ol, *N*-acetylhexosaminitol.

¹ To whom correspondence should be addressed (e-mail Yann.Guerardel@univ-lille1.fr).

been identified in databanks ([14] and R. Oriol, unpublished work). One of these, CEFT-1, has been cloned; it can fucosylate unsialylated type 2 acceptors to generate the Le^x antigen *in vitro*, although *C. elegans* does not express such an antigen [15]. This example points out the difficulties of characterizing the exact activity of an isolated glycosyltransferase without any knowledge of its endogenous glycan substrate.

This work aimed to define the structure of major O-glycans synthesized by *C. elegans*, to give insight into the endogenous activity of glycosyltransferases cloned from this worm and to facilitate studies about their exact role during embryogenesis. Considering the extreme structural diversity of O-glycans, and the absence of real conserved motifs, the analysis of their detailed structure in a new model may be very complex. Thus, in this first study, we have worked on a total extract of a mixed-stage wild-type strain to obtain enough material for multiple NMR experiments. O-linked glycans were released from the *C. elegans* extract by a reductive alkaline treatment. Through the use of homo- and heteronuclear NMR experiments, electrospray ionization MS (ESI-MS) and GC/MS analyses, we identified two families of unusual mucin-type O-glycans. Both families were characterized by the original presence of multiple β -Glc residues in non-reducing terminal positions. By analysing the GAG chains released by β -elimination, this study confirmed that *C. elegans* synthesizes the linker sequence GlcA(β 1-3)Gal(β 1-3)Gal(β 1-4)Xyl-ol (where Xyl-ol is xylitol) that initiates the biosynthesis of most GAGs, except dermatan sulphate and hyaluronate. The identification of short chondroitin-type oligosaccharides with a GalNAc residue in the reductive terminal position also raised the possibility of GAG chains directly linked to the protein moieties through a GalNAc. The release of oligosaccharides in mild alkali also permitted us to isolate mannosyl-type N-glycans synthesized in large quantities by *C. elegans*.

EXPERIMENTAL

C. elegans cultures

C. elegans wild-type strain N2 Bristol was grown in either solid or liquid cultures. Solid cultures were achieved by growing the nematodes on NGM agar plates seeded with *Escherichia coli* OP50, a leaky uracil-requiring strain [16]. For yields of worms that are in the range of grams, *C. elegans* is conveniently grown in bulk by using *E. coli* strain NA22 as a food source [17]. Briefly, the bacteria were inoculated in 2 l flasks containing 500 ml of 3XD medium (10.5 g/l of Na₂HPO₄, 4.5 g/l of KH₂PO₄, 0.6 g/l of NH₄Cl, 15 g/l of casein hydrolysate, 24 g/l of glycerol and 3 ml/l of 1 M MgSO₄) [18], and incubated in an orbital shaker overnight at 37 °C and 250 rev./min. The bacteria were harvested by spinning for 15 min at 4500 g in sterile 500 ml centrifuge bottles. The supernatant was discarded and the bacteria weighed.

The bacterial pellets were resuspended in 500 ml of S complete medium (5.9 g/l NaCl, 50 ml/l of 1 M potassium phosphate, pH 6.0, 5 mg/l cholesterol, 10 ml/l of 1 M potassium citrate, pH 6.0, 10 ml/l trace metals solution, 3 ml/l of 1 M CaCl₂ and 3 ml/l of 1 M MgSO₄) [19] at a final density of 30–50 g/l in 2 l flasks. Then, worms were washed off with sterile M9 buffer (6 g/l Na₂HPO₄, 3 g/l KH₂PO₄, 5 g/l NaCl and 0.25 g/l MgSO₄·7H₂O) [16] from fresh, uncontaminated solid cultures and inoculated to the bacteria.

The cultures were incubated in an orbital shaker at 20 °C and 200 rev./min for several days until they were clear of bacteria and had ceased to grow. The cultures were collected and allowed to settle overnight at 4 °C into 500 ml glass cylinders. Most of the medium was then aspirated, and the worms were transferred to 50 ml Falcon tubes, pelleted by spinning for 5 min at 1000 g, and

washed three times with 50 ml of M9 buffer. Then nematodes were harvested and decontaminated by sucrose floatation [17]. Briefly, the washed worms were resuspended in 50 ml Falcon tubes in 25 ml of ice-cold 0.1 M NaCl, and then 25 ml of ice-cold 60% (w/w) sucrose was added and mixed by inversion. Quickly, the tubes were centrifuged at 1500 g for 10 min and the cap of floating nematodes was collected using a Pasteur pipette and distributed into two 15 ml pre-weighed Falcon tubes. The worms were washed three times with M9 buffer, weighed and stored at –80 °C. Freshly cleared cultures produced a yield of worms equal to about half the weight of the bacteria used [20].

Some nematode strains used in this work were provided by the *Caenorhabditis* Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources (NCRR), Buenos Aires, Argentina.

Isolation of oligosaccharide-alditols

Freeze-dried *C. elegans* were ground and repeatedly delipidated by chloroform/methanol (2:1) extraction. The delipidated material was submitted to alkaline reductive degradation in 100 mM NaOH containing 1.0 M sodium borohydride at 37 °C for 72 h. The reaction was stopped by addition of Dowex 50 × 8 cation-exchange resin (25–50 mesh, H⁺ form) at 4 °C until pH 6.5, and after evaporation to dryness, boric acid was distilled as methyl ester in the presence of methanol. Total material was submitted to cation-exchange chromatography on a Dowex 50 × 2 column (200–400 mesh, H⁺ form) to remove residual peptides and desalted on a Bio-Gel P2 column (Bio-Rad). The oligosaccharide fraction was then purified from macromolecular material on a Bio-Gel P6 column (Bio-Rad).

Fractionation of oligosaccharide-alditols

Compounds were fractionated by HPLC on a primary amine-bonded silica column (Supelcosyl, LC-NH₂, 4.6 × 250 mm; Supelco, Bellefonte, PA, U.S.A.) using a mixture of acetonitrile/30 mM KH₂PO₄/water (80:0:20–50:50:0 in 65 min) at a flow rate of 1 ml/min. Oligosaccharides were detected by UV spectroscopy at 206 nm using a variable-wavelength detector (Spectra Monitor D, Milton Roy, Riviera Beach, FL, U.S.A.) connected to a Spectra-Physics Model 4100 computing integrator.

Matrix-assisted laser-desorption ionization–time-of-flight (MALDI-TOF) MS

The molecular mass of the oligosaccharides was measured by MALDI on a Vision 2000 time-of-flight instrument (Finnigan Mat) equipped with a 337 nm UV laser. Samples were dissolved in water at a concentration of 100 pmol/μl; 1 μl of the solution was mixed with an equal volume of the matrix solution on the target and then allowed to crystallize at room temperature. We used 2,5-dihydroxybenzoic acid (10 mg/ml; dissolved in water/methanol, 4:1, v/v) and 3-aminoquinoline for the neutral and acidic oligosaccharides, respectively.

ESI-MS

All MS measurements were carried out in positive-ion mode on a triple quadrupole instrument (Micromass, Altrincham, Cheshire, U.K.) fitted with an atmospheric pressure ionization electrospray source. A mixture of polypropylene glycol was used to calibrate the quadrupole mass spectrometer. The sample (2 μl), dissolved in methanol/water (50:50) and 1% formic acid

at a concentration of 10 pmol/ μ l, was infused using the nanoflow probe at 50 nl/min. Quadrupole was scanned from 600 to 1800 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 set at 70 V. For collision-induced dissociation (CID) experiments, the pressure of argon in the cell was set at 4×10^{-3} mbar and the collision energy was set to values ranging from 40 to 75 eV.

