Glycobiology vol. 11 no. 5 pp. 395–406, 2001

Selective expression of different fucosylated epitopes on two distinct sets of *Schistosoma mansoni* cercarial O-glycans: identification of a novel core type and Lewis X structure

Hung-Hsiang Huang[2,3], Pei-Lun Tsai[2], and Kay-Hooi Khoo[1,2,3]

[2]Institute of Biological Chemistry, Academia Sinica,128, Academia Road Sec 2, Nankang, Taipei 115, Taiwan, R.O.C, and [3]Institute of Biochemical Sciences, National Taiwan University, Taipei, Taiwan

Received on October 20, 2000; revised on December 11, 2000; accepted on January 4, 2001

The glycobiology of *Schistosoma mansoni* is dominated by developmentally regulated expression of various fucosylated structures, most notably the Lewis X epitope and a multifucosylated sequence, Fucα1→2Fucα1→, in its various forms. For the infective cercarial stage, Lewis X has been structurally identified on glycosphingolipids and N-glycans of total glycoprotein extracts, and a population of multifucosylated glycoproteins were found to carry a unique terminal sequence, ±Fucα1→2Fucα1→[3GalNAcβ1→4(Fucα1→2Fucα1→2Fucα1→3)GlcNAcβ1→3Galα1→]$_n$, on their O-glycans. Using a mass spectrometry approach coupled with chromatographic separation, sequential exoglycosidase digestion, periodate oxidation, and other chemical derivatization, we demonstrate that Lewis X could also be carried on the cercarial O-glycans, but the two distinctive sets of fucosylated epitopes were conjugated to two different core structures. Lewis X, lacNAc, or single GlcNAc was found to attach directly to the →3Galβ1→3GalNAc core and indirectly via another β-Gal residue branching off from C6 of the reducing end GalNAc to give a biantennary-like structure. The →3(±Galβ1→6)Galβ1→3(→3Galβ1→6)GalNAc core thus characterized represents a novel core type for O-glycans. In contrast, the previously characterized multifucosylated terminal sequences were carried on conventional type 1 and 2 cores. The smallest structures of the reductively released O-glycans were defined as GalNAcβ1→4GlcNAcβ1→3Galβ1→3GalNAcitol with a total of two to four fucoses attached to the terminal lacdiNAc. α-Galactosylation of the nonreducing terminal β-GalNAc instead of fucose capping leads to further elongation with another lacdiNAc unit that could also extend directly from C6 of the reducing end GalNAc and similarly elongated or terminated.

Key words: Schistosoma mansoni/Lewis X/O-glycosylation/mass spectrometry/structural analysis

Introduction

Schistosomiasis is a parasitic disease inflicted by the digenetic blood flukes of the genus *Schistosoma*. The infectious, free-living stage of schistosome, called cercariae, is covered by a dense 1-μm-thick glycocalyx, which is an extremely potent immunogen with an unusually high fucose content (Samuelson and Caulfield, 1985; Caulfield *et al.*, 1987). Recent studies have demonstrated that at least two distinct sets of fucosylated epitopes are synthesized by the *Schistosoma mansoni* cercariae. A uniquely multifucosylated structure, ±Fucα1→2Fucα1→[3GalNAcβ1→4(Fucα1→2Fucα1→2Fucα1→3)GlcNAcβ1→3Galα1→]$_n$, was found to characterize a very heterogeneous population of the cercarial O-glycans (Khoo *et al.*, 1995). These unusual structures constitute the immunodominant fucosylated epitopes that cross-react with the egg antigens (Weiss and Strand, 1985; Weiss *et al.*, 1986; Dalton *et al.*, 1987), but their expression is apparently down-regulated as the larvae transforms and develops into adult worm (Köster and Strand, 1994). Another source of cross-reacting fucosylated epitopes that are commonly present in the extracts of the cercaria, adults, and eggs is Galβ1→4(Fucα1→3)GlcNAc or Lewis X (Le^x) (Weiss and Strand, 1985; Ko *et al.*, 1990; Köster and Strand, 1994). In the *S. mansoni* cercariae, Le^x epitope has been identified as the predominant terminal epitope on the N-glycans (Khoo *et al.*, 2001) and glycosphingolipids (Wuhrer *et al.*, 2000), although probing with monoclonal antibody revealed that it was probably not exposed on the cercarial surface (Köster and Strand, 1994).

Over the last decade, significant advances have been made in understanding the *N*-glycosylation pattern in parasitic helminths, which include the trematodes and the nematodes (Cummings and Nyame, 1999; Dell *et al.*, 1999). Accumulating structural data indicate that the initial steps in *N*-glycosylation as defined for the mammalian systems are probably well conserved within the helminthic phyla, resulting in the trimannosyl core or truncated version of it in all cases examined. Structural variation in this trimannosyl chitobiose core structure usually resides in the presence or otherwise of α6-, α3-linked Fuc and/or β2-linked Xyl, all of which have been identified on the schistosomal N-glycans (Khoo *et al.*, 1997, 2001). In contrast, much less is known about *O*-glycosylation in parasitic helminths. Unlike the N-glycans, which are en bloc transferred to nascent polypeptides and then further processed by trimming and extension, O-glycans are synthesized by adding one glycosyl residue at a time (Brockhausen, 1995). Part of the difficulties in characterizing the O-glycans in detail therefore stems from the fact that *O*-glycosylation does not follow the conserved pattern of *N*-glycosylation to give the

[1]To whom correspondence should be addressed

well-characterized subsets of high mannose, hybrid, or complex-type structures based on the same trimannosyl core. To date, up to seven well-accepted core types have been identified for the O-glycans (Brockhausen, 1995), with probably a few more unusual ones remaining to be uncovered by refined analytical techniques (Mårtensson *et al.*, 1998). It is anticipated that *O*-glycosylation pathways in the lower organisms, including the parasitic helminths, may be equally diverse and possibly novel.

Despite several detailed structural analysis of unusual glycosyl chains from the schistosomal O-glycans (Bergwerff *et al.*, 1994; van Dam *et al.*, 1994; Khoo *et al.*, 1995), their respective core structures have not been rigorously defined. In most cases, it was suggested that these are conjugated to conventional type 1 and 2 cores, namely, →Galβ1→3GalNAc and →Galβ1→3(→GlcNAcβ1→6)GalNAc, respectively. However, a mass spectrometry (MS) mapping of the total O-glycans from the schistosomal egg extracts (Khoo *et al.*, 1997) has demonstrated that novel core structures may exist beside types 1 and 2. In particular, a $Hex_2HexNAc_1$itol structure was found in both *S. mansoni* and *S. japonicum* egg O-glycans (Khoo *et al.*, 1997), and a branched core structure, Hex-HexNAc-Hex-(Hex)HexNAcitol, has also been identified in the O-glycans from the circulating cathodic antigens excreted by the adult worms (van Dam *et al.*, 1994).

Based on our preliminary observation that the *S. mansoni* cercarial O-glycans comprise a subset of smaller O-glycans not related to those carrying the multifucosylated structures, and that their molecular compositions as defined by MS analysis point to novel core structures, we initiated detailed analysis of this set of O-glycans. Our results indicate that the Le^x epitope is carried on a set of O-glycans with novel branched cores that give rise to biantennary-like structures mimicking the N-glycans. In addition, the recovery of smaller O-glycans belonging to the previously identified multifucosylated series allowed us to further delineate its fucosylation and branching pattern.

Results

Identification of two distinct sets of O-glycans

The *Anguilla anguilla* (AA) lectin–binding glycoprotein fraction from *S. mansoni* cercarial extracts has been shown to carry a series of large, highly fucosylated O-glycans (Khoo *et al.*, 1995) but very little or no N-glycans. The diagnostic markers afforded by fast-atom-bombardment mass spectrometry (FAB-MS) analysis of the permethyl derivatives of these multifucosylated O-glycans are the abundant fragment ions at *m/z* 434 and 608, corresponding to nonreducing terminal mono- and difucosylated GalNAc, respectively (Khoo *et al.*, 1995).

In a concerted effort to examine the total protein glycosylation profile of the cercarial stage, we have since found that the total glycoprotein extracts not further purified by the AA lectin affinity column do carry an abundance of N-glycans (Khoo *et al.*, 2001). Interestingly, the de-*N*-glycosylated peptide sample was found to yield a series of novel O-glycans quite distinct from those characterized previously, the permethyl derivatives of which afforded very little of the characteristic fragment ions at *m/z* 434 and 608. On the other hand, if the total extracts were first passed through the AA lectin column, the multifucosylated components could be detected in the bound fractions eluted with 100 mM fucose, whereas the flow-through fraction yielded O-glycans similar to those observed when the total O-glycans were examined. The majority of N-glycans were also found to be contained within this flow-through fraction. To facilitate MS identification of both sets of O-glycans, a prior AA-lectin fractionation is therefore required. The two subsets recovered are denoted as glycoprotein fractions retained and not retained by the AA lectin column, respectively.

Total O-glycan profile from the glycoprotein fraction not retained by the lectin column

The *S. mansoni* cercarial glycoprotein extracts that failed to bind the AA lectin consistently yielded a series of O-glycans of simpler composition, ranging from 2 to 10 glycosyl residues, as defined by FAB-MS analysis of their permethyl derivatives (Figure 1, see Table I for assignment). Because the O-glycans were released by reductive elimination, they were recovered as oligoglycosyl alditol, denoted as "-itol." Sugar analysis indicated that the only alditol found is GalNAcitol, whereas linkage analysis showed that both 3-linked GalNAcitol and 3,6-linked GalNAcitol could be detected. Thus, the *O*-linked glycans are attached to the peptide via a GalNAc residue at the reducing end and that they could either form a linear R→3GalNAc chain or a branched R_1→3(R_2→6)GalNAc structure.

