The EMBO Journal Vol.16 No.14 pp.4302–4310, 1997

The solution NMR structure of glucosylated *N*-glycans involved in the early stages of glycoprotein biosynthesis and folding

A.J.Petrescu, T.D.Butters, G.Reinkensmeier, S.Petrescu, F.M.Platt, R.A.Dwek[1] and M.R.Wormald[1]

Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

[1]Corresponding authors
e-mail: mark@oxglua.glycob.ox.ac.uk

Glucosylated oligomannose *N*-linked oligosaccharides ($Glc_xMan_9GlcNAc_2$ where $x = 1–3$) are not normally found on mature glycoproteins but are involved in the early stages of glycoprotein biosynthesis and folding as (i) recognition elements during protein *N*-glycosylation and chaperone recognition and (ii) substrates in the initial steps of *N*-glycan processing. By inhibiting the first steps of glycan processing in CHO cells using the α-glucosidase inhibitor *N*-butyl-deoxynojirimycin, we have produced sufficient $Glc_3Man_7GlcNAc_2$ for structural analysis by nuclear magnetic resonance (NMR) spectroscopy. Our results show the glucosyl cap to have a single, well-defined conformation independent of the rest of the saccharide. Comparison with the conformation of $Man_9GlcNAc_2$, previously determined by NMR and molecular dynamics, shows the mannose residues to be largely unaffected by the presence of the glucosyl cap. Sequential enzymatic cleavage of the glucose residues does not affect the conformation of the remaining saccharide. Modelling of the $Glc_3Man_9GlcNAc_2$, $Glc_2Man_9GlcNAc_2$ and $Glc_1Man_9GlcNAc_2$ conformations shows the glucose residues to be fully accessible for recognition. A more detailed analysis of the conformations allows potential recognition epitopes on the glycans to be identified and can form the basis for understanding the specificity of the glucosidases and chaperones (such as calnexin) that recognize these glycans, with implications for their mechanisms of action.

Keywords: glucosylated glycans/NMR/oligosaccharide structure/protein folding

Introduction

N-linked glycosylation of proteins is a highly conserved process in eukaryotic evolution. The oligosaccharides $Glc_xMan_9GlcNAc_2$ (Figure 1), where $x = 1–3$, are involved in a number of important steps during the biosynthesis and folding of glycoproteins, including: (i) the *N*-glycosylation of the nascent peptide chain in the endoplasmic reticulum (ER) lumen [involving transfer of $Glc_3Man_9GlcNAc_2$ to the peptide by the oligosaccharyltransferase (OST) complex]; (ii) the initial steps of glycan processing on the glycoprotein in the ER (involving removal of the terminal glucose residue from $Glc_3Man_9GlcNAc_2$ by α-glucosidase I and from $Glc_2Man_9GlcNAc_2$ and $Glc_1Man_9GlcNAc_2$ by α-glucosidase II); and (iii) the chaperone-dependent folding of glycoproteins in the ER (involving recognition of $Glc_1Man_9GlcNAc_2$ by proteins such as calnexin and calreticulin).

Although the presence of an oligosaccharide may influence the physical and biological characteristics of a protein (Dwek, 1996), a primary role for *N*-linked glycosylation in many simple eukaryotes may have been to retain the protein in the ER so that it folds correctly. Oligosaccharide motifs, based on $Glc_xMan_9GlcNAc_2$ structures, would then act as recognition signals for lectins/chaperones located in the ER.

The requirement for terminal glucose residues as part of the oligosaccharide during protein *N*-glycosylation is species dependent. In mammalian cells, only those glycans containing three glucose residues are transferred to protein, even in mutant cell lines that synthesize truncated $Glc_3Man_5GlcNAc_2$ oligosaccharides (Kornfeld *et al.*, 1979). The non-conditional *alg3* mutant yeast strain can only synthesize lipid-linked $Man_5GlcNAc_2$, resulting in reduced glycoprotein biosynthesis but no reduction in total secreted protein (Verostek *et al.*, 1993). Trypanosomatid protozoa cannot glucosylate oligosaccharides, resulting in the normal transfer of non-glucosylated oligosaccharides to the peptide (Parodi, 1993).