NMR spectroscopy

The NMR experiments were performed on Bruker® ASX400 and DMX600 spectrometers both equipped with a 5 mm $^1\text{H}/^{13}\text{C}$ mixed-probe head, operating in the pulse Fourier-transform mode controlled by an Aspect 3000 computer. Each oligosaccharide was dissolved in 400 μ l of $^2\text{H}_2\text{O}$ after three exchanges with $^2\text{H}_2\text{O}$ at pH 7 (99.97% atom ^2H , Euriso-top, CEA group, Gif-sur-Yvette, France) and intermediate lyophilizations. The oligosaccharides were analysed at 300 K. The chemical shifts (δ) were referenced to internal acetone ($\delta^1\text{H} = 2.225$ and $\delta^{13}\text{C} = 31.55$ p.p.m. under the conditions used). Two-dimensional homonuclear [COSY90, single-relayed COSY and double-relayed COSY, TOCSY, rotating-frame Overhauser enhancement spectroscopy (ROESY)] and heteronuclear multiple quantum coherence (HMQC) experiments were performed by using standard Bruker® pulse programs (cosy, cosyrl, cosyrl, mlevtp, roesyprsh and invbtp respectively). The main pulses and variable delays were optimized for each pulse program and sample.

Composition of monosaccharides

Monosaccharides were systematically analysed by GC/MS as three different derivatives: per-trimethylsilyl [21], per-heptafluorobutyl [22] and alditol-acetate derivatives. For alditol-acetates analysis, oligosaccharide-alditols were hydrolysed in 4 M trifluoroacetic acid for 4 h at 100 °C. They were reduced with sodium borodeuterite in 0.05 M NH_4OH for 4 h. Reduction was stopped by dropwise addition of acetic acid until pH 6 was reached and borate salts were co-distilled by repetitive evaporation in dry methanol. Per-acetylation was performed in acetic anhydride at 100 °C for 2 h.

Carboxyl reduction

Before methylation, GlcA-containing oligosaccharide-alditols were alternatively treated as follows. Compounds were incubated in a mixture of dry DMSO (100 μ l) and iodomethane (100 μ l) for 2 h to convert carboxylic acid groups into methyl esters. After lyophilization, methyl ester function was reduced in a solution of sodium borodeuterite in dry methanol for 12 h. Acetic acid was added to pH 6 and boric acid was removed as methyl borate by repetitive evaporation in methanol. Dried material was further purified by gel filtration on Sephadex G-10 (Pharmacia) and subjected to methylation analysis.

Methylation

Native oligosaccharide-alditols or carboxyl-reduced oligosaccharides were per-methylated according to Ciucanu and Kerek [23]. Briefly, compounds were incubated overnight in a suspension of 200 mg/ml NaOH in dry DMSO (100 μ l) and iodomethane (100 μ l). The methylated products were extracted in chloroform and washed with water. Lyophilized product was subjected to methanolysis in 500 μ l of 0.5 M HCl in anhydrous methanol at 80 °C for 20 h. After drying, products were per-acetylated in 200 μ l of acetic anhydride and 50 μ l of pyridine overnight at

room temperature. The reagents were evaporated, and the sample was dissolved in chloroform before analysis in GC/MS.

RESULTS

Purification

The purification scheme led to the isolation of 12 major fractions by normal-phase HPLC (Figure 1), containing 15 different oligosaccharides. Unlabelled peaks contained either mixtures that were too complex or compounds in quantities too small to be analysed exhaustively. Three fractions (fractions 6, 10 and 12) contained a mixture of two oligosaccharides that were then analysed as a blend. All these compounds were classified in three types, depending on their nature: compounds 2, 3, 4, 5, 6A, 6B and 7 appeared to be mucin-type O-glycans, compounds 1, 8, 12A and 12B were GAG-derived oligosaccharides and compounds 9, 10A, 10B and 11 were oligomannosyl-type N-glycans. Each one was analysed by NMR, and the chemical shifts of their constituent ^1H , and for some of them ^{13}C , are given in Table 1 for mucin-type O-glycans, Table 2 for GAG-type oligosaccharides and Table 3 for N-glycans.

Mucin-type O-glycan analysis

The ^1H NMR spectrum of the oligosaccharide-alditols 2, 3, 4, 5, 6A, 6B and 7 were assigned using two-dimensional COSY, TOCSY and ROESY, and ^1H -detected ^1H - ^{13}C HMQC experiments also allowed the assignment of the ^{13}C chemical shifts for compounds 4, 5, 6A and 6B. After structural analysis, it appeared that oligosaccharide-alditols 2, 3, 4, 6A, 6B and 7 were part of a homologous family of O-glycans, whereas compound 5 presented largely distinct features. Thus, analysis of this last compound will be discussed separately at the end of this section.

Based on the $^3J_{\text{H,H}}$ coupling constant values, the three sugar-spin systems of β -glucopyranose (β -Glc), β -glucopyranuronic acid and β -galactopyranose (β -Gal) were identified in all five samples 2, 3, 4, 6 and 7. The *N*-acetylgalactosaminol (GalNAc-ol) unit was identified owing to the assignment of its characteristic ^1H and ^{13}C atom resonances, and the confirmation was given by GC/MS of the corresponding hexosaminol acetate. The two-dimensional COSY spectrum of compound 4 (Figure 2) clearly showed the presence of 3,6-di-substituted GalNAc-ol, as indicated by the significant downfield displacements of H-3 and H-6,6' at δ 4.073, 3.986 and 3.788 p.p.m. respectively. The absence of downfield glycosylation shifts for the ring protons of β -Glc and β -GlcA indicated that these two units were terminal. On the contrary, downfield displacements were observed for the H-2, H-3 and H-4 atom resonances of β -Gal. The H-5 resonance of β -Gal was assigned from the H-1,5 cross-peak in the ROESY spectrum (Figure 3), and consequently found at δ 3.97 p.p.m. in the ^1H - ^{13}C HMQC spectrum. Comparison of the ^{13}C chemical shifts for β -Gal with literature values for β -methyl galactoside [24] identified the substitution position at C-3. In agreement with the significant deshielding of H-5 and the light shielding of C-5, a second substitution position was established at C-6. Finally, the H-5, H-6, C-5 and C-6 resonances of β -Gal were assigned by elimination after all the other ^1H and ^{13}C resonances had been allocated.

The ROESY spectrum (Figure 3) showed the following contacts: (i) between GlcA III H-1 and signals at δ 3.52 p.p.m. (GlcA III H-3), δ 3.721 p.p.m. (GlcA III H-5) and δ 3.834 p.p.m. (Gal II H-3), (ii) between Glc IV H-1 and signals at δ 3.987 p.p.m. (Gal II H-6') and δ 4.073 p.p.m. (Gal II H-6), (iii) between Gal II H-1 and signals at δ 4.073 p.p.m. (GalNAc-ol H-3), δ 3.834 p.p.m. (Gal II H-3) and δ 3.97 p.p.m. (Gal II H-5) and

Figure 1 HPLC profile of the oligosaccharides released by mild alkali treatment

Material was eluted from a primary amine-bonded silica column by a KH_2PO_4 gradient in a mixture of water and acetonitrile.