However, the identification of a major $Hex_2HexNAc_1$-itol structure (*m/z* 716) and the collision-induced dissociation (CID) MS/MS data on selected molecular ions afforded by the permethyl derivatives (Figure 2) gave a first indication that

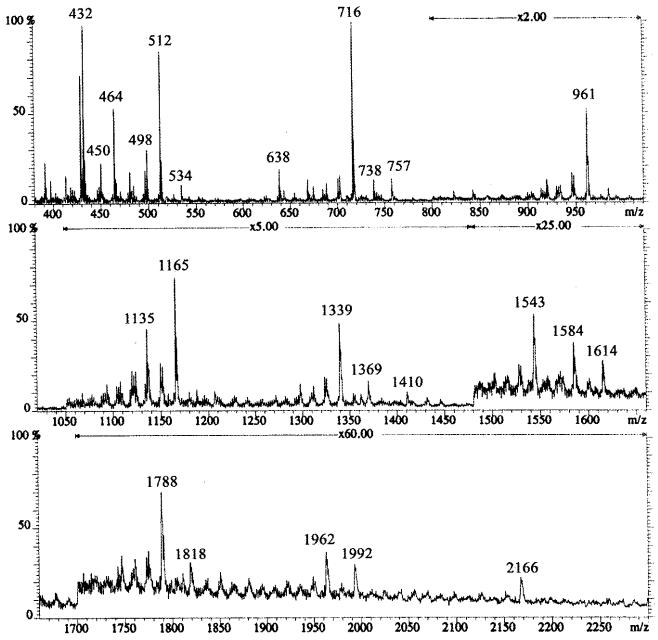

Fig. 1. FAB-mass spectrum of the permethyl derivatives of *S. mansoni* cercarial O-glycans from glycoprotein fraction not retained by the lectin column. Assignment of the $[M+H]^+$ molecular ion signals are listed in Table I. Signals at *m/z* 534 and 738 are the sodiated molecular ions. Signals at *m/z* 432, 464, 498, and 638 are fragment ions, the assignment of which are shown on the schematic drawings in Figure 2. Fragment ion signal at *m/z* 260 is also present at the low mass end.

Table I. *S. mansoni* cercarial O-glycans from glycoprotein fraction not retained by the lectin column

No. (GU)[a]	[b][M+H]$^+$	Composition	Key structural data[c]	Deduced structures[d]
#1 (3.2u)	512	$Hex_1HexNAc_1$-itol	LA : t-Gal, 3-GalNAcitol IO_4 : Hex-C_4	Gal-3GalNAcitol
#2 (4.5u)	716	$Hex_2HexNAc_1$-itol	LA : t-Gal, 3,6-GalNAcitol IO_4 : Hex-C_2, Hex-C_4	Gal-6 Gal-3GalNAcitol
#3 (4.1u)	757	$Hex_1HexNAc_2$-itol	IO_4 : $Hex_1HexNAc_1$-C_4	GlcNAc-3Gal-3GalNAcitol
#4 (5.3u)	961	$Hex_2HexNAc_2$-itol	MS/MS : Fig 2A IO_4 : $Hex_2HexNAc_1$-C_4, Hex-C_2, $Hex_1HexNAc_1$-C_4 $Hex_1HexNAc_1$-C_2, Hex-C_4	Gal-4GlcNAc-3Gal-3GalNAcitol Gal-6 GlcNAc-3Gal-3GalNAcitol GlcNAc-3Gal-6 Gal-3GalNAcitol
#5 (6.1u)	1135	$Fuc_1Hex_2HexNAc_2$-itol	MS/MS : Fig 2C LA : t-Fuc, t-Gal, 3-Gal, 3,4-GlcNAc, 3-GalNAcitol	Gal-4GlcNAc-3Gal-3GalNAcitol Fuc3
#6 (6.5u)	1165	$Hex_3HexNAc_2$-itol	MS/MS : Fig 2B IO_4 : $Hex_2HexNAc_1$-C_2, Hex-C_4 $Hex_2HexNAc_1$-C_4, Hex-C_2 $Hex_3HexNAc_1$-C_4	Gal-4GlcNAc-3Gal-6 Gal-3GalNAcitol Gal-6 Gal-4GlcNAc-3Gal-3GalNAcitol Gal-6 Gal-4GlcNAc-3Gal-3GalNAcitol
#7a (7.4u) #7b #7c	1339 1369 1410	$Fuc_1Hex_3HexNAc_2$-itol (major within #7) $Hex_4HexNAc_2$-itol $Hex_3HexNAc_3$-itol	MS/MS : Fig 2D LA : t-Fuc, t-Gal, 3-Gal, 3,6-Gal, 3,4-GlcNAc, 3-GalNAcitol, 3,6-GalNAcitol	Gal-6 Gal-4GlcNAc-3Gal-3GalNAcitol Fuc3 Gal-6 Gal-4GlcNAc-3Gal-3GalNAcitol Fuc3
#8a (8.6u) #8b #8c	1543 1584 1614	$Fuc_1Hex_4HexNAc_2$-itol (major within #8) $Fuc_1Hex_3HexNAc_3$-itol $Hex_4HexNAc_3$-itol	MS/MS : Fig 2E LA : t-Fuc, t-Gal, 3,6-Gal, 3,4-GlcNAc, 3,6-GalNAcitol	Gal-6 Gal-6 Gal-4GlcNAc-3Gal-3GalNAcitol Fuc3
#9a (9.5u) #9b	1788 1818	$Fuc_1Hex_4HexNAc_3$-itol (major within #9) $Hex_5HexNAc_3$-itol	MS/MS : *m/z* 432, 450, 464, 638 IO_4 : $Hex_2HexNAc_1$-C_2, $Hex_2HexNAc_1$-C_4, $Fuc_1Hex_2HexNAc_1$-C_2, $Fuc_1Hex_2HexNAc_1$-C_4, $Hex_3HexNAc_1$-C_4	Gal-4GlcNAc-3Gal-6 3GalNAcitol Gal-4GlcNAc-3Gal- Fuc3
#10a (10.5u) #10b	1962 1992	$Fuc_2Hex_4HexNAc_3$-itol (major within #10) $Fuc_1Hex_5HexNAc_3$-itol	MS/MS : *m/z* 432, 450, 638 IO_4 : $Fuc_1Hex_2HexNAc_1$-C_2, $Fuc_1Hex_2HexNAc_1$-C_4 $Hex_3HexNAc_1$-C_4 (minor)	Fuc3 Gal-4GlcNAc-3Gal-6 Gal-4GlcNAc-3Gal-3GalNAcitol Fuc3
#11 (11.4u)	2166	$Fuc_2Hex_5HexNAc_3$-itol	IO_4 : $Fuc_1Hex_2HexNAc_1$-C_2 $Fuc_1Hex_3HexNAc_1$-C_4	Fuc3 Gal-4GlcNAc-3Gal-6 Gal-6 Gal-4GlcNAc-3Gal-3GalNAcitol Fuc3

[a] Pooled HPLC fractions were numbered according to their elution order. The corresponding glucose unit (GU) is given in parenthesis. #7–10 were found to contain more than one components differing in molecular weights.
[b] *m/z* of the protonated molecular ions for the permethyl derivatives (Figure 1).
[c] Linkage analysis (LA) and mild periodate oxidation (IO_4) were performed on selected fractions. In the case of #7–10, the data obtained correspond mainly to the major component within each fraction.
[d] Deduced structures for the major component within each fraction.