The full role of the glucose residues during glycosylation is not well understood. *In vitro* experiments have shown that these residues produce a 10-fold increase in the apparent affinity constant of OST for the oligosaccharide (Breuer and Bause, 1995). The authors also suggest that the recognition and binding of outer arm glucose residues induce conformational changes in the active site of OST, influencing the association constant of the peptide substrate (Breuer and Bause, 1995).

The first stages of glycan processing involve sequential removal of the terminal glucose residues from the precursor *N*-glycan $Glc_3Man_9GlcNAc_2$ by α-glucosidase I and II. The $Glc_1Man_9GlcNAc_2$ glycan, formed after cleavage by α-glucosidase I and II, mediates interaction of the glycoprotein with the ER-resident chaperones calnexin and calreticulin (Ou *et al.*, 1993; Hebert *et al.*, 1995; Ware *et al.*, 1995). These chaperones both retain the glycoprotein in the ER during folding, and promote folding and oligomerization (Bergeron *et al.*, 1994; Arunachalam and Cresswell, 1995; Zhang *et al.*, 1995). Removal of the final glucose residue by α-glucosidase II results in a loss of interaction of the glycoprotein with the chaperones. However, incorrectly folded proteins can be reglucosylated in the ER and so be recycled through the chaperone system (Sousa *et al.*, 1992; Helenius, 1994).

Glucosidase inhibitors, notably *N*-butyl-deoxynojirimy-

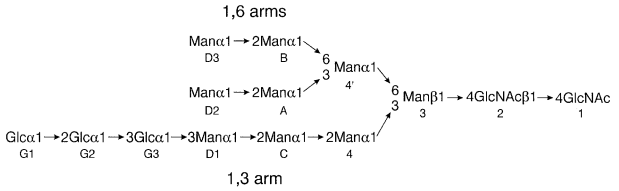

Fig. 1. Schematic representation of the $Glc_3Man_9GlcNAc_2$ structure showing the primary sequence and the residue numbering scheme used. The $Glc_3Man_7GlcNAc_2$ structure is missing residues D2 and D3. The $Glc_2Man_7GlcNAc_2$ and $Glc_1Man_7GlcNAc_2$ structures are missing residues G1 and G2 respectively as well. The $Glc_3Man_4GlcNAc_1$ structure is missing residues D1, D2, A, B, 4′ and 1.

cin (NB-DNJ), prevent glycan processing and thus prevent glycoproteins interacting with chaperones such as calnexin and calreticulin. This has been shown to produce proteins which are misfolded, leading to alterations in biological activity (Fischer *et al.*, 1996; Mehta *et al.*, 1997; Petrescu *et al.*, 1997).

We have utilized NB-DNJ-treated cells to produce a glucosylated oligosaccharide ($Glc_3Man_7GlcNAc_2$) in sufficient quantities for conformational analysis by NMR spectroscopy. We find that the triglucosyl cap has a well-defined, compact and relatively rigid conformation. The conformation of the mannose residues is very similar to that previously determined for $Man_9GlcNAc_2$ (R.J.Woods, A.Pathiaseril, M.R.Wormald, C.J.Edge and R.A.Dwek, submitted). Further analysis shows that sequential removal of each glucose residue by α-glucosidase I and II has no discernible structural effect on the remaining residues. A detailed analysis of the structural features of these saccharides has significant implications in understanding the specificity of the interactions which occur between this *N*-glycan structure and the proteins which interact with this structure during glycoprotein folding and biosynthesis.