(iv) between Glc II' H-1 and signal at δ 3.788 p.p.m. (GalNAc-ol H-6'). On the basis of the data obtained, it is concluded that compound 4 has the following structure:

Compound 2 showed a $[\text{M}-\text{H}]^-$ at m/z 721 in MALDI-TOF. In accordance with composition analysis, it suggested the presence of two hexose residues, one hexosaminitol residue and one hexuronic acid residue. Integration of signals from the two-dimensional COSY spectrum of compound 2 confirmed that it was composed of one β -Glc, one β -Gal, one β -GlcA and one GalNAc-ol. Based on the set of its H-3, H-5 and H-6,6' atom resonances it was concluded that the GalNAc-ol unit is 3-linked. The H-2, H-3, H-4 and H-6,6' atom resonances of β -Gal were identical to those observed in compound 4, as well as the NMR parameters of terminal β -Glc and β -GlcA. Thus the NMR evidence permits us to write the structure of compound 2 as follows:

Compound 3 presented the same $[\text{M}-\text{H}]^-$ at m/z 721 as the latter compound. Based on the relayed COSY spectrum (Figure 2) as well as on the composition analysis, the oligosaccharide-alditol 3 was defined as a tetramer containing two β -Gal, one β -GlcA and one GalNAc-ol unit. The H-2, H-3 and H-4 atom resonances of the β -Gal II unit were observed at δ values that signify a 3-substitution with β -GlcA, whereas the corresponding signals of β -Gal II' are characteristic of a terminal position. Resonance positions of the signals for GalNAc-ol H-3, H-6 and H-6' at δ

4.102, 3.977 and 3.793 p.p.m. respectively showed a 3,6-substitution. The ROESY spectrum showed contacts between (i) GlcA III H-1 and Gal II H-3 (strong), (ii) Gal II H-1 and GalNAc-ol H-3 and (iii) Gal II' H-1 and GalNAc-ol H6, H6'. Therefore, compound 3 has the following structure:

From the MALDI-TOF analysis, fraction 6 was found to be composed of two compounds, 6A and 6B, containing respectively four hexose (Hex), one hexuronic acid (HexA) and one *N*-acetylhexosaminitol (HexNAc-ol; $[\text{M} + \text{Na}]^+$ at m/z 1040) and four Hex and one HexNAc-ol ($[\text{M} + \text{Na}]^+$ at m/z 894). Based on the intensity of the H-1 signals and the $^3J_{\text{H,H}}$ coupling constants, the three sugar-spin systems of β -Glc, β -Gal and β -GlcA were identified in the ratio 6:2:1. Moreover, two types of H-3, H-4 and H-6,6' signal related to the GalNAc-ol unit were clearly observed in the COSY90 spectrum. A ROESY experiment (Figure 4) showed inter-residue cross-peaks between Gal II_A H-1 and GalNAc-ol I_A H-3 and Gal II_B H-1 and GalNAc-ol I_B H-3. Therefore, it is concluded that oligosaccharide-alditols 6A and 6B have the following compositions: three β -Glc, one β -Gal and one β -GlcA and one GalNAc-ol for 6A and three β -Glc, one β -Gal and one GalNAc-ol for 6B. From the NMR spectra, 3,6- and 3,4,6-substituted GalNAc-ol units were characterized. The TOCSY spectrum included signals at δ 4.077, δ 3.556, δ 3.89 and δ 3.78 p.p.m., easily assignable to the H-3, H-4, H-6 and H-6' resonances of the 3,6-disubstituted GalNAc-ol I_A unit, whereas the signals at δ 4.194, δ 3.784, δ 4.05 and δ 3.86 p.p.m. could therefore be attributed to the H-3, H-4, H-6 and H-6' resonances of the 3,4,6-trisubstituted GalNAc-ol I_B unit. The nature of the substitution of the Gal II_A and II_B units was deduced from

Table 1 ^1H and ^{13}C NMR chemical shifts of the mucin-type O-glycans (compounds 2, 3, 4, 5, 6A, 6B and 7)

Compound	Residue	H/C	Chemical shift of ^1H or ^{13}C (p.p.m.)									
			1	1'	2	3	4	5	6	6'	NAc	OCH ₃
2	GalNAc-ol I	H	3.802	3.747	4.399	4.064	3.515	4.201	3.67	3.67	2.052	—
	Gal(β 1-3) II	H	4.558	—	3.732	3.828	4.208	3.97	n.d.	n.d.	—	—
	GlcA(β 1-3) III	H	4.674	—	3.421	3.52	3.52	3.723	—	—	—	—
	Glc(β 1-6) IV	H	4.493	—	3.31	3.505	3.38	3.46	3.918	3.718	—	—
3	GalNAc-ol I	H	≈ 3.8	3.75	4.413	4.102	3.574	4.38	3.977	3.793	2.046	—
	Gal(β 1-3) II	H	4.537	—	3.726	3.828	4.186	n.d.	n.d.	n.d.	—	—
	GlcA(β 1-3) III	H	4.671	—	3.425	3.52	3.52	3.723	—	—	—	—
	Gal(β 1-6) II'	H	4.439	—	3.543	3.659	3.922	n.d.	n.d.	n.d.	—	—
4	GalNAc-ol I	H	≈ 3.8	3.742	4.401	4.073	3.55	4.37	3.986	3.788	2.045	—
		C	60.8	—	51.73	76.97	69.24	68.42	71.26	—	22.23	—
	Gal(β 1-3) II	H	4.564	—	3.734	3.834	4.208	3.97	4.073	3.987	—	—
		C	104.21	—	70.34	82.45	68.17	73.93	70.01	—	—	—
	GlcA(β 1-3) III	H	4.669	—	3.42	3.52	3.52	3.721	—	—	—	—
		C	104.17	—	73.57	75.7	72.12	76.53	n.d.	—	—	—
	Glc(β 1-6) IV	H	4.49	—	3.312	3.505	3.383	3.462	3.92	3.718	—	—
		C	103.08	—	73.55	75.7	≈ 70	76.23	61.08	—	—	—
	Glc(β 1-6) II'	H	4.494	—	3.312	3.505	3.383	3.462	3.92	3.718	—	—
		C	103.08	—	73.55	75.7	≈ 70	76.23	61.08	—	—	—
5	GlcNAc-ol I	H	3.74	3.65	4.191	3.89	3.797	3.83	n.d.	n.d.	2.060	—
		C	62.27	—	53.66	69.82	81.58	74.3	≈ 62.3	—	23.7	—
	GalNAc(β 1-4) II	H	4.501	—	3.964	3.94	4.075	3.87	3.96	3.82	2.081	—
		C	103.61	—	52.63	77.45	70.08	74.79	69.37	—	23.7	—
	Gal(β 1-3) III	H	4.676	—	3.761	3.966	4.282	3.813	4.058	3.93	—	—
		C	103.01	—	75.60	84.36	69.37	74.3	70.31	—	—	—
	Gal(β 1-3) IV	H	4.606	—	3.63	3.63	3.918	3.690	n.d.	n.d.	—	—
		C	105.30	—	72.25	73.9	69.95	76.29	≈ 62.3	—	—	—
	Gal(β 1-6) V	H	4.542	—	3.612	3.820	.920	3.866	4.088	3.87	—	—
		C	103.26	—	77.17	72.8	69.95	74.79	70.87	—	—	—
	Gal(β 1-6) VI	H	4.454	—	3.553	3.665	3.924	3.705	n.d.	n.d.	—	—
		C	104.56	—	72.01	73.9	69.95	76.29	≈ 62.3	—	—	—
	Glc(β 1-6) VII	H	4.538	—	3.310	3.508	3.389	3.471	3.951	3.724	—	—
		C	103.64	—	74.37	77.02	70.76	77.25	61.98	—	—	—
	Fuc(α 1-2) VIII	H	5.325	—	3.815	3.617	3.695	4.185	1.207	—	—	—
		C	100.17	—	69.2	70.45	72.98	67.92	16.9	—	—	—
6A	2-O-Me-Fuc(α 1-2) IX	H	5.443	—	3.502	3.879	3.837	4.284	1.223	—	—	3.511
		C	97.59	—	78.66	70.15	72.84	67.56	16.9	—	—	58.59
	GalNAc-ol IA	H	3.82	3.82	4.415	4.077	3.556	4.38	3.89	3.78	2.047	—
		C	61.61	—	52.49	78	69.95	69.3	71.86	—	23.08	—
	Gal(β 1-3) IIA	H	4.578	—	3.843	3.93	4.415	3.989	≈ 4.1	4.01	—	—
		C	105.03	—	71.28	82.59	76.8	74.44	71.5	—	—	—
	GlcA(β 1-3) IIIA	H	4.673	—	3.419	3.51	3.51	≈ 3.7	—	—	—	—
		C	105.17	—	74.32	76.58	73.11	77.57	n.d.	—	—	—
	Glc(β 1-4) IVA	H	4.832	—	3.306	3.535	≈ 3.4	≈ 3.4	3.93	3.72	—	—
		C	103.56	—	74.3	76.7	70.7	76.57	61.8	—	—	—
	Glc(β 1-6) VA	H	4.476	—	3.28	≈ 3.5	≈ 3.4	3.45	3.93	3.72	—	—
		C	103.95	—	74.3	76.7	70.7	76.9	61.8	—	—	—
6B	Glc(β 1-6) II'A	H	4.501	—	3.33	≈ 3.5	≈ 3.4	3.45	3.93	3.72	—	—
		C	103.73	—	74.3	76.7	70.7	76.9	61.8	—	—	—
	GalNAc-ol IB	H	3.82	3.82	4.45	4.194	3.784	4.38	4.05	3.86	2.067	—
		C	61.61	—	59.93	76.6	78.55	69.3	73.57	—	23.08	—
	Gal(β 1-3) IIB	H	4.552	—	3.603	3.696	3.93	3.973	4.077	3.89	—	—
		C	105.5	—	71.85	73.16	69.55	74.85	70.51	—	—	—
	Glc(β 1-6) VB	H	4.501	—	3.33	≈ 3.5	≈ 3.4	3.45	3.93	3.72	—	—
		C	103.73	—	74.3	76.7	70.7	76.9	61.8	—	—	—
	Glc(β 1-4) VIB	H	4.442	—	3.33	n.d.	n.d.	n.d.	3.93	3.72	—	—
		C	103.44	—	74.3	n.d.	n.d.	n.d.	61.8	—	—	—
	Glc(β 1-6) II'B	H	4.479	—	3.28	≈ 3.5	≈ 3.4	3.45	3.93	3.72	—	—
		C	103.95	—	3.33	76.7	n.d.	76.9	61.8	—	—	—
7	GalNAc-ol I	H	≈ 3.77	≈ 3.77	4.392	4.078	3.509	4.194	≈ 3.64	≈ 3.64	2.046	—
	Gal(β 1-3) II	H	4.529	—	3.700	3.813	4.142	n.d.	n.d.	n.d.	—	—
	GlcA(β 1-3) III	H	4.660	—	3.448	3.604	3.728	n.d.	—	—	—	—
	GlcNAc(β 1-4) IV	H	4.555	—	3.728	3.567	n.d.	n.d.	n.d.	n.d.	2.046	—