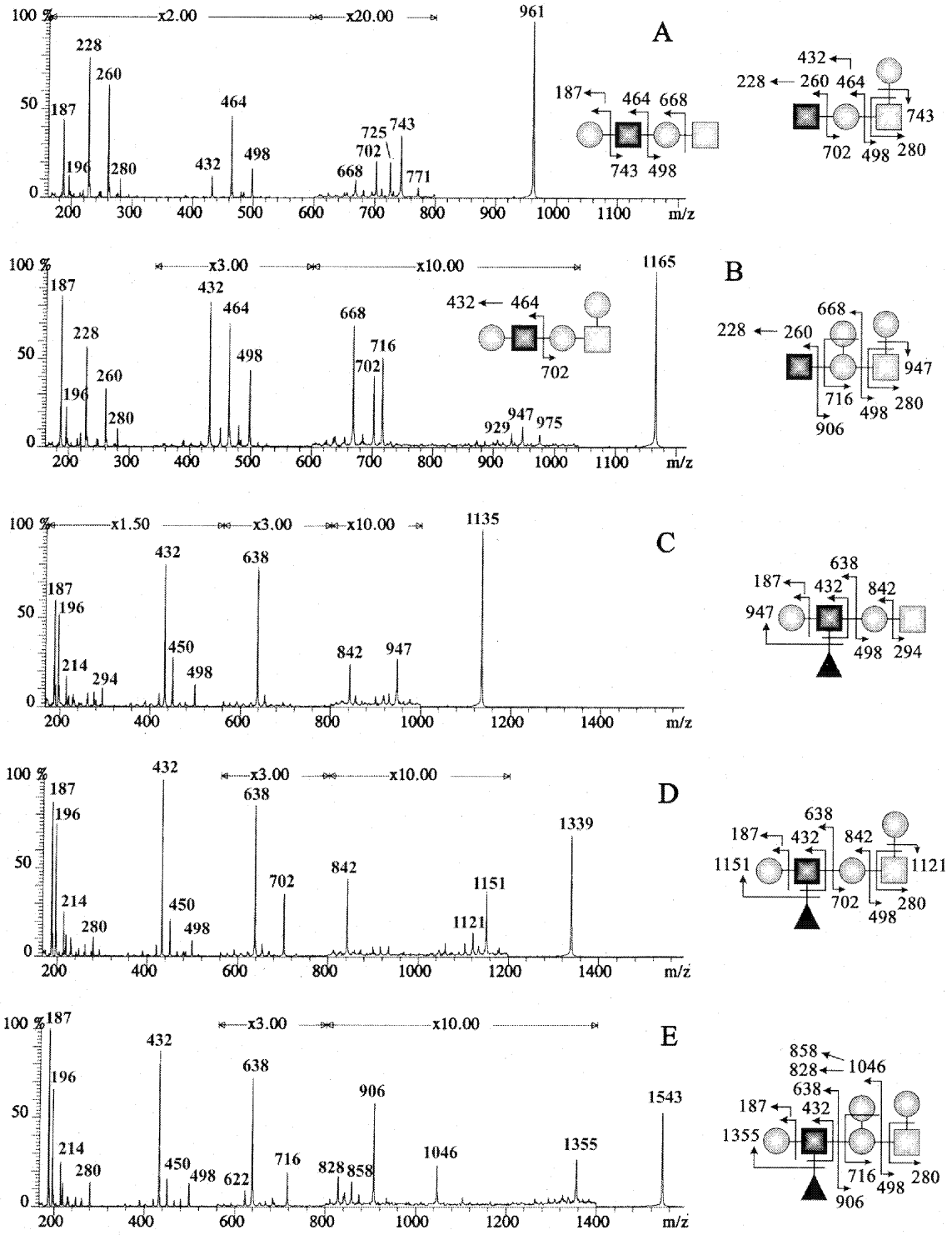

Fig. 2. CID MS/MS spectra of the [M+H]+ parent ions at m/z 961 (**A**), 1165 (**B**), 1135 (**C**), 1339 (**D**), and 1543 (**E**) afforded by the permethyl derivatives of *S. mansoni* cercarial O-glycans. Daughter ions are mainly of two types: (1) nonreducing terminal oxonium ions, for example, *m/z* 260, 464, 638, 842, 1046, most were accompanied by further elimination of the substituent at position 3 of the residue at cleavage site; (2) reducing terminal β-cleavage ions, for example, *m/z* 280, 498, 702, 906. The assignment of the daughter ions are schematically shown on the right. More than one structural isomers were implicated for (**A**) and (**B**) to account for all ions observed. Symbols used: circle, Hex; dark square, HexNAc; triangle, Fuc; light square, GalNAcitol.

these deceptively simple O-glycans may not be based on conventional type 1 and 2 core structures. Instead, the daughter ion profiles afforded by MS/MS analysis implicated a novel branched core. The absence, in most cases, of β-cleavage ion at

m/z 294 that represents $(HO)_1$GalNAcitol appeared to be compensated by the ions at *m/z* 280 and 498, corresponding to $(HO)_2$GalNAcitol and $(OH)_1Hex_1$GalNAcitol, respectively. Likewise, the daughter ions at *m/z* 716 and 906 for the *m/z* 1543 parent (Figure 2E) in place of the β-cleavage ion at *m/z* 702 for the *m/z* 1339 parent (Figure 2D) were best rationalized by evoking a second branch point on the Hex directly linked to the branched GalNAcitol. Thus, a probable core structure may be represented as R-(±Hex)Hex-3(±Hex-6)GalNAc-itol, where R defines the nonreducing terminal sequences.

Nonreducing terminal sequences

Nonreducing end fragment ions corresponding to terminal $HexNAc^+$, Hex-$HexNAc^+$, and Hex-(Fuc-3)$HexNAc^+$ were afforded by direct FAB-MS analysis (Figure 1) at *m/z* 260, 464, and 638, respectively, resulting from facile A-type oxonium cleavage at permethylated HexNAc residues (Dell, 1987). The absence of an ion at *m/z* 606 but the presence instead of an abundant ion at *m/z* 432 is consistent with the elimination of a Fuc residue rather than an *O*-methyl moiety from position 3 of the HexNAc residue at the cleavage site. These fragment ions were also formed with CID MS/MS analysis (Figure 2) although under CID, the cleavages observed were not restricted to HexNAc. The additional daughter ion at *m/z* 842 coupled with the β-cleavage ions at *m/z* 702 and 498 for the *m/z* 1339 parent (Figure 2D) indicated that the Hex-(Fuc)HexNAc terminus was directly linked to a -Hex-(Hex)HexNAcitol core. Likewise, the ion at *m/z* 1046 instead of *m/z* 842, coupled with the corresponding β-cleavage ions at *m/z* 906 and 498 for the *m/z* 1543 parent (Figure 2E), supported a Hex-(Fuc)HexNAc-(Hex)Hex-(Hex)HexNAcitol structure. In addition, treatment with α3/4-specific fucosidase abolished all components assigned as carrying one or two Fuc residues. Thus, in comparison with Figure 1, the permethyl derivatives of the fucosidase-digested samples afforded only sodiated molecular ions at *m/z* 534, 738, 779, 983, 1187, 1391, 1636, and 1840 (data not shown), which were 22 mass units higher than the protonated molecular ions of the nonfucosylated components (Figure 1).

To facilitate a more definitive characterization, the O-glycans were fractionated by normal-phase high-pressure liquid chromatography (HPLC) on a Palpak type N column, calibrated with pyridylaminated Glc oligomer standards. One-milliliter fractions were collected and screened by MS to identify the eluted components. Successive fractions containing the same oligoglycosyl alditols were pooled and numbered according to their elution order and corresponding Glc units (Table I). All components detected by direct MS analysis (Figure 1) were identified although several of them coeluted within the same fractions collected.

The fucose-containing oligoglycosyl alditols #5, #7, and #8 were subjected to sequential enzyme digestions and linkage analysis. Sequential removal of one α-Fuc, β-Gal, and β-GlcNAc by specific α3,4-fucosidase, β4-galactosidase, and *N*-acetyl-β-glucosaminidase, respectively, as detected by shifts in mass by MS analysis, as well as the detection of t-Fuc, t-Gal, and 3,4-GlcNAc in linkage analysis, together with the corresponding MS/MS data, firmly established the presence of Le^x. Based on the daughter ions at *m/z* 432 and 638 but not 606, it was deduced that the larger components, #9a and #10a, similarly carried Le^x as the terminal epitope. The yield of an additional ion at *m/z* 464 (Hex-$HexNAc^+$) from the parent ion at *m/z* 1788 (#9a) but not that at *m/z* 1962 (#10a) suggested that a nonfucosylated lacNAc terminus existed along with the Le^x terminus in #9a. Other nonreducing terminal structures attached to the smaller O-glycans could be considered as lacking either the Fuc or both Fuc and Gal to form lacNAc or single terminal GlcNAc, respectively.

The detection of a 3-linked Gal in #5 and #7, and 3,6-linked Gal in #7 and #8 indicated that the Le^x epitope was attached through a 3-linked Gal to the GalNAcitol core, and that this 3-linked Gal may be further substituted with another Gal at C6 position to give the 3,6-linked Gal identified in linkage analysis. This additional Gal residue, as well as the one implicated to be present at the C6 position of the GalNAcitol, were found to be resistant to β-galactosidase from both bovine testes or *Diplococcus pneumoniae*. Thus, the aforementioned sequential enzyme digestion of #5, #7, and #8 gave a major product corresponding to Hex_1HexNAcitol, Hex_2HexNAcitol, and Hex_3HexNAcitol, respectively, after the removal of the Le^x moiety. Further linkage analysis on the digested product of #8 gave terminal Gal, 6-linked Gal, and 3,6-linked GalNAcitol, consistent with the conversion of a 3,6-linked Gal to 6-linked Gal due to removal of the terminal Le^x structure from position 3 of the branched Gal residue. Eventually, it was found that the resulting Hex_2HexNAcitol and Hex_3HexNAcitol products were susceptible to β-galactosidase from jack bean, which is known to be most reactive against β1→6 bond. Both Hex_2HexNActiol and Hex_3HexNActiol were converted to Hex_1HexNActiol. Using the same enzyme under the same conditions, both the Hex_1HexNAcitol digestion product from #5 and the nondigested Hex_1HexNAcitol from #1 remained undigested. It was therefore concluded that all additional terminal Gal on the Gal-GalNAcitol core were of β-anomeric configuration.

Periodate oxidation defined the branched core

The size and molecular composition of the larger O-glycans (#9–11) implicated the presence of two Le^x units or a Le^x and a lacNAc units. The absence of larger fragment ions corresponding to a tandem of two such units indicated that they were not arranged as polylactosamine-like structures. Instead, the additional lacNAc or Le^x unit was likely to be attached to one of the two identified branched points on the -(Gal)Gal-(Gal)GalNacitol core. To resolve this issue, a strategy based on periodate oxidation was employed.

Periodate oxidation has been widely used to oxidize two adjacent hydroxyl groups to aldehydes. In the case of the O-glycans with HexNAcitol at the reducing end, the glycosyl alditol will be cleaved into two series, one that is based on the remnant C5–C6 from GalNAcitol denoted as $R\text{-}C_2$, and the other, which is based on the remnant C1–C4 denoted as $R\text{-}C_4$. This provided a very effective way to determine which residues were attached to the 6 and 3 positions of GalNAc, respectively. It has been shown that under mild conditions with reduced concentration of periodate and shorter reaction time, this oxidation can be restricted to the adjacent hydroxyl groups of linear alditol (Stoll *et al.*, 1990). Thus, in the case of the oligoglycosyl alditols, the only bond that would be oxidized and cleaved is between C4 and C5 of the GalNAcitol, as mentioned. On the other hand, under normal periodate oxidation conditions followed by mild acid treatment (Smith degradation), all residues with adjacent OH groups, including all terminal Fuc,

terminal Gal, terminal GlcNAc, will be cleaved and removed, whereas 4-linked GlcNAc and 3-linked Gal will remain intact.

When the nonfractionated total O-glycans were subjected to Smith degradation, the products obtained were identified by MS analysis as Hex-C_2, HexNAc-Hex-C_2, Hex-C_4, and HexNAc-Hex-C_4 (data not shown). The data confirmed that the internal Gal of the backbone chain was indeed 3-linked as deduced by linkage analysis. Though the production of the C_4 fragments were expected, the recovery of the C_2-containing pieces indicated that the Gal residue 6-linked to the GalNAcitol could be further extended. A more complete range of products were obtained via the mild oxidation conditions, which did not remove the terminal residues.