Results

Isolation of triglucosylated saccharides from NB-DNJ-treated cells

Transformed Chinese hamster ovary (CHO) cells secreting recombinant gp120 were cultivated for 1 week in the presence of the α-glucosidase inhibitor NB-DNJ to obtain a culture medium enriched in hyperglucosylated glycoproteins. Glycoproteins with oligomannose-type glycans were separated by lectin affinity chromatography and the glycans were released using hydrazine and labelled with 2-aminobenzamide (2-AB). Only a small proportion of the total 2-AB-labelled material is hyperglucosylated oligosaccharide (Figure 2a). The hyperglucosylated species found were $Glc_3Man_9GlcNAc_2$, $Glc_3Man_8GlcNAc_2$ and $Glc_3Man_7GlcNAc_2$ in the ratio 1:1.5:7 respectively. Only $Glc_3Man_7GlcNAc_2$ could be prepared in sufficient quantity for NMR structural analysis. There are two exomannosidases capable of hydrolysing α1→2-linked terminal mannose residues resident in the ER (Bause *et al.*, 1992). Thus, $Glc_xMan_7GlcNAc_2$ structures are potential natural substrates for glucosidase II, calnexin and calreticulin.

Pure glucosylated oligosaccharides were obtained by initial fractionation of the hydrazine-released material using Bio-Gel P4 chromatography, followed by normal phase HPLC of the appropriate fraction (Figure 2b). This

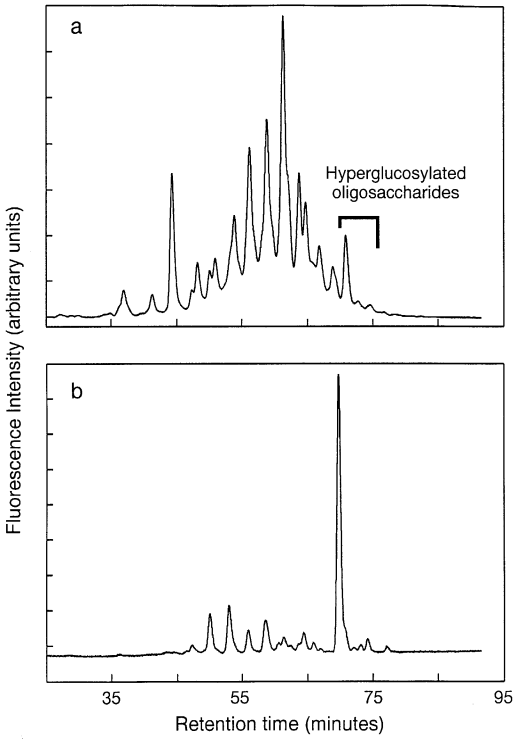

Fig. 2. (**a**) Analytical normal phase HPLC trace of the hydrazine-released and 2-AB-labelled material obtained from NB-DNJ-treated CHO cells after lectin affinity chromatography. The hyperglucosylated oligosaccharides elute between 70 and 75 min. (**b**) Analytical normal phase HPLC trace of the $Glc_3Man_7GlcNAc_2$ fraction obtained by Bio-Gel P4 chromatography. $Glc_3Man_7GlcNAc_2$ elutes at 70 min.

procedure was carried out on a preparative scale using unlabelled material and identifying fractions by mass spectrometry. The final sample purity was checked by mass spectrometry and NMR spectroscopy. All samples used for NMR analysis were >95% pure. From 10 g (dry weight) of material taken for hydrazinolysis, 3 mg of $Glc_3Man_7GlcNAc_2$ were obtained.

Linkage conformation analysis of the Glc_3Man structural unit

Resonance spin-system assignments (Table I) were obtained from two-dimensional COSY and RELAY spectra. This enabled most of the resonances from the three glucose residues and the C1H to C4H resonances of the mannose residues to be identified. A few further resonances could be assigned to the glucose spin-systems from the TOCSY spectrum. Sequence-specific assignments were made on the basis of the pattern of cross-linkage nuclear Overhauser effects (NOEs) (Table II). For the mannose residues, these could be confirmed by comparison of the anomeric proton chemical shifts with those of oligomannose oligosaccharides (see Materials and methods).