the HMQC spectrum (Figure 4), which allowed us to distinguish the 3,4,6-trisubstituted β -Gal II_A unit (C-3, C-4, C-6 observed at δ 82.59, δ 76.8 and δ 71.5 p.p.m. respectively) from the 6-

mono-substituted β -Gal II_B unit (C-3, C-4, C-6 observed at δ 73.16, δ 69.55 and δ 70.51 p.p.m. respectively). Similarly, the HMQC spectrum also confirmed the C-3, C-4 or C-6 substitu-

Figure 2 Extension of ^1H 400 MHz double-relayed COSY spectrum (left) of oligosaccharide 4 ($3.25\text{--}4.75 \times 3.25\text{--}4.75$ p.p.m.) and of ^1H 400 MHz single-relayed COSY spectrum (right) of oligosaccharide 3 ($3.3\text{--}4.8 \times 3.3\text{--}4.8$ p.p.m.)

NMR analysis. On the basis of the data obtained, it is concluded that the compounds 6A and 6B have the following structures:

On the basis of composition analysis and integration of ^1H NMR analysis, compound 7 is composed of one β -GlcNAc, one β -GlcA, one β -Gal and one GalNAc-ol residue. Chemical shifts of GalNAc-ol H-6 and H-6', both observed at δ 3.64 p.p.m., and H-3 at 4.078 p.p.m. are indicative of a mono-substitution in C-3. The Gal II residue shows identical parameters to compound 2 and 3. The H-4 of GlcA III is deshielded to 3.728 p.p.m. compared with 3.51–3.52 p.p.m. in compounds 2, 3, 4 and 6A, suggesting that this residue is substituted by the β -GlcNAc residue in a (1-4) linkage. Thus, we propose the following structure for compound 7:

Compound 5 presented features that were distinct from the six oligosaccharide-alditols analysed above. It showed pseudo-molecular ions $[\text{M} + \text{Na}]^+$ and $[\text{M} + \text{K}]^+$ at m/z 1566 and m/z 1582 on a MALDI-TOF analyser. In accordance with these data and composition analysis it was concluded that this compound constituted one GlcNAc-ol, one GalNAc, one Glc, three Gal, one Fuc and one *O*-Me-Fuc residue. The GalNAc residue was unambiguously distinguished from the GlcNAc-ol residue in GC/MS as an alditol-acetate derivative after a reduction with sodium borodeuteride. Indeed, in electronic-impact mode GlcNAc-ol residue showed two ions at m/z 144 and 145, attributed to $\text{CH}_2\text{OCOCH}_3\text{-CHNH}^+\text{COCH}_3$ and $\text{CH}_2\text{OCOCH}_3\text{-CHO}^+\text{COCH}_3$ fragments, respectively. In contrast, deuterated GalNAc-ol residue showed both fragments $\text{C}^2\text{HHOCOCH}_3\text{-CHNH}^+\text{COCH}_3$ and $\text{CH}_2\text{OCOCH}_3\text{-CHO}^+\text{COCH}_3$ at m/z 145. These data indicated that the non-deuterated GlcNAc-ol residue occurred in the reducing terminal position compared with the internal position of the GalNAc residue. Fuc and *O*-Me-Fuc residues were discriminated as alditol-acetate derivatives in chemical-ionization mode. Fuc was characterized by a $[\text{M} + \text{NH}_4]^+$ at m/z 395 and *O*-Me-Fuc by a $[\text{M} + \text{NH}_4]^+$ at m/z 367. The position of the methyl group in the *O*-Me-Fuc residue was obtained in GC/MS by analysis of its fragmentation pattern in electronic-impact mode. It was characterized by an intense ion at m/z 118 corresponding to the mono-methylated, $\text{C}^2\text{HHOAc-CH}_2\text{O}^+\text{Me}$ fragment, as well as an ion at m/z 275 corresponding to the mono-methylated $\text{CH}_2\text{OAc-(CH}_2\text{OAc)}_3\text{-CH}_2\text{O}^+\text{Me}$ fragment. From these data, we concluded that it was substituted by

Figure 3 ^1H 600 MHz ROESY spectrum of oligosaccharide 4

The ROESY spectrum was performed with a mixing time of 400 ms and with mild presaturation during relaxation time to suppress the residual water signal.

a methyl group in the C-2 position. These results were confirmed by ^1H NMR experiments.