As shown in Figure 3, the smallest oxidized products found were Hex-C_2 ($[M+Na]^+$ at *m/z* 472) and Hex-C_4 (*m/z* 618). It showed that a single Hex could be directly attached to GalNAc-itol at the C3 and C6 positions as would be expected for #1 and #2 which were also individually oxidized and characterized (Table I). Going up in size, the $[M+Na]^+$ molecular ions of the perdeuteroacetylated C_2-containing products could be assigned as HexNAc-Hex-C_2 (*m/z* 765), Hex-HexNAc-Hex-C_2 (*m/z* 1062), and Hex-(Fuc)HexNAc-Hex-C_2 (*m/z* 1298). MS/MS sequencing of the last product using the protonated parent ion at *m/z* 1276 yielded daughter ions that supported the sequence (Figure 4A). Further linkage analysis demonstrated that the internal Gal was also 3-linked. The data therefore confirmed that a Le^x structure was attached to the C3 position of the Gal on the 6-arm of the GalNAcitol.

For the C_4-containing series, the other molecular ions detected were HexNAc-Hex-C_4 (*m/z* 911), Hex-HexNAc-Hex-C_4 (*m/z* 1208), Hex-(Fuc)HexNAc-Hex-C_4 (*m/z* 1444), Hex-HexNAc-(Hex)Hex-C_4 (*m/z* 1505), and Hex-(Fuc)HexNAc-(Hex)Hex-C_4 (*m/z* 1741). The deduced sequences were likewise supported by MS/MS sequencing, as shown for the two

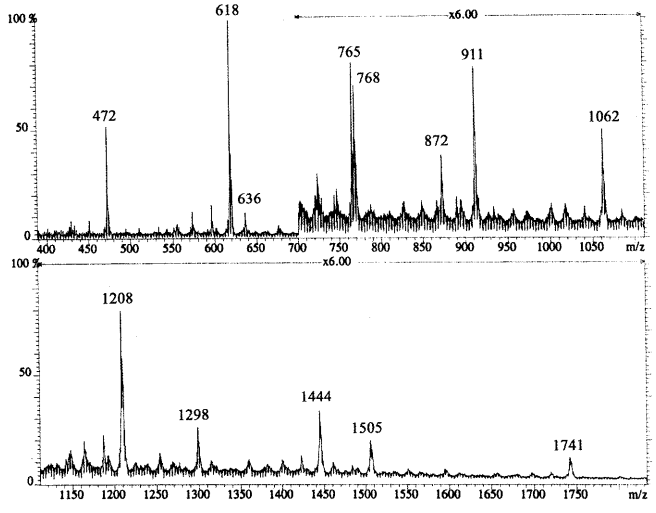

Fig. 3. FAB-mass spectrum of the perdeuteroacetyl derivatives of mild periodate oxidized *S. mansoni* cercarial O-glycans from glycoprotein fraction not retained by the lectin column. Signals at *m/z* 636 and 872 correspond to oxonium type fragment ions Hex-HexNAc$^+$ and $Fuc_1Hex_1HexNAc^+$, respectively. The signal at *m/z* 768 corresponds to $[M+Na]^+$ of nonoxidized Hex-HexNAcitol.

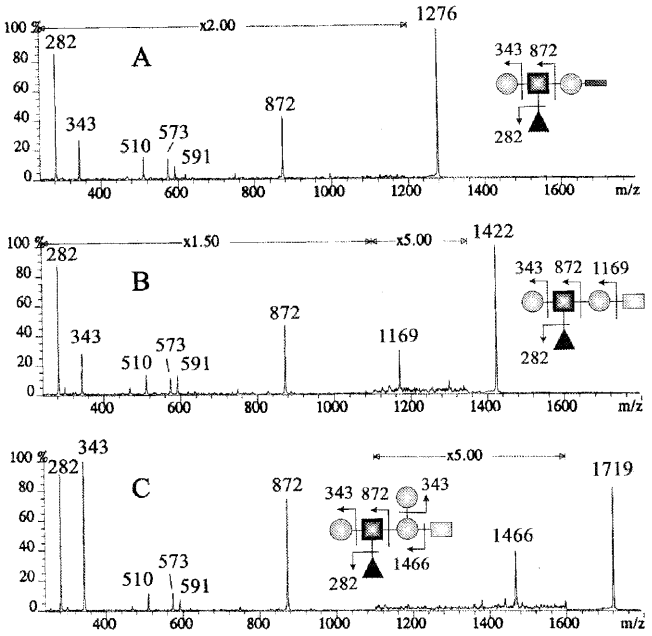

Fig. 4. CID MS/MS spectra of the $[M+H]^+$ parent ions at *m/z* 1276 (**A**), 1422 (**B**), and 1719 (**C**) afforded by the perdeuteroacetyl derivatives of mild-periodate-oxidized *S. mansoni* cercarial O-glycans. The predominant daughter ions produced correspond to nonreducing terminal oxonium type fragment ions as shown on the schematic drawings. Elimination of the Fuc from C3 position of the HexNAc of Hex-(Fuc)HexNAc$^+$ (*m/z* 872) yielded the ion at *m/z* 573. A further loss of a deuteroacetic acid moiety yielded the ion at *m/z* 510, whereas *m/z* 591 corresponds to Hex-(OH)HexNAc$^+$. Notably in all three cases, fully perdeuteroacetylated Hex-HexNAc$^+$ (*m/z* 636) was not detected. Symbols used: circle, Hex; dark square, HexNAc; triangle, Fuc; light square, C_4; dark bar, C_2.

Fuc-containing components (Figure 4). The critical difference between the daughter ion spectra afforded by the *m/z* 1422 (Figure 4B) and *m/z* 1719 (Figure 4C) parents was the absence of *m/z* 1169 in the latter, substituted by an ion at *m/z* 1466. This corroborated the deduced branched sequence, which was also reflected by the higher abundance of *m/z* 343 in Figure 4C because two terminal Hex were present instead of only one.

Together with other data described above, the mild periodate oxidation data therefore unambiguously defined the presence of GlcNAc, lacNAc, and Le^x structures extending from both 3- and 6-arms of the GalNAcitol via an internal 3-linked Gal. The oligoglycosyl chain on the 3-arm could have an extra Gal on the C6-position of the internal 3-linked Gal, whereas this additional terminal Gal stub was not found on the 6-arm. Application of the same mild periodate oxidation to the individual HPLC fractions containing the various O-glycans (Table I) resulted in an additional confirmation of the structures deduced. It also demonstrated that the O-glycans #4 and #6 were a mixture of all possible structural isomers. For the larger O-glycans (#9–11), it established the presence of biantennary-like structures instead of polylactosamine-like structures. The deduced structures for each of the detected components are as shown in Table I. All GlcNAc and Gal were shown to be β except the Gal 3-linked to the GalNAcitol, the anomeric configuration of which could not be established by exoglycosidase digestion.

Further confirmation of the anomeric configurations of the core

As mentioned, the Gal on the 6-position of GalNAcitol could be removed by β-galactosidase from jack bean, whereas the Gal on the 3-position of GalNAcitol was not susceptible to any of the β-galactosidase tested or the coffee bean α-galactosidase. Thus, one Gal residue was removed from Gal-3(Gal-6)GalNAcitol (#2), whereas Gal-3GalNAcitol was resistant to digestion. The β-galactosidase digestion product of #2 yielded Hex-C_4 and not Hex-C_2 after mild periodate oxidation, as detected by MS, which further confirmed that it was the Gal β6-linked to the GalNAcitol in #2 which was removed.

To prove that the Gal 3-linked to GalNAcitol is also in β configuration, a different strategy was used. Gal-3GalNAcitol (#1) and Gal-3(Gal-6)GalNAcitol (#2) were first mild periodate oxidized to become Gal-C_4 and/or Gal-C_2 before perdeutroacetylation, and then subjected to chromium trioxide (CrO_3) oxidation. For controls, authentic Galβ1-3GalNAcitol and Galα1-3Galactitol standard were also processed in exactly the same way. As expected, the Galβ1-3GalNAcitol standard was oxidized (by periodate) first to Galβ1-C_4 ($[M+H]^+$ at *m/z* 596), and then oxidized by CrO_3 to *m/z* 610, 14 u higher than *m/z* 596. In contrast, Galα1-3Galactitol was oxidized by periodate to Galα1-C_4 ($[M+Na]^+$, *m/z* 622), and Galα1-C_5 ($[M+Na]^+$, *m/z* 697), but remained resistant to CrO_3 oxidation. As shown in Figure 5A, Gal-3GalNAcitol (#1) and its mild-periodate-oxidized product, Gal-C_4, behaved exactly like the authentic Galβ1-3GalNAcitol standard, that is, oxidized by CrO_3 to 14-u increment. Likewise, the Gal-C_2 and Gal-C_4 products from #2 were oxidized by CrO_3 (Figure 5B). Further MS/MS analysis on the parent ions of the oxidized products (*m/z* 464 and 610) showed that instead of a dominant *m/z* 343 (Hex^+) daughter ion, a signal at *m/z* 357 was produced. These data therefore conclusively demonstrate a β anomeric configuration for both Gal that were oxidized by CrO_3 to ketoester.

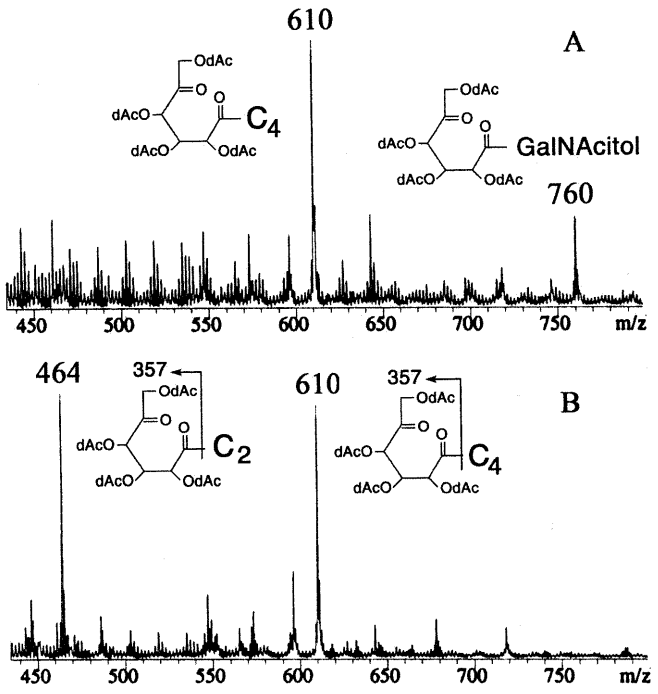

Fig. 5. FAB-mass spectra of the perdeuteroacetyl derivatives of mild periodate oxidized Gal-GalNAcitol (**A**) and Gal-(Gal)GalNAcitol (**B**) after further CrO_3 oxidation. In (**A**), the original Gal-GalNAcitol was not completely cleaved by periodate oxidation to give Gal-C_4. MS/MS analysis of the two products in (**B**) gave mainly an *m/z* 357 daughter ion from both, as shown schematically.