Figure 3 shows the cross-sections through the NOESY spectrum at the chemical shifts of the three glucose anomeric protons. The initial build-up rates, together with the distance constraints derived from the NOE data, are given in Table II. All three linkages (Glcα1→2Glcα, Glcα1→3Glcα and Glcα1→3Manα) are unusual in that only one NOE is observed for each linkage at short mixing times (<200 ms). Thus, the conformational information available from the absence of NOEs is of great importance,

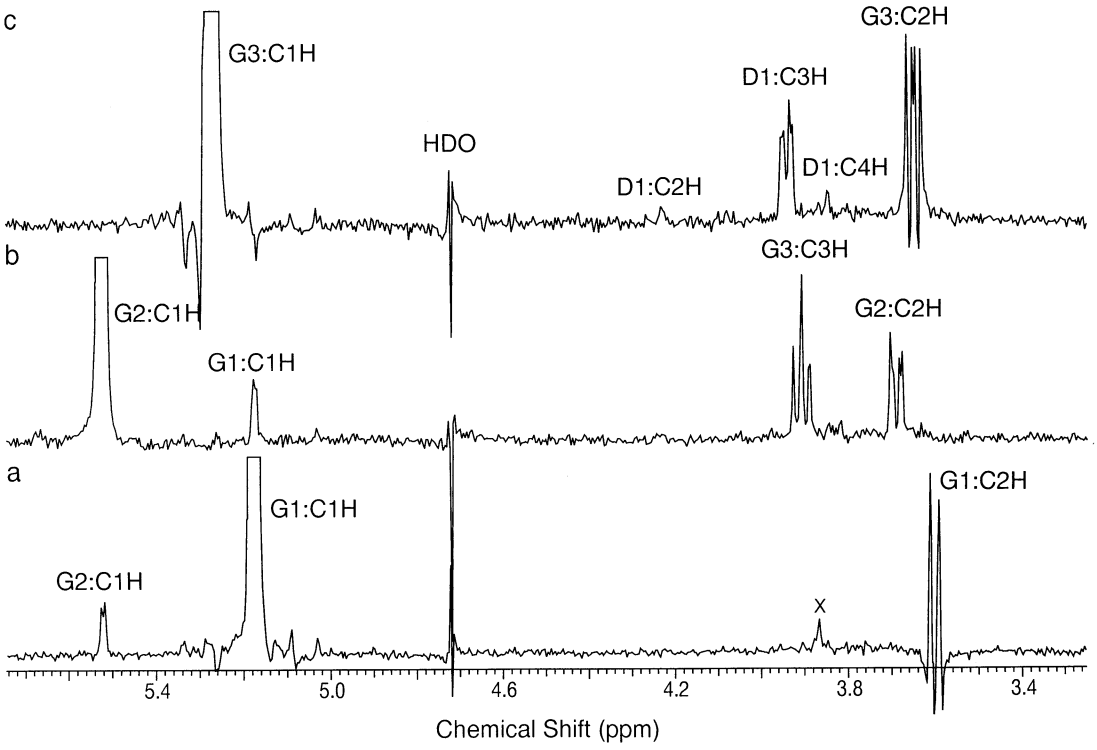

Fig. 3. Selected traces, parallel to F2, through the 200 ms mixing time NOESY spectrum of $Glc_3Man_7GlcNAc_2$ in ^2H_2O, pH = 8.05, T = 30°C, showing the cross-linkage NOEs (data summarized in Table II). (**a**) Trace through the G1:C1H resonance. (**b**) Trace through the G2:C1H resonance. (**c**) Trace through the G3:C1H resonance. The spectrum is phased with a positive diagonal (negative NOEs are shown as positive). X, spectral artefact peak.

Table I. Resonance assignments for residues on the 1,3 arm in $Glc_3Man_7GlcNAc_2$ (pH = 8.05), $Glc_2Man_7GlcNAc_2$ (pH = 7.85), $Glc_1Man_7GlcNAc_2$ (pH = 8.60) and $Glc_3Man_4GlcNAc_1$ (pH = 8.44) in ^2H_2O, T = 30°C, referenced to internal acetone at 2.217 p.p.m.