The methylation results indicated that compound 5 contained terminal Gal and Glc residues, a 4-linked GlcNAc-ol, a 2,6-linked Gal and a 2,3,6-linked Gal. Fuc and 2-*O*-Me-Fuc residues were both observed as permethyl-Fuc derivatives.

Fragmentation of the ion $[\text{M} + \text{Na}]^+$ at m/z 1566 in ESI-MS/MS showed a series of informative ions (Figure 6). First, we observed an intense ion at m/z 1342 corresponding to the C ion $[\text{M-HexNAc-ol} + \text{Na}]^+$. This confirmed that the GlcNAc-ol residue in the terminal reducing position was mono-substituted. This is one of the only two fragments containing a Fuc residue that we identified, suggesting an important instability of the fucose-

residue linkage during fragmentation. Ions at m/z 1419 and 1196 correspond to the fragments $[\text{M-Fuc} + \text{Na}]^+$ and $[\text{M-Fuc-HexNAc-ol} + \text{Na}]^+$. Ions at m/z 1034, 872, 712, 550, 388 and 226 all correspond to decomposition fragments of the ion at m/z 1196, lacking both Fuc and HexNAc-ol residues. X ions at m/z 611 and 449 are characteristic of the sequences Hex-HexNAc-HexNAc-ol and HexNAc-HexNAc-ol respectively. These data, concomitantly with methylation analysis, prove the presence of the dimer $\text{GalNAc}(\alpha/1-4)\text{GlcNAc-ol}$ in the terminal reducing position. The C and B ions at m/z 995 and 977 both indicate a fragment containing four Hex, one Fuc and one *O*-Me-Fuc residue. Intense C and B ions at m/z 849 and 831 correspond to the previous fragment lacking its Fuc residue. It is

Figure 4 Extension (50–110 × 65–3.1 p.p.m.) of HMQC (A) and extension (4.8–3.2 × 4.85–4.35 p.p.m.) of ROESY (B) of a mixture of oligosaccharide-alditols 6A and 6B

The ROESY spectrum was performed with a mixing time of 300 ms.

noteworthy that we did not observe the equivalent fragment devoid of *O*-Me-Fuc, suggesting a greater stability of this residue compared with the Fuc residue. Ions at *m/z* 687, 669, 525, 507, 365 and 347 correspond to further fragmentation of ions 849 and 831. Taken together, these data suggest that the GalNAc linked to the GlcNAc-ol residue is disubstituted in C-3 and C-6 by a sequence containing four Hex, one Fuc and one 2-*O*-Me-Fuc residues on one branch and by the remaining Hex residue on the other.

Assignment of the ¹H and ¹³C resonances of the sugar residues and the alditol was made from COSY, two-step relayed COSY, HMQC and ROESY spectroscopy. The H-1 to H-6 resonances for β -Glc_p and the H-1 to H-4 resonances for β -Gal_p and β -GalNAc residues were assigned by referring to the two-step relayed COSY spectrum. The H-5 resonances for β -Gal III, IV, V and VI were detected by ROESY at 3.813, 3.69, 3.866 and 3.705 p.p.m., respectively, suggesting that β -Gal units III and V are 6-linked. The relative low-field position of some ¹³C resonances (Figure 7), compared with their corresponding non-

substituted monosaccharides, indicated the presence of Glc (VII), Gal (IV), Gal (VI), Fuc (VIII), 2-*O*-Me-Fuc (IX) in the terminal position of 2,6-linked Gal (V) and 2,3,6-linked Gal (III). The GalNAc(β 1-4)GlcNAc-ol linkage gave rise to a strong nuclear Overhauser effect (NOE) signal between H-1 of GalNAc and H-4 of GlcNAc-ol (Table 4). Other strong NOE connectivities were observed between H-1 of Gal (V) and Gal (VI) and signals at 4.058 and 4.088 p.p.m., which were assigned as the H-6 resonances of Gal (III) and Gal (V). The exact location of Fuc VIII on Gal II was unambiguously established according to the successive NOE contacts observed between Fuc VIII H-1 and Gal III H-2, and between Gal III H-1 and GalNAc II H-3 (Table 4). Then 2-*O*-Me-Fuc IX was located on the remaining C-2 of the 2,6-disubstituted Gal V, which was confirmed by the NOE contact between 2-*O*-Me-Fuc IX H-1 and Gal V H-2. The resonances which remained to be assigned [namely C-5 of GalNAc (II) and GlcNAc-ol (I) as well as C-6 of GlcNAc-ol (I), Gal (IV) and Gal (VI)] were detected on the HMQC spectrum by comparison with ¹³C NMR chemical shifts of monosaccharides

Figure 5 GC/MS profile of methylation analysis of compounds 6A and 6B

(a) From top to bottom, is shown chromatogram reconstitution using the search for ions specific for 1,5-Me₂-4-Ac-3,6-anhydro-GalNAc(Me)-ol residue (ion at m/z 244) and 3,4,6-trisubstituted GalNAc-ol residue (ion at m/z 216) and a total ion count (TIC) chromatogram. (b) Fragmentation pattern of 1,5-Me₂-4-Ac-3,6-anhydro-GalNAc(Me)-ol. (c) Fragmentation pattern of 1,4,5-Me₃-3,6-anhydro-GalNAc(Me)-ol.

Figure 6 Fragmentation pattern of the proposed structure of oligosaccharide-alditol 5 in ESI-MS

(a) Fragmentation of the pseudo-molecular ion $[M + Na]^+$ at m/z 1565.6. Only the most informative fragments are shown. (b) ESI-MS/MS spectrum.

Figure 7 Extension (110–50 × 5.7–3.1 p.p.m.) of inverse-detected ^1H - ^{13}C 600 MHz HMQC spectrum of oligosaccharide 5

Each experiment is an accumulation of 48 scans, and recorded in the phase-sensitive mode using states-TPPI (time-proportional phase incrementation) method.

Table 4 Summary of inter-residue contacts observed in homonuclear ROESY experiment for the oligosaccharide-alditol, compound 5

Inter-residue NOE connectivity

GalNAc II H-1/GlcNAc-ol H-4
Gal III H-1/GalNAc II H-3
Gal IV H-1/Gal III H-3
Gal V H-1/Gal III H-6
Gal VI H-1/Gal V H-6
Glc VII H-1/GalNAc II H-6
Fuc VIII H-1/Gal III H-2
2-O-Me-Fuc H-1/Gal V H-2

GAG-type oligosaccharide analysis

The second series of compounds (compounds 1, 8, 12A and 12B) is characterized by the presence of the dimer GalNAc(β1-4)GlcA. From the MALDI-TOF analysis and the sugar ratio determination by routine GC/MS analysis, compound 8 was found to be composed of one GalNAc, one GlcA, two Gal and one Xyl-ol residue. The ^1H NMR spectra led us to propose that this compound presented the following sequence:

[24]. Thus, on the basis of the above results, the structure of the compound 5 oligosaccharide may be represented as follows:

Figure 8 Extension (3.3–4.8 × 3.3–4.8 p.p.m.) of ^1H 400 MHz single-relayed COSY spectrum of oligosaccharide-alditol 8

Indeed, the partial sequence $\text{GlcA}(\beta 1-3)\text{Gal}(\beta 1-3)\text{Gal}(\beta 1-4)\text{Xyl-ol}$ was easily identified on the basis of the ^1H NMR parameters, owing to a compilation of previous data [26]. The presence of the additional terminal $\text{GalNAc}(\beta 1-4)$ unit was established on the basis of the 3J vicinal coupling constants of the GalNAc unit and the deshielding of GlcA H-3 and H-4 atom resonances to δ 3.623 and δ 3.75 p.p.m. respectively, as shown on the COSY spectrum (Figure 8).