Total O-glycan profile from the glycoprotein fraction retained by the lectin column

In comparison with the FAB-mass spectrum of the permethyl derivatives of the O-glycans from glycoprotein fraction not retained by the lectin column (Figure 1), those from the bound fraction afforded additional signals (Figure 6A). In particular, signals at *m/z* 434, 608, 1027, 1201, 1375, 1650, and 1824 corresponded to A-type oxonium fragment ions, which defined the previously characterized ±Fuc-2Fuc-3GalNAc-(Fuc-2Fuc-2Fuc-3)4GlcNAc-3Gal-3GalNAc- termini for the cercarial glycocalyx (Khoo *et al.*, 1995). Additional molecular ion

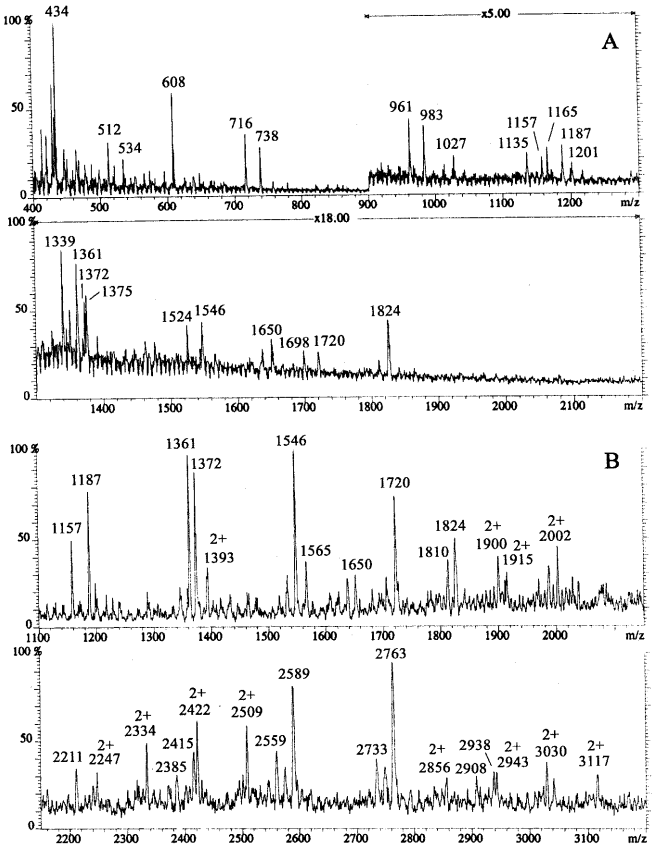

Fig. 6. FAB-mass spectrum (**A**) and ESI-mass spectrum (**B**) of the permethyl derivatives of *S. mansoni* cercarial O-glycans from glycoprotein fractions retained by AA lectin column chromatography. Doubly charged molecular ions observed in (**B**) are marked with the annotation "2+." These were differentiated from singly charged ions by virture of the mass difference between their isotope peaks. The molecular ion signals at *m/z* 512/534 ($[M+H]^+$/$[M+Na]^+$), 716/738, 961/938, 1135/1157, 1165/1187, and 1339/1361 in (**A**) and *m/z* 1157, 1187, 1361, 1565, 1810 in (**B**) corresponded to the same series of O-glycans characterized for the glycorprotein fraction not retained by the lectin column (Figure 1).

signals at *m/z* 1350/1372, 1524/1546, and 1698/1720 could be attributed to the smallest O-glycans, which carried the unique multifucosylated termini, without any of the repeating unit. Thus, *m/z* 1698 could be assigned as $[M+H]^+$ of $HexNAc_2Gal$-3GalNAcitol with four fucoses attached. MS/MS analysis afforded two major daughter ions at *m/z* 608 and 1201, corresponding to $Fuc_2HexNAc^+$ and $Fuc_2HexNAc\text{-}Fuc_2HexNAc^+$, respectively. The absence of *m/z* 434 ($Fuc_1HexNAc^+$) indicated that both HexNAc carried two Fuc. One of the terminal Fuc was missing in the *m/z* 1524 component, which gave fragment ions at *m/z* 434 ($Fuc_1HexNAc^+$), 608 ($Fuc_2HexNAc^+$), and 1027 ($Fuc_3HexNAc_2^+$). Interestingly, the smallest molecular ion in this series that could be detected carried a total of two Fuc (*m/z* 1350), whereas component with five or more Fuc could not be found.

For larger components, electrospray ionization mass spectrometry (ESI-MS) analysis afforded a better signal-to-noise ratio (Figure 6B) as well as doubly charged species that gave an apparent lower *m/z* ratio. The strong signals at *m/z* 2589 and 2763 corresponded to the previously identified smallest multifucosylated O-glycans that carried an additional GalNAc-GlcNAc-Gal repeat unit as compared to the molecular ions at *m/z* 1372/1546/1720. This [GalNAc-GlcNAc-Gal]-$HexNAc_2Gal$-3GalNAcitol set of glycans comprised species with a total of four, five, six, and seven Fuc, which afforded the $[M+Na]^+$ molecular ion signals at *m/z* 2415, 2589, 2763, and 2938, respectively. The next one up in the series, $[GalNAc\text{-}GlcNAc\text{-}Gal]_2$-$HexNAc_2Gal$-3GalNAcitol, could be detected as doubly charged $[M+2Na]^{2+}$ signals at *m/z* 1915 and 2002 with eight and nine Fuc, respectively, whereas fucosylated $[GalNAc\text{-}GlcNAc\text{-}Gal]_3$-$HexNAc_2Gal$-3GalNAcitol was not found.

Other signals detected were consistent with O-glycans with a branched core 2 structure in which the glycan chain attached to C6 of the reducing end GalNAcitol could be represented as $[GalNAc\text{-}GlcNAc\text{-}Gal]_n$-$HexNAc_2$-. Thus, the signals could be assigned as $(GalNAc\text{-}GlcNAc\text{-}Gal)_m HexNAc_2Gal$-3[$(GalNAc\text{-}GlcNAc\text{-}Gal)_n HexNAc_2$-6]GalNAcitol. Weak $[M+Na]^+$ signals for the smallest glycans in this series, that is, $m + n = 0$, were detected at *m/z* 2211, 2385, 2559, 2733, and 2908 carrying a total of four, five, six, seven, and eight Fuc, respectively. For $m + n = 1$, $[M+2Na]^{2+}$ signals were observed at *m/z* 1900 and 1987, carrying a total of 10 and 11 Fuc. For $m + n = 2$, $[M+2Na]^{2+}$ signals were observed at *m/z* 2247, 2334, 2422, and 2509, carrying a total of 9 to 12 Fuc. For $m + n = 3$, $[M+2Na]^{2+}$ signals were observed at *m/z* 2856, 2943, 3030, and 3117, carrying a total of 12 to 15 Fuc. This conclusion was further supported by analysis of the mild-periodate-oxidized products. As listed in Table II, the smallest C_2 and C_4 fragments detected both carried a maximum of four Fuc. Up to three additional repeating units for each could be detected within the sensitivity limit.

Because the larger components detected were virtually identical to those found in previous studies using FAB-MS (Khoo *et al.*, 1995), they were not further pursued here. Instead, we focused on the additionally identified smaller core structures to address a few outstanding structural issues. Smith degradation led to isolation of three major products from normal phase HPLC fractionation, which were identified as $HexNAc_2Hex\text{-}C_4$, $(HexNAc_2Hex)\text{-}HexNAc_2\text{-}C_2$, and $(HexNAc_2Hex)\text{-}HexNAc_2Hex\text{-}C_4$ by MS analysis. The first, $HexNAc_2Hex\text{-}C_4$, was subjected to linkage analysis, which resulted in the identification of terminal GalNAc, 4-linked GlcNAc, and 3-linked Gal. Termination with a lacdiNAc sequence was therefore not restricted to longer chains carrying the repeating units and that the first HexNAc extending from the Gal-GalNAcitol core was indeed a GlcNAc. The other two products were subjected to β-*N*-acetylhexosaminidase digestion to first remove the nonreducing terminal β-HexNAc and then followed by α-galactosidase and another round of β-*N*-acetylhexosaminidase. It was found that the C_2 component was completely digested, whereas the C_4 component yielded $Hex\text{-}C_4$, as shown by MS analysis of the perdeuteroacetyl derivatives. A final CrO_3 oxidation caused the $[M+Na]^+$ molecular ion signal at *m/z* 618 to shift to *m/z* 632, thereby indicating a β linkage. Taken together, the data demonstrated that the smaller multifucosylated O-glycans were indeed based on type 1 and, most likely, type 2 core structures, although the latter could still not be rigorously proven. Interestingly, the β-GlcNAc most proximal to the GalNAcitol was shown here to carry a maximum of only two Fuc, whereas the β-GlcNAc within the repeating unit carried a maximum of three, as shown

Table II. Mild-periodate-oxidized products of the multifucosylated *S. mansoni* cercarial O-glycans