Residue	C1H	C2H	C3H	C4H	C5H	C6H	C6H′
$G_3M_7N_2$							
4	5.336	4.088	3.992	*3.648*			
C	5.292	4.105	3.954	*3.692*			
D1	5.032	4.232	3.946	*3.846*			
G3	5.266	3.650	*3.910*	{3.622}	{3.852}	{3.752}	
G2	5.524	3.692	*3.840*	3.510	4.060	3.810	
G1	5.174	3.604	*3.786*	3.454	3.964	{3.842}	
$G_2M_7N_2$							
4	5.336	4.091	3.992	3.666			
C	5.295	4.106	3.947	3.705			
D1	5.034	4.233	3.941	3.840			
G3	5.262	3.655	3.908	3.621	3.852	3.756	
G2	5.351	3.553	3.745	3.445	4.003		
$G_1M_7N_2$							
4	5.336	4.088	3.993	3.679			
C	5.295	4.108	3.948	3.730			
D1	5.031	4.231	3.935	3.842			
G3	5.248	3.550	3.772	3.394	3.830		
$G_3M_4N_1$							
4	5.337	4.088	3.990	3.680			
C	5.287	4.106	3.948	3.716			
D1	5.035	4.230	3.941	3.856			
G3	5.267	3.653	3.910	3.624	3.849	3.755	
G2	5.524	3.695	3.830	3.510	4.060	3.810	
G1	5.174	3.609	3.787	3.452	3.964	3.841	

For the $Glc_3Man_7GlcNAc_2$ assignments, *italics* indicate peaks assigned via RELAY spectra (cannot be resolved in COSY) and brackets {...} indicate peaks assigned to a spin-system via TOCSY spectra. All other assignments were made from TOCSY spectra by comparison with $Glc_3Man_7GlcNAc_2$.

Table II. NOE results for the 1,3 arm linkages in $Glc_3Man_7GlcNAc_2$ and the derived distance constraints used to generate the torsion angle maps in Figure 4

Proton pair	Initial NOE build-up rate		Distance constraint (Å)
G1–G2 : Glcα1→2Glcα linkage			
G1:C1H	G1:C2H	1.19	2.30 (calibration)
	G2:C1H	0.66	2.40–3.00
	G2:C2H	not observed	⩾3.5
	G2:C3H	not observed	⩾3.5
G1:C5H	G2:C1H	not observed	⩾3.5
	G2:C2H	obscured	–
G2–G3 : Glcα1→3Glcα linkage			
G2:C1H	G2:C2H	2.66	2.30 (calibration)
	G3:C2H	not observed	⩾3.5
	G3:C3H	3.10	2.10–2.30
	G3:C4H	not observed	⩾3.5
G2:C5H	G3:C2H	not observed	⩾ 3.5
	G3:C3H	obscured	–
	G3:C4H	not observed	⩾3.5
G3–D1 : Glcα1→3Manα linkage			
G3:C1H	G3:C2H	3.67	2.30 (calibration)
	D1:C2H	not observed	⩾3.5
	D1:C3H	2.86	2.33–2.50
	D1:C4H	not observed	⩾3.5
G3:C5H	D1:C2H	obscured	–
	D1:C3H	obscured	–
	D1:C4H	obscured	–
D1–C : Manα1→2Manα linkage			
D1:C1H	D1:C2H	1.67	2.50 (calibration)
	C:C1H	0.63	2.85–3.15
	C:C2H	3.56	2.15–2.30
	C:C3H	0.47	2.90–3.40
	C:C4H	not observed	⩾3.5
D1:C2H	C:C2H	0.39	3.1–3.6
D1:C5H	C:C1H	obscured	–
C–4 : Manα1→2Manα linkage			
C:C1H	C:C2H	medium (overlapping)	2.50 (calibration)
	4:C1H	obscured	–
	4:C2H	strong (overlapping)	<2.4
	4:C3H	very weak	>2.8
C:C5H	4:C1H	obscured	–

Not observed indicates that the peak is not seen at short mixing times (<200 ms); some of these peaks are observed at longer mixing times (in the spin-diffusion regime).