On the basis of composition and MALDI-TOF analyses,

compounds 1, 12A and 12B contain one GalNAc-ol residue, together with GlcA and GalNAc residues. Compound 1 showed a $[\text{M} + \text{Na}]^+$ at m/z 623 and a $[\text{M} + \text{K}]^+$ at m/z 639. Compounds 12A and 12B were analysed as a mixture and showed $[\text{M} + \text{Na}]^+$ at m/z 799 and 1002, and $[\text{M} + \text{K}]^+$ at m/z 815 and 1018. Thus, these compounds were respectively characterized as tri-, tetra- and pentasaccharide-alditols. In each compound, the H-3 (δ 4.036–4.039 p.p.m.) and H-6,6' (δ 3.618–3.65 p.p.m. and δ 3.588–3.59 p.p.m.) atom resonances of the GalNAc-ol unit indicated a mono 3-O-substitution of this residue (Figure 9). The attachment of a GlcA residue to the GalNAc-ol unit in compounds 12A and 12B was proved by a ROESY experiment (Figure 9). Starting from the H-4 atom resonance of GlcA IV_A and IV_B of compounds

Figure 9 Extension (3.15–4.8 × 3.15–4.8 p.p.m.) of ^1H 400 MHz double-relayed COSY spectrum (left) and extended ROESY (right; same scale as COSY) of mixture 12

The ROESY spectrum was performed with a mixing time of 400 ms and with mild presaturation during relaxation time to suppress the residual water signal.

12A and 12B, it was possible to assign the H-1 resonance of these two units. In a second step, a strong NOE effect was observed between these anomeric protons and both GalNAc III_A and III_B H-3, at δ 3.803 p.p.m. and 3.793 p.p.m. respectively. In addition, the ROESY spectrum led us to distinguish and to assign the H-4 resonances of the GalNAc III_A and GalNAc III_B units, and consequently to determine the precise chemical shifts of their anomeric protons. Owing to a HMQC experiment, the 4-O-substitution of GlcA II_A, II_B and IV_B was proved by the deshielding of their C-4 atom resonances to δ 80.88 p.p.m., compared with terminal GlcA IV_A residue C-4 atom resonance at δ 73.29 p.p.m. Similarly, the C-3 atom resonances of GalNAc III_A and III_B were deshielded to δ 81.79 p.p.m. compared with the C-3 atom resonances of terminal GalNAc V_B at δ 76.4 p.p.m. From these observations, the sequences of compounds 12A and 12B were established as follows:

As shown by the COSY90 spectrum, the terminal non-reducing GalNAc III unit of compound 1 possesses identical H-2, H-3 and H-4 atom resonances to the GalNAc V_B unit of compound 12B. Similarly, the GlcA II protons show identical parameters to GlcA II_A and II_B of compounds 12A and 12B. Thus, we propose the following sequence for compound 1:

N-glycan analysis

MALDI-TOF and composition analyses indicated that fractions 9, 10 and 11 contained Man, GlcNAc and GlcNAc-ol in the proportions 7:1:1, 8:1:1 and 9:1:1. From one-dimensional ^1H NMR analysis, the ^1H NMR parameters matched perfectly those of oligomannoside-type N-glycans [27]. In particular, fraction 9 contained a single Man-7 N-glycan, fraction 10 consisted of a mixture of two Man-8 N-glycans (compounds 10A and 10B) and fraction 11 contained a Man-9 N-glycan. Thus the structures of these compounds were established as follows:

DISCUSSION

In this study, we have fully sequenced seven major highly unusual mucin-type O-glycans. Six of them were part of a homologous family, synthesized in abundance by *C. elegans*. They presented either type-1 core Gal(β1-3)GalNAc or the newly described core type Gal(β1-6)[Gal(β1-3)]GalNAc. Five out of six were substituted with a GlcA unit and showed the trisaccharidic sequence GlcA(β1-3)Gal(β1-3)GalNAc. Most of these structures had (β1-6)Glc and/or (β1-4)Glc residues on the GalNAc and/or Gal residues. To our knowledge, these are the only glucose-containing mucin-type O-glycans ever described. Similarly, compound 6B represents the only example of a trisubstituted GalNAc-ol residue ever described. All these unusual features contribute to the uniqueness of the glycosylation of *C. elegans*. Indeed, this few structures suggest the expression in *C. elegans* of five novel glycosyltransferase activities: four β-glucosyltransferase activities and a N-acetylgalactosamine β-1,6-galactosyltransferase activity.

The last mucin-type O-glycan we described (compound 5) presented a GlcNAc-ol residue in the terminal position, which suggests that it would be linked to the protein moieties through a GlcNAc and not a GalNAc residue. O-glycosidically linked GlcNAc-bound oligosaccharides have already been described in glycoproteins from *Trypanosoma cruzi* [28] and act as sialic acid acceptors for the α-2,3-trans-sialidase [29,30]. They present a simpler sequence than the one described in *C. elegans*, and are characterized by the oligomerization of β-Galf and β-Galp residues on the GlcNAc unit. Compound 5 was also characterized by the simultaneous presence of Fuc(β1-2) and 2-O-Me-Fuc(β1-2) residues, as well as a Glc(β1-6) residue. 2-O-Me-Fuc residues were identified previously in the major mucin-type O-glycans substituting the TES protein of the parasitic nematodes *T. canis* and *T. cati* [5]. However, the overall, complex structure of compound 5 shows no equivalence with any other known O-glycan.

Considering the large amounts of these O-glycans that we have purified (up to 500 μg each/10 g of dry worm), it is likely that

C. elegans overexpresses highly O-glycosylated glycoproteins. Previous studies on the parasitic nematode *T. canis* showed that it overexpresses and secretes a surface-associated highly O-glycosylated mucin-type glycoprotein, TES 120 [1,31]. A gene encoding a mucin-type glycoprotein, *let-653* [32], as well as typical mucin-type cysteine-rich domains [33], have also been characterized in *C. elegans*, suggesting the occurrence of such glycoproteins. Moreover, preliminary results from the fractionation of the total components of *C. elegans* through isopycnic ultracentrifugation (results not shown) gave evidence for the presence of such components. They were characterized by a high density ($\rho \geq 1.4 \text{ g} \cdot \text{ml}^{-1}$), an electrophoretic mobility in agarose gel similar to purified mucins and a susceptibility to Toluidine Blue and periodic acid-Schiff reagents. Previous work showed that mutation of the *let-653* gene in *C. elegans* resulted in larval death due to the dysfunction of the secretory/excretory apparatus [32], confirming the importance of such mucin-type glycoproteins in *C. elegans*.

The fact that we have released four major N-linked glycans among O-glycans by mild alkali treatment is not surprising. Indeed, if it is accepted that the N-glycosidic linkages are more stable under mild alkali conditions than are the O-glycosidic linkages, studies have pointed out that such treatments lead also to a slow degradation of GlcNAc-asparagine linkages [34] and to the release of a low percentage of total intact N-glycans.