Composition and deduced sequence	No. of Fuc (x)[a]	*m/z* of the $[M+Na]^+$ molecular ions observed[b]
$Fuc_x[HexNAc_2\text{-}Hex]\text{-}C_4$	2 to 4 (3)	1284, **1458**, 1632
$Fuc_x[(HexNAc_2\text{-}Hex)_1\text{-}HexNAc_2\text{-}Hex]\text{-}C_4$	2 to 7 (5)	1979, 2154, 2328, **2502**, 2676, 2850
$Fuc_x[(HexNAc_2\text{-}Hex)_2\text{-}HexNAc_2\text{-}Hex]\text{-}C_4$	4 to 10 (6, 7)	3022, 3196, **3370**, **3544**, 3718, 3893, 4067
$Fuc_x[(HexNAc_2\text{-}Hex)_3\text{-}HexNAc_2\text{-}Hex]\text{-}C_4$	7 to 10[c]	4239, 4413, 4587, 4762
$Fuc_x[HexNAc_2]\text{-}C_2$	2 to 4 (3)	951, **1125**, 1299
$Fuc_x[(HexNAc_2\text{-}Hex)_1\text{-}HexNAc_2]\text{-}C_2$	2 to 6 (5)	1646, 1820, 1994, **2169**, 2343
$Fuc_x[(HexNAc_2\text{-}Hex)_2\text{-}HexNAc_2]\text{-}C_2$	5 to 9 (6, 7)	2863, **3037**, **3211**, 3385, 3559
$Fuc_x[(HexNAc_2\text{-}Hex)_3\text{-}HexNAc_2]\text{-}C_2$	8 to 11[c]	4080, 4254, 4428, 4603

[a]The number of total Fuc carried on the most abundant component(s) within each series is shown in parentheses.
[b]The most intense peak(s) detected within each series is in bold print.
[c]The detection of a full range of fucosylated components was limited by sensitivity.

previously. The composite structures for the multifucosylated O-glycans are shown in Figure 7B and 7C, together with a representative structure for the Le^x-containing O-glycans (Figure 7A).

Discussion

In addition to various core modified N-glycans, the digenetic schistosomes apparently synthesized a remarkable range of developmental stage–specific O-glycans of unusual structures. Pioneering work by Cummings's group (Nyame *et al.*, 1987, 1988) have demonstrated that the *S. mansoni* adult worms and schistosomula synthesize only short and simple O-GlcNAc, GalNAc, and Gal-GalNAc, based on characterization of metabolically radiolabeled glycans from total glycoprotein extracts. Our own MS analysis of the total O-glycans from the adult stage indicates that the major molecular ions detected correspond to Hex-HexNAcitol and $Hex_2HexNAc_2$itol (unpublished data). Notably, no fucosylated structure has been detected, nor has the $Hex_2HexNAc$itol structure characterized in this work. However, other work has demonstrated that the adult worm is capable of *O*-glycosylating its secreted protein in a highly unusual manner.

The gut-associated antigens regularly released by the adult worms into the host circulation have long been recognized to carry highly immunoreactive glycans. Both the circulating cathodic antigen (CCA) and anodic antigen (CAA) isolated from adult *S. mansoni* worms have been characterized in detail with respect to their *O*-glycosylation (Bergwerff *et al.*, 1994; van Dam *et al.*, 1994). From CCA, the O-glycans released by alkaline reductive elimination were fractionated by gel-filtration into two main pools—a major polysaccharide fraction and a smaller amount of oligosaccharides. The oligosaccharides were characterized as being based on mucin-type core 1 and 2 structures elongated by lacNAc and possibly Le^x structures. Among the incompletely defined structures was the unusual $Hex_2HexNAc$itol branched core extended by a Hex-HexNAc unit. The branching nature was indicated by MS/MS analysis but not further defined (van Dam *et al.*, 1994). The polysaccharide fraction was shown to comprise poly-*N*-acetyllactosamine chains where the internal and nonreducing terminal repeating lacNAc units are, respectively, near 100% and 80% fucosylated to yield poly-Le^x epitopes. It was suggested that these polysaccharide chains of an estimated 25 repeating units are conjugated to the protein backbone via core types 1 and 2, namely, an extension from the characterized oligosaccharides. The CAA, on the other hand, was shown to contain polymeric chains of 6-linked β-GalNAc backbone with a β-glucuronic acid residue attached to its C3 position. The mode of linkage to protein core was not established (Bergwerff *et al.*, 1994).

The possible presence of both conventional type 1 and 2 cores as well as novel branched cores in the schistosomal O-glycans was further noted by MS analysis of the total egg O-glycans (Khoo *et al.*, 1997). In particular, it was shown that $Hex_2HexNAc$itol is a prominent component in both *S. mansoni* and *S. japonicum* egg extracts; this observation was subsequently extended to the miracidia stage (unpublished data). Using the periodate oxidation strategy as employed here, it was further demonstrated that for *S. mansoni* egg O-glycans, a stretch of up to four HexNAcs could be directly linked to the C6 position of the GalNAcitol, probably through core 2 structure. However, it was not clear from the data if extension on the 6-arm could proceed via a Hex in a manner similar to the one characterized here. The molecular composition defined was interpreted then as a series of variably fucosylated $Hex_1HexNAc_1$, $HexNAc_2$, and $HexNAc_2Hex_1$ units extending from a Hex-HexNAcitol core (Khoo *et al.*, 1997). In retrospect, this simplistic interpretation belies a more complicated picture in the light of current findings, because all possible core structures and branching patterns may exist in the egg O-glycans.

The cercarial O-glycans as characterized here and previously (Khoo *et al.*, 1995) are remarkable in many aspects. First, the two sets of O-glycans are clearly based on two distinct core structures. The multifucosylated chains are attached to type 1 and 2 cores, whereas the Le^x structures are attached via the novel core. Furthermore, these two sets appeared to be carried on two distinct populations of glycoproteins that could be separated by the AA lectin column. It also indicated that for the glycoproteins to be retained by the AA lectin, a higher degree of fucosylation than just one or two Le^x epitopes are required.

Second, extreme heterogeneity was associated with each set. In the case of the multifucosylated O-glycans, heterogeneity

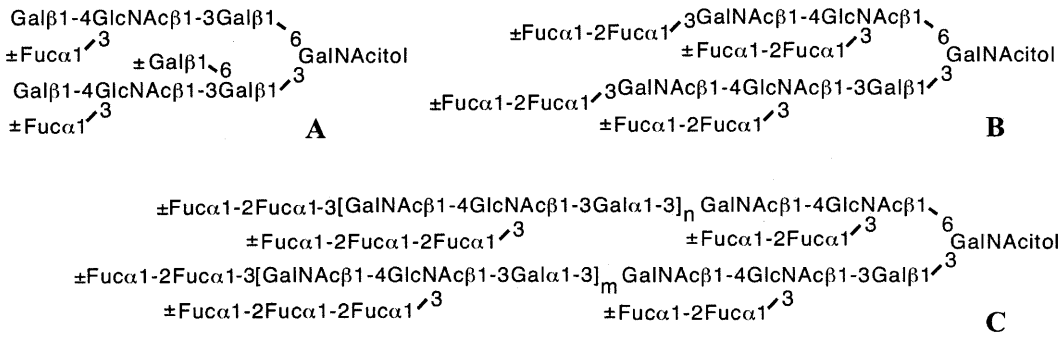

Fig. 7. *S. mansoni* cercarial O-glycans. The smaller Le^x-containing O-glycans (**A**) are based on a novel core. The 6-arm could be absent or extended with a single β-Gal. Glycoproteins containing only this type of O-glycans were not retained by the AA lectin column. The multifucosylated larger O-glycans (**C**) were characterized in previous work (Khoo *et al.*, 1995) but the smaller ones (**B**) without the additional repeat units were found only in this work. Structures in (**B**) with and without the 6-arm were found to contain a maximum of eight and four Fuc residues, respectively. Unlike the GlcNAc in the repeating units of (**C**), the innermost GlcNAc in (**B**) appeared to carry only a maximum of 2 Fuc.

was primarily due to (1) different degree of fucosylation, (2) presence or absence of the glycosyl chain on the 6-arm, and (3) the number of repeating units on each arm. This work extended the findings from previous work (Khoo *et al.*, 1995) by identifying the smallest structures within this series. It is remarkable that the smallest branched structure carries a maximum of eight Fuc on a hexasaccharide that would probably qualify for glycoprotein derived glycans with a highest density of fucosylation. Intriguingly, the inner HexNAc does not carry a third Fuc. Assuming a maximum of two Fuc for the innermost HexNAc, three Fuc for other GlcNAcs, and two Fuc for the nonreducing terminal GalNAc, the range of fucosylated components detected implies that further extension of the smaller structures with the repeating units seems to be associated with higher degree of underfucosylation. It is also possible that the underfucosylated structures represent degradative products due to actions of fucosidase during sample extraction and preparation.

For the Le^x-containing O-glycans, the heterogeneity is best viewed as a series of incompletely glycosylated structures as compared to the largest structure characterized, namely, fully fucosylated and fully branched on both Gal and GalNAcitol (Figure 7A). O-glycans with a single β-Gal on the C6 position of the GalNAcitol of type 1 core structures have been identified in human gastric mucins (Slomiany *et al.*, 1984a,b). To our knowledge, further extension via this β-Gal has not been previously reported. The success in identifying such biantennary-like structures were largely due to the mild periodate strategy employed. Prior to this work, normal Smith degradation has been employed to probe the backbone structures of glycosyl chains attached respectively to the 3- and 6-arm of the GalNAcitol (Khoo *et al.*, 1995, 1997). By adopting the mild periodate oxidation method used to selectively cleave only the C4–C5 bond of the GalNAcitol residue (Stoll *et al.*, 1990; Chai *et al.*, 1993), we have unambiguously demonstrated the existence of branching on the C6 of GalNAcitol via a Gal residue. Under the experimental conditions employed, the single Gal residue survived as Gal-C_2, which could be readily detected by MS analysis as perdeuteroacetyl derivatives.