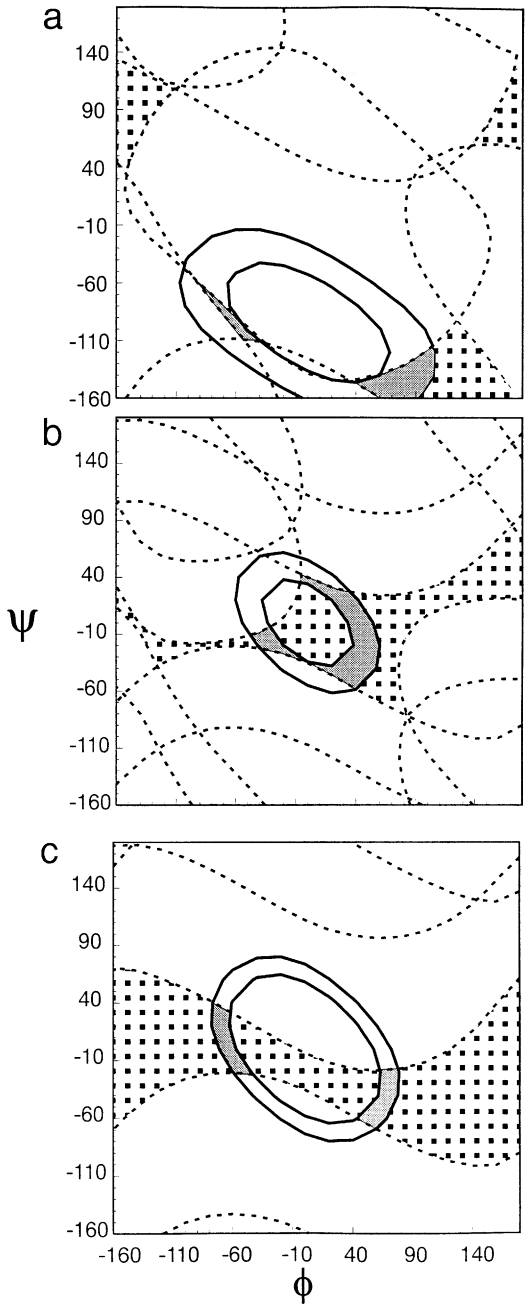

Fig. 4. Torsion angle maps for (**a**) Glcα1→2Glcα (**b**) Glcα1→3Glcα and (**c**) Glcα1→3Manα glycosidic linkages. The solid lines give the regions of conformational space consistent with the distance constraints from the observed NOEs and the dashed lines the constraints from the absent NOEs (see Table II). The grey areas give the regions consistent with all the NOE data. The unshaded areas are those which cannot be significantly populated, due to the absence of characteristic NOEs (Wormald and Edge, 1993).

and in each case this allows large regions of conformational space to be excluded as not significantly populated (Wormald and Edge, 1993). Figure 4 shows the torsion angle maps, derived from the NOE distance, for the three glycosidic linkages in the Glc_3Man structural unit. In addition, further areas of conformational space can be excluded on the basis of unfavourable steric interactions predicted by molecular modelling.

G1–G2 (Glcα1→2Glcα) linkage. Figure 4a shows two regions of conformational space consistent with all the available NOE data, at ($\phi = -60 \pm 20$, $\psi = -90 \pm 20$) and ($\phi = +80 \pm 30$, $\psi = -145 \pm 30$). The former region is consistent with the known conformational preference of an α-linkage for $\phi \top -60$, due to the exo-anomeric effect (Woods *et al.*, 1995). The latter region can be excluded because the resulting structure involves severe steric interactions between G1:O2 and G2:O1.