We have also described fragments of an unsulphated chondroitin-like GAG by NMR. This component, as well as a minute amount of heparan sulphate-type GAG, have been studied previously by enzymic degradation [8,9]. We did not isolate any heparan sulphate-type GAG, probably due to the fact that the methodology we used only permits access to major components. However, the present study permitted us to isolate the tetrasaccharide GAG core substituted by a single GalNAc residue (compound 8), which was not achieved in previous work. This directly confirmed the presence of the ubiquitous GAG protein moiety linkage sequence GlcA(β1-3)Gal(β1-3)Gal(β1-4)Xyl in *C. elegans*. It is noteworthy that this compound does not seem to be the result of a chemical cleavage releasing the GAG chain from the core sequence. Indeed, to our knowledge, mild alkali degradation under reductive conditions cannot cleave the glycosidic bond in the GlcA(β1-3)GalNAc dimer by β-elimination reaction. Similarly, stability studies of chondroitin sulphate demonstrated that no desulphation of this compound occurred in 0.1 M NaOH at 30 °C over a period of 1000 h [35]. This suggests that the unsulphated pentasaccharide we have isolated is definitely not a degradation product, and is abundantly synthesized by *C. elegans*, either as a precursor for longer GAG chains or as a final product.

The exact origin of compounds 1, 12A and 12B is still unclear. A previous study has demonstrated the possibility of a very slow decomposition process of chondroitin sulphate chains under non-reducing mild alkali conditions (0.1 M NaOH, at 30 and 60 °C) through a β-elimination reaction, forming a 4-deoxy-α-L-threohex-4-enopyranosyluronic acid (ΔHexA) at the non-reducing end of each newly formed chain [35]. To assess this possibility, we subjected a chondroitin sulphate to the reductive alkaline conditions that we used on *C. elegans* extract but failed to detect any degradation product or ΔHexA-containing compound (results not shown). Furthermore, it is noteworthy that we did not detect any ΔHexA-substituted chondroitin-like oligosaccharide in *C. elegans*. Therefore, compounds 1, 12A and 12B were the result of either the selective release of oligosaccharide fragments exclusively from the non-reducing end of an unsulphated chondroitin-like GAG, or the release by β-elimination directly from the peptide backbone of a glycoprotein. This last

hypothesis does not exclude the existence of classical non-sulfated chondroitin chains.

Previous studies isolated mutations that perturbed the worm vulval invagination without affecting vulval cell lineage [36]. Out of the eight genes involved in this defect *sqv-3* and *sqv-8* respectively encode a galactosyltransferase and a glucuronyltransferase directly involved in the biosynthesis of the tetrasaccharide core GlcA(β 1-3)Gal(β 1-3)Gal(β 1-4)Xyl [37–39]. *Sqv-8* could also transfer GlcA to the Gal(β 1-3)GalNAc- α -O-benzyl sequence, although this activity was 10-fold lower than that observed on a Gal(β 1-3)Gal/ β -O-naphthalenemethanol. Considering the quantitative importance of the GlcA(β 1-3)Gal(β 1-3)GalNAc sequence in *C. elegans*, we believe that it is necessary to determine whether *Sqv-8* is also involved in the biosynthesis of these mucin-type O-glycans. If it is involved, then inactivating mutations of *sqv-8* would lead to the synthesis of truncated homologues of the mucin-type O-glycans we have observed in the present study. In this last case, modifications in the synthesis of these compounds might also be involved in the abnormal phenotypes observed in *sqv* mutants.

By defining major oligosaccharide structures expressed by *C. elegans*, our studies reported here should facilitate research concerning the specificity of glycosyltransferases in this organism and the role of glycosylation during development in multicellular organisms. Only knowledge of the *C. elegans* oligosaccharides will permit us to design proper substrates to assess the endogenous specificity of each glycosyltransferase cloned, especially in the case of unusual activities such as the ones observed in *C. elegans*. Similarly, comparison of stage-specific O-glycosylation patterns will be useful. Then, compilation of detailed data obtained from *C. elegans* development and specific-stage glycosyltransferase activities will offer the unique possibility in a multicellular organism to follow the synthesis of glycans throughout its entire embryogenesis.

We are deeply indebted to the late Professor André Verbert for his continuous support and encouragement. We thank Dr J. M. Wieruszkeski, P. Timmerman and G. Ricard for valuable technical assistance. This research was supported by the Centre National de la Recherche scientifique (UMR 8576, Glycobiologie structurale et fonctionnelle) and by the Ministère de l'Enseignement Supérieur et de la Recherche.