It is not known if such novel core structure is wide spread in nature because the only other reported case in human gastric mucins has been called into question (Hanisch *et al.*, 1993). However, at least in *S. mansoni*, such novel core structures could be detected among the O-glycans synthesized by the adult worms (in the excreted CCA), eggs, and miracidia, in addition to the cercaria. Our MS analyses have further indicated that the Hex_2HexNAcitol entity could also be found among the O-glycans of the cercarial and egg extracts of the other two schistosome species, that is,. *S. japonicum* and *S. haematobium* (unpublished data). Thus, this core type is probably as common as types 1 and 2 for the trematodes, although further data are needed to support this observation. The identification of an entirely new core structure also serves as a cautionary note to structural studies of lower organisms, including other parasitic helminths, for which very little structural information with respect to the O-glycans is available.

Comparing the Le^x-containing cercarial O-glycan structures characterized in this work with those of the cercarial N-glycans (Khoo *et al.*, 2001), it is interesting to note that the →Galβ1→3 (→Galβ1→6)GalNAc core sequence essentially takes the place of the →Manα1→3(→Manα1→6)Manβ1→ sequence in the trimannosyl core of N-glycans. Just as the monoantennary or hybrid type N-glycans could have the single Man on the 6-arm not further extended, the single Gal on the 6-arm in this series of O-glycans could be similarly not extended. Furthermore, in a fashion analogous to chain elongation in N-glycans, the β-Gal on both 3- and 6-arms could be extended by lacNAc and then fucosylated to give the Le^x structures.

The identification of Le^x on the cercarial O-glycans now completed the picture that Le^x can be carried on N-glycans (Khoo *et al.*, 2001), O-glycans, and glycosphingolipids (Wuhrer *et al.*, 2000). The exposure of this epitope on the cercarial surface may, however, be masked by the more extensively fucosylated larger structures carried on conventional type 1 and 2 structures. It is predicted that the egg O-glycans would likewise make up a mixture of these two distinct sets of structures, as well as possibly other as yet characterized cores and terminal epitopes. On the other hand, despite some indication by monoclonal antibody probing (van Die *et al.*, 1999; van Remoortere *et al.*, 2000), we have consistently failed to detect any unusually fucosylated or modified cores on the glycans derived from the adult worms. The implicated developmental regulation is therefore quite distinctive, if not absolute.

Materials and methods

Parasite materials

Biomphalaria glabrata albino snails infected with a Puerto Rican strain of *S. mansoni* were obtained from Dr. Fred Lewis of the Bethesda Biomedical Research Institute. Cercariae were shed by exposing the snails to light for 2 h and collected by centrifugation.

Extraction of glycoprotein and preparation of O-glycans

Cercarial glycoproteins were extracted by phenol/water partition method as described (Xu *et al.*, 1994). Dialyzed and lyophilized glycoprotein extracts were then redissolved in 20 mM Tris buffer (pH 7.6–7.8, with 1 mM each of $CaCl_2$, $MgCl_2$, $MnCl_2$, and 0.1 M NaCl) and loaded onto a lectin affinity column (*A. anguilla* gel, EY laboratories) equilibrated in the same buffer. Nonbound glycoproteins were eluted with two column volumes of the same Tris buffer, while bound glycoproteins were subsequently eluted with 100 mM fucose in the same Tris buffer. Elution was monitored online by UV absorbance at 280 nm. Fractions collected were pooled accordingly, dialyzed against water, and then lyophilized.

The crude glycoprotein extracts obtained were first digested with non–TPCK treated trypsin (Roche) (1:50 enzyme: protein ratio, w/w; 37°C for 4 h in 50 mM ammonium hydrogen carbonate buffer, pH 8.4) followed by non–TLCK treated chymotrypsin in the same conditions. The digested glycopeptides were loaded onto C18 Sep-pak® cartridge (Waters). Glucan polymer and other hydrophilic contaminants were washed off with 5% aqueous acetic acid, and the bound peptides were eluted with a step gradient of 20%, 40%, and 60% 1-propanol in water. All the eluted fractions were pooled, dried down, and then incubated with *N*-glycosidase F (5 units, Roche) overnight at 37°C in 50 mM ammonium bicarbonate buffer, pH 8.4. Released N-glycans were separated from peptides/glycopeptides using the same C18 Sep-pak procedure. O-glycans were reductively eliminated from the pooled 20–60% 1-proponal

fractions with 1 M sodium borohydride solution in 0.05 M aqueous sodium hydroxide (45°C, overnight) and desalted by passing through a Dowex (50W-X8, 50–100 mesh, protonated form) column in 5% acetic acid. Borates were removed by repeated coevaporation with 10% acetic acid in methanol.

Normal-phase HPLC separation of O-linked glycans

HPLC analysis was performed on a Hewlett Packard 1100 series LC equipped with a thermostatted column compartment. The *O*-linked glycans were size fractionated on a PalPak type N column (Takara, 4.6 × 250 mm) at a flow rate of 1 ml/min, 30°C. Buffers A and B contain 25:75 and 50:50 stock buffer: acetonitrile (v/v), respectively, where stock buffer was aqueous solution containing 10% acetonitrile and 3% acetic acid, titrated to pH 7.3 with triethylamine. Column was equilibrated and maintained at 10% B for 5 min, then linearly increased to 100% B in 40 min, and kept at 100% B for another 15 min. In normal-phase HPLC analysis, PA-labeled isomalto-oligomers prepared from partially hydrolyzed dextran were used to calibrate the elution positions as Glc units.

Sequential exoglycosidase digestions

The O-glycans were digested with exo-glycosidases using the following conditions: α1-3,4 fucosidase (from *Xanthomonas manihotis*, New England Biolabs): 5 U in 55 µl of 50 mM sodium citrate buffer, pH 6.0; β-galactosidase (from *Diplococcus pneumoniae*, Roche): 5 mU in 50 µl of 50 mM sodium acetate buffer, pH 6.0; β-galactosidase (from jack bean, Seikagaku, Japan): 125 mU in 50 µl of 50 mM sodium acetate buffer, pH 3.5 for 48 h; *N*-acetyl-β-D-glucosaminidase (from *Diplococcus pneumoniae*, Roche): 5 mU in 50 µl of 50 mM sodium acetate buffer, pH 6.0. All digestions were carried out at 37°C for 24 h, except for β-galactosidase from jack bean. Each enzyme digestion was desalted by passing through a mixed bed ion exchange column packed with 1 ml each of Dowex (50W-X8, 50–100 mesh, protonated form) and AG 3-X4 (AG 3-X4, 100–200 mesh, free base form, Bio-Rad) resins prior to MS analysis.

Periodate oxidation

For mild oxidation, the oligoglycosyl alditols were treated with 200 nmol of sodium meta-periodate (Merck) in 40 mM imidazole–HCl (pH 6.5) at 4°C in the dark for 45 min. Excess oxidant was reacted with ethylene glycol (2 µmol) under the same conditions for an additional 40 min. The reaction product was then lyophilized and reduced with 10 mg/ml sodium borohydride in 2 M NH_4OH for 2 h at room temperature. The reaction was then terminated by adding a few drops of glacial acetic acid and desalted by passing through a Dowex (50W-X8, 50–100 mesh, protonated form) column in 5% acetic acid. Borates were removed by repeated coevaporation with 10% aqueous acetic acid in methanol. For normal periodate oxidation, the samples were incubated with 50 µl of 50 mM sodium m-periodate in 100 mM ammonium acetate buffer (pH 5.5) at room temperature for 4 h in the dark. The reaction was then terminated and processed as with mild periodate oxidation.

Chemical derivatization and MS analysis

Samples were permethylated using the NaOH/dimethyl sulfoxide slurry method as described by Dell *et al.* (1994) or perdeuteroacetylated using 1:1 pyridine:deuteroacetic anhydride (2 h at 80°C). For FAB-MS analysis, chemical derivatives of the O-glycans were redissolved in CH_3OH for loading onto the probe tip coated with 1-monothioglycerol as matrix for positive ion modes. Glycerol:*m*-nitrobenzylalcohol: trifluoroacetic acid (50:50:1, v/v/v) matrix was used in MS-MS studies to promote $[M+H]^+$ species as the preferred parent ions. FAB-mass spectra were acquired on an Autospec orthogonal acceleration-time of flight mass spectrometer (Micromass, UK) fitted with a cesium ion gun operating at 26 kV. CID MS-MS was performed by introducing argon gas to the collision cell to a reading of ~1.2×10^{-6} millibars on the time of flight ion gauge. The source accelerating voltage was at 8 kV, with a push-out frequency of 56 kHz for orthogonal sampling. A 1-s integration time per spectrum was chosen for the time of flight analyzer with a 0.1-s interscan delay. Individual spectra were summed for data processing. ESI-MS analysis was performed on the same instrument fitted with an electrospray ionization source in place of the FAB source and operated at 4 kV accelerating voltage. Derivatized samples were introduced into the electrospray source by loop injection and delivered at a flow rate of 5 µl /min in 50:50:1 (v/v/v) methanol/water/acetic acid for positive-ion mode analysis.

Chromium trioxide oxidation

Perdeuteroacetylated oligosaccharide samples were dissolved in 100 µl of glacial acetic acid, followed by the addition of ~10 mg of chromium trioxide (CrO_3, RDH). The resulting suspension was stirred at 50°C for 2 h and then quenched with water (2 ml). The reaction mixtures were extracted with chloroform and washed several times with water until the aqueous layer was colorless. The resulting samples were dried down under a stream of nitrogen and redissolved in methanol for FAB-MS analysis.

Monosaccharide composition and linkage analysis

For gas chromatography (GC)-MS linkage analysis, partially methylated alditol acetates were prepared from permethyl derivatives by hydrolysis (2 M trifluoroacetic acid, 121°C, 2 h), reduction (10 mg/ml $NaBH_4$, 25°C, 2 h), and acetylation (acetic anhydride, 100°C, 1 h). GC-MS was carried out using a Hewlett-Packard Gas Chromatograph 6890 connected to a HP 5973 Mass Selective Detector. The sample was dissolved in hexane prior to splitless injection into a HP-5MS fused silica capillary column (30 m × 0.25 mm I.D., HP). The column head pressure was maintained at around 8.2 psi to give a constant flow rate of 1 ml/min using helium as carrier gas. Initial oven temperature was held at 60°C for 1 min, increased to 90°C in 1 min, and then to 290°C in 25 min. For monosaccharide composition analysis, released *N*-glycans were methanolyzed with 0.5 M methanolic-HCl (Supelco) at 80°C for 16 h; re-*N*-acetylated with 500 µl of methanol, 10 µl of pyridine, and 50 µl of acetic anhydride; and then treated with the Sylon HTP® trimethylsilylating reagent (Supelco) for 20 min at room temperature, dried down, and redissolved in hexane. GC-MS analysis of the trimethylsilylated derivatives was performed on the same HP system using a temperature gradient of 60°C to 140°C at 25°C/min, increased to 250°C at 5°C/min, and then increased to 300°C at 10°C/min.

Acknowledgments

We gratefully acknowledge Dr. John P. Caulfield (Roche Bioscience, Palo Alto, CA) for provision of the *S. mansoni* cercariae sample and Dr. Delphi Chatterjee (Department of Microbiology, Colorado State University, CO) for her initial involvement in the preparation of cercarial glycoprotein extracts. This work was supported by the Academia Sinica and National Science Council (Taiwan, R.O.C.) Grants NSC 87–2311-B-001–091 and NSC 88–2311-B-001–117 to KHK.

Abbreviations

AA, *Anguilla anguilla*; CAA, circulating anodic antigen; CCA, circulating cathodic antigen; CID, collision induced dissociation; ESI, electrospray ionization; FAB, fast-atom-bombardment; GC-MS, gas chromatography mass spectrometry; HPLC, high-pressure liquid chromatography; LacNAc, *N*-acetyllactosamine; Le^x, Lewis X; MS, mass spectrometry.

References

Bergwerff, A.A., van Dam, G.J., Rotmans, J.P., Deelder, A.M., Kamerling, J.P., and Vliegenthart, J.F. (1994) The immunologically reactive part of immunopurified circulating anodic antigen from *Schistosoma mansoni* is a threonine-linked polysaccharide consisting of ∅ 6)-(β-D-Glc*p*A-(1∅ 3))-β-D-Gal*p*NAc-(1∅ repeating units. *J. Biol. Chem.*, **269**, 31510–31517.

Brockhausen, I. (1995) Biosynthesis of O-glycans of the N-acetylgalactosamine-α-Ser/Thr linkage type. In Montreuil, J., Vliegenthart, J.F.G., and Schachter, H. (eds), *Glycoproteins*. Elsevier *Science* B.V., Amsterdam, Vol. 29a, pp. 201–259.

Caulfield, J.P., Cianci, C.M.L., McDiarmid, S.S., Suyemitsu, T., and Schmid, K. (1987) Ultrastructure, carbohydrate, and amino acid analysis of two preparations of the cercarial glycocalyx of *Schistosoma mansoni*. *J. Parasitol.*, **73**, 514–522.

Chai, W., Stoll, M.S., Cashmore, G.C., and Lawson, A.M. (1993) Specificity of mild periodate oxidation of oligosaccharide-alditols: relevance to the analysis of the core-branching pattern of O-linked glycoprotein oligosaccharides. *Carbohydr. Res.*, **239**, 107–115.

Cummings, R.D., and Nyame, A.K. (1999) Schistosome glysoconjugates. *Biochim. Biophys. Acta*, **1455**, 363–374.

Dalton, J.P., Lewis, S.A., Aronstein, W.S., and Strand, M. (1987) *Schistosoma mansoni*: immunogenic glycoproteins of the cercarial glycocalyx. *Exp. Parasitol.*, **63**, 215–226.

Dell, A. (1987) F.A.B.-mass spectrometry of carbohydrates. *Adv. Carbohydr. Chem. Biochem.*, **45**, 19–72.

Dell, A., Haslam, S.M., Morris, H.R., and Khoo, K.H. (1999) Immunogenic glycoconjugates implicated in parasitic nematode diseases. *Biochim. Biophys. Acta*, **1455**, 353–362.

Dell, A., Reason, A.J., Khoo, K.H., Panico, M., McDowell, R.A., and Morris, H.R. (1994) Mass spectrometry of carbohydrate-containing biopolymers. *Methods Enzymol.*, **230**, 108–132.

Hanisch, F.G., Chai, W., Rosankiewicz, J.R., Lawson, A.M., Stoll, M.S., and Feizi, T. (1993) Core-typing of O-linked glycans from human gastric mucins. Lack of evidence for the occurrence of the core sequence Gal1-6GalNAc. *Eur. J. Biochem.*, **217**, 645–655.

Khoo, K.-H., Chatterjee, D., Caulfield, J.P., Morris, H.R., and Dell, A. (1997) Structural mapping of the glycans from the egg glycoproteins of *Schistosoma mansoni* and *Schistosoma japonicum*: identification of novel core structures and terminal sequences. *Glycobiology*, **7**, 663–677.

Khoo, K.-H., Huang, H.-H., and Lee, K.-M. (2001) Characteristic structural features of schistosome cercarial N-glycans : expression of Lewis X and core xylosylation. *Glycobiology*, in press.

Khoo, K.-H., Sarda, S., Xu, X., Caulfield, J.P., McNeil, M.R., Homans, S.W., Morris, H.R., and Dell, A. (1995) A unique multifucosylated – 3GalNAcβ1∅4GlcNAcβ1∅3Galα1- motif constitutes the repeating unit of the complex *O*-glycans derived from the cercarial glycocalyx of *Schistosoma mansoni*. *J. Biol. Chem.*, **270**, 17114–17123.

Ko, A.I., Drager, U.C., and Harn, D.A. (1990) A *Schistosoma mansoni* epitope recognized by a protective monoclonal antibody is identical to the stage-specific embryonic antigen 1. *Proc. Natl Acad. Sci. USA*, **87**, 4159–4163.

Köster, B., and Strand, M. (1994) *Schistosoma mansoni*: immunolocalization of two different fucose-containing carbohydrate epitopes. *Parasitology*, **108**, 433–446.

Mårtensson, S., Levery, S.B., Fang, T.T., and Bendiak, B. (1998) Neutral core oligosaccharides of bovine submaxillary mucin. Use of tetraacetate in the cold for establishing branch position. *Eur. J. Biochem.*, **258**, 603–622.

Nyame, K., Cummings, R.D., and Damian, R.T. (1987) *Schistosoma mansoni* synthesizes glycoproteins containing terminal *O*-linked *N*-acetylglucosamine residues. *J. Biol. Chem.*, **262**, 7990–7995.

Nyame, K., Cummings, R.D., and Damian, R.T. (1988) Characterization of the *N*- and *O*-linked oligosaccharides in glycoproteins synthesized by *Schistosoma mansoni* schistosomula. *J. Parasitol.*, **74**, 562–572.

Samuelson, J.C., and Caulfield, J.P. (1985) The cercarial glycocalyx of *Schistosoma mansoni*. *J. Cell Biol.*, **100**, 1423–1434.

Slomiany, A., Zdebska, E., and Slomiany, B.L. (1984a) Structures of the neutral oligosaccharides isolated from A-active human gastric mucin. *J. Biol. Chem.*, **259**, 14743–14749.

Slomiany, B.L., Zdebska, E., and Slomiany, A. (1984b) Structural characterization of neutral oligosaccharides of human H^+Le^{b+} gastric mucin. *J. Biol. Chem.*, **259**, 2863–2869.

Stoll, M.S., Hounsell, E.F., Lawson, A.M., Chai, W.G., and Ten, F.Z. (1990) Microscale sequencing of O-linked oligosaccharides using mild periodate oxidation of alditols, coupling to phospholipid and TLC-MS analysis of the resulting neoglycolipids. *Eur. J. Biochem.*, **189**, 499–507.

van Dam, G.J., Bergwerff, A.A., Thomas-Oates, J.E., Rotmans, J.P., Kamerling, J.P., Vliegenthart, J.F., and Deelder, A.M. (1994) The immunologically reactive O-linked polysaccharide chains derived from circulating cathodic antigen isolated from the human blood fluke *Schistosoma mansoni* have Lewis X as repeating unit. *Eur. J. Biochem.*, **225**, 467–482.

van Die, I., Gomord, V., Kooyman, F.N., van den Berg, T.K., Cummings, R.D., and Vervelde, L. (1999) Core α1∅3-fucose is a common modification of *N*-glycans in parasitic helminths and constitutes an important epitope for IgE from *Haemonchus contortus* infected sheep. *FEBS Lett.*, **463**, 189–193.

van Remoortere, A., Hokke, C.H., van Dam, G.J., van Die, I., Deelder, A.M., and van den Eijnden, D.H. (2000) Various stages of *Schistosoma* express $Lewis^x$, LacdiNAc, GalNAcβ1–4(Fucα1–3)GlcNAc and GalNAcβ1–4(Fucα1–2Fucα1–3)GlcNAc carbohydrate epitopes: detection with monoclonal antibodies that are characterized by enzymatically synthesized neoglycoproteins. *Glycobiology*, **10**, 601–609.

Weiss, J.B., and Strand, M. (1985) Characterization of developmentally regulated epitopes of *Schistosoma mansoni* egg glycoprotein antigens. *J. Immunol.*, **135**, 1421–1429.

Weiss, J.B., Magnani, J.L., and Strand, M. (1986) Identification of *Schistosoma mansoni* glycolipids that share immunogenic carbohydrate epitopes with glycoproteins. *J. Immunol.*, **136**, 4275–4282.

Wuhrer, M., Dennis, R.D., Doenhoff, M.J., Lochnit, G., and Geyer, R. (2000) *Schistosoma mansoni* cercarial glycolipids are dominated by Lewis X and pseudo-Lewis Y structures. *Glycobiology*, **10**, 89–101.

Xu, X., Stack, R.J., Rao, N., and Caulfield, J.P. (1994) *Schistosoma mansoni*: Fractionation and characterization of the glycocalyx and glycogen-like material from cercariae. *Exp. Parasitol.*, **79**, 399–409.