REFERENCES

- Maizels, R. M., Kennedy, M. W., Meghji, M., Robertson, B. D. and Smith, H. V. (1987) Shared carbohydrate epitopes on distinct surface and secreted antigens of the parasitic nematode *Toxocara canis*. *J. Immunol.* **139**, 207–214.
- Haslam, S. M., Houston, K. M., Harnett, W., Reason, A. J., Morris, H. R. and Dell, A. (1999) Structural studies of N-glycans of filarial parasites. Conservation of phosphorylcholine-substituted glycans among species and discovery of novel chito-oligomers. *J. Biol. Chem.* **274**, 20953–20960.
- Reason, A. J., Ellis, L. A., Appleton, J. A., Wisniewski, N., Grieve, R. B., McNeil, M., Wassom, D. L., Morris, H. R. and Dell, A. (1994) Novel tyvelose-containing tri- and tetra-antennary N-glycans in the immunodominant antigens of the intracellular parasite *Trichinella spiralis*. *Glycobiology* **4**, 593–603.
- Haslam, S. M., Coles, G. C., Munn, E. A., Smith, S., Smith, H. F., Morris, H. R. and Dell, A. (1996) *Haemonchus contortus* glycoproteins contain N-linked oligosaccharides with novel highly fucosylated core structures. *J. Biol. Chem.* **271**, 30561–30570.
- Khoo, K. H., Maizels, R. M., Page, A. P., Taylor, G. W., Rendell, N. B. and Dell, A. (1991) Characterization of nematode glycoproteins: the major O-glycans of *Toxocara* excretory-secretory antigens are O-methylated trisaccharides. *Glycobiology* **12**, 163–171.
- Gerdts, S., Lochnit, G., Dennis, R. D. and Geyer, R. (1997) Isolation and structural analysis of three neutral glycosphingolipids from a mixed population of *Caenorhabditis elegans* (Nematoda: Rhabditida). *Glycobiology* **7**, 265–275.
- Gerdts, S., Dennis, R. D., Borgonie, G., Schnabel, R. and Geyer, R. (1999) Isolation, characterization and immunolocalization of phosphorylcholine-substituted glycolipids in developmental stages of *Caenorhabditis elegans*. *Eur. J. Biochem.* **266**, 952–963.
- Yamada, S., Van Die, I., Van den Eijnden, D. H., Yokota, A., Kitagawa, H. and Sugahara, K. (1999) Demonstration of glycosaminoglycans in *Caenorhabditis elegans*. *FEBS Lett.* **459**, 327–331.
- Toyoda, H., Kinoshita-Toyoda, A. and Selleck, S. B. (2000) Structural analysis of glycosaminoglycans in *Drosophila* and *Caenorhabditis elegans* and demonstration that tout-velu, a *Drosophila* gene related to EXT tumor suppressors, affects heparan sulfate *in vivo*. *J. Biol. Chem.* **275**, 2269–2275.
- Chen, S., Zhou, S., Sarkar, M., Spence, A. M. and Schachter, H. (1999) Expression of three *Caenorhabditis elegans* N-acetylglucosaminyltransferase I genes during development. *J. Biol. Chem.* **274**, 288–297.
- Chen, S., Spence, A. M. and Schachter, H. (2001) Phenotypes of *Caenorhabditis elegans* UDP-GlcNAc: α -3-D-mannoside β 1,2-N-acetylglucosaminyltransferase I (GnT I) null mutants. *Glycoconj. J.* **17**, 71.
- Hagen, F. K. and Nehrke, K. (1998) cDNA cloning and expression of a family of UDP-N-acetyl-D-galactosamine: polypeptide N-acetylglucosaminyltransferase sequence homologs from *Caenorhabditis elegans*. *J. Biol. Chem.* **273**, 8268–8277.
- Gentile, K., Layden, M., Berbach, K., Schwartz, A. and Hagen, F. (2001) O-linked glycoproteins and renal tubule development in *C. elegans*. *Glycoconj. J.* **17**, 33.
- Oriol, R., Mollicone, R., Cailleau, A., Balanzino, L. and Breton, C. (1999) Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology* **9**, 323–334.
- De Bose-Boyd, R. A., Nyame, A. K. and Cummings, R. D. (1998) Molecular cloning and characterization of an alpha 1,3 fucosyltransferase, CEFT-1, from *Caenorhabditis elegans*. *Glycobiology* **8**, 905–917.
- Brenner, S. (1974) The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94.
- Sulston, J. E. and Hodgkin, J. (1988) Methods. In *The Nematode Caenorhabditis elegans* (Wood, W. B., ed.), pp. 587–606, Cold Spring Harbor Press, Cold Spring Harbor.
- Epstein, H. F., Waterston, R. H. and Brenner, S. (1974) A mutant affecting the heavy chain of myosin in *Caenorhabditis elegans*. *J. Mol. Biol.* **90**, 291–300.
- Sulston, J. E. and Brenner, S. (1988) The DNA of *Caenorhabditis elegans*. *Genetics* **77**, 95–104.
- Lewis, J. A. and Fleming, J. T. (1995) Basic culture methods. In *Caenorhabditis elegans: Modern Biological Analysis of an Organism*, Methods in Cell Biology, vol. 48 (Epstein, H. F. and Shakes, D. C., eds.), pp. 3–29, Academic Press, Orlando.
- Kamerling, J. P., Gerwig, G. J., Vliegthart, J. F. and Clamp, J. R. (1975) Characterization by gas-liquid chromatography-mass spectrometry and proton-magnetic-resonance spectroscopy of pertrimethylsilyl methyl glycosides obtained in the methanolysis of glycoproteins and glycopeptides. *Biochem. J.* **151**, 491–495.
- Zanetta, J. P., Timmerman, P. and Leroy, Y. (1999) Gas-liquid chromatography of the heptafluorobutyrate derivatives of the O-methyl-glycosides on capillary columns: a method for the quantitative determination of the monosaccharide composition of glycoproteins and glycolipids. *Glycobiology* **9**, 255–266.
- Ciucanu, I. and Kerek, F. (1984) A simple and rapid method for the permethylation of carbohydrates. *Carbohydr. Res.* **131**, 209–217.
- Bock, K. and Pedersen, C. (1983) Carbon-13 nuclear magnetic resonance spectroscopy of monosaccharides. *Adv. Carbohydr. Chem. Biochem.* **41**, 27–66.
- Wieruszkeski, J. M., Michalski, J. C., Montreuil, J., Strecker, G., Peter-Katalinic, J., Egge, H., van Halbeek, H., Mutsaers, J. H. and Vliegthart, J. F. (1987) Structure of the monosialyl oligosaccharides derived from salivary gland mucin glycoproteins of the Chinese swiftlet (genus *Collocalia*). Characterization of novel types of extended core structure, Gal beta(1-3)[GlcNAc beta(1-6)] GalNAc alpha(1-3)GalNAc(ol), and of chain termination, [Gal alpha(1-4)]0-1[Gal beta(1-4)]2GlcNAc beta(1-). *J. Biol. Chem.* **262**, 6650–6657.
- Sugahara, K., Ohl, Y., Harada, T., de Waard, P. and Vliegthart, J. F. (1992) Structural studies on sulfated oligosaccharides derived from the carbohydrate-protein linkage region of chondroitin 6-sulfate proteoglycans of shark cartilage. I. Six compounds containing 0 or 1 sulfate and/or phosphate residues. *J. Biol. Chem.* **267**, 6027–6035.
- Vliegthart, J. F. G., Dorland, L. and van Halbeek, H. (1983) High-resolution, 1H-nuclear magnetic resonance spectroscopy as a tool in the structural analysis of carbohydrates related to glycoproteins. *Adv. Carbohydr. Chem. Biochem.* **41**, 209–374.
- Prevato, J. O., Jones, C., Goncalves, L. P., Wait, R., Travassos, L. R. and Mendonça-Prevato, L. (1994) O-glycosidically linked N-acetylglucosamine-bound oligosaccharides from glycoproteins of *Trypanosoma cruzi*. *Biochem. J.* **301**, 151–159.

- 29 Vandekerckhove, F., Schenkman, S., Carvalho, L. P., Tomlinson, S., Kiso, M., Toshida, M., Hasegawa, A. and Nussenzweig, V. (1993) Substrate specificity of the *Trypanosoma cruzi* trans-sialidase. *Glycobiology* **2**, 541–548
- 30 Ferrero-Garcia, M. A., Trombetta, S. E., Sanchez, D. O., Reglero, A., Frasch, A. C. and Parodi, A. J. (1993) The action of *Trypanosoma cruzi* trans-sialidase on glycolipids and glycoproteins. *Eur. J. Biochem.* **213**, 765–771
- 31 Maizels, R. M., de Savigny, D. and Ogilvie, B. M. (1984) Characterisation of surface and excretory-secretory antigens of *Toxocara canis* infective larvae. *Parasite Immunol.* **6**, 23–37
- 32 Jones, S. J. and Baillie, D. L. (1995) Characterization of the *let-653* gene in *Caenorhabditis elegans*. *Mol. Gen. Genet.* **248**, 719–726
- 33 Gems, D. and Maizels, R. M. (1996) An abundantly expressed mucin-like protein from *Toxocara canis* infective larvae: the precursor of the larval surface coat glycoproteins. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1665–1670
- 34 Ogata, S. I. and Lloyd, K. O. (1982) Mild alkaline borohydride treatment of glycoproteins. A method for liberating both N- and O-linked carbohydrate chains. *Anal. Biochem.* **119**, 351–359
- 35 Volpi, N., Mucci, A. and Schenetti, L. (1999) Stability studies of chondroitin sulfate. *Carbohydr. Res.* **315**, 345–349
- 36 Herman, T., Hartwig, E. and Horvitz, H. R. (1999) Sqv mutants of *Caenorhabditis elegans* are defective in vulval epithelial invagination. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 968–973
- 37 Herman, T. and Horvitz, H. R. (1999) Three proteins involved in *Caenorhabditis elegans* vulval invagination are similar to components of a glycosylation pathway. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 974–979
- 38 Okajima, T., Yoshida, K., Kondo, T. and Furukawa, K. (1999) Human homolog of *Caenorhabditis elegans* sqv-3 gene is galactosyltransferase I involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *J. Biol. Chem.* **274**, 22915–22918
- 39 Bulik, D. A., Wei, G., Toyoda, H., Kinoshita-Toyoda, A., Waldrup, W. R., Esko, J. D., Robbins, P. W. and Selleck, S. B. (2000) Sqv-3, -7, and -8, a set of genes affecting morphogenesis in *Caenorhabditis elegans*, encode enzymes required for glycosaminoglycan biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10838–10843

Received 12 February 2001/26 March 2001; accepted 19 April 2001