G2–G3 (Glcα1→3Glcα) linkage. Figure 4b shows two regions of conformational space consistent with all the available NOE data, at ($\phi = -35 \pm 15$, $\psi = -15 \pm 15$) and a large region around ($\phi = +50$, $\psi = -10$). The region around ($\phi = +40 \pm 20$, $\psi = -30 \pm 20$) can be excluded because of steric interactions between G2:O2 and G3:O4; however, the region at ($\phi = +20 \pm 25$, $\psi = +10 \pm 20$) is feasible on steric grounds. The region at ($\phi = -35 \pm 15$, $\psi = -15 \pm 15$) is consistent with the known conformational preference of an α-linkage for ϕ

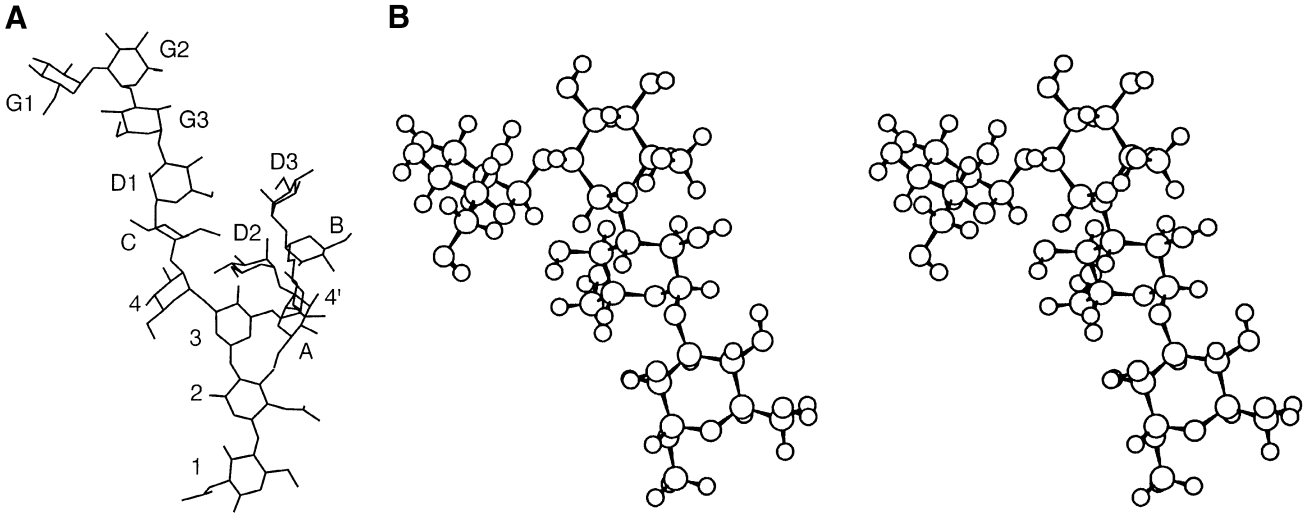

Fig. 5. (**A**) Molecular model of $Glc_3Man_9GlcNAc_2$, based on the NMR structure for the Glc_3Man unit and the NMR/molecular dynamics structure for $Man_9GlcNAc_2$ (R.J.Woods, A.Pathiaseril, M.R.Wormald, C.J.Edge and R.A.Dwek, submitted). (**B**) Stereo-image of the structure of the Glc_3Man unit from $Glc_3Man_7GlcNAc_2$, fully consistent with the NMR, X-ray crystallographic and molecular modelling data.

$\top$ –60. The disaccharide Glcα1→3Glcα-OMe has been crystallized (Neuman *et al.*, 1980) and this gives torsion angles of (ϕ = –18.2, ψ = –14.9).

G3–D1 (Glcα1→3Manα) linkage. Figure 4c shows two regions of conformational space consistent with all the available NOE data, at (ϕ = –65 ± 15, ψ = +10 ± 30) and (ϕ = +65 ± 20, ψ = –45 ± 25). The former region is consistent with the known conformational preference of an α-linkage for ϕ $\top$ –60. The latter region can be excluded because of steric interactions between G3:O2 and D1:O4.

Thus, for each linkage there is a single, sterically allowed conformation consistent with all the available NMR data, X-ray data and the known linkage conformational preferences. This is in contrast to the Man–Man linkages in $Man_9GlcNAc_2$ (R.J.Woods, A.Pathiaseril, M.R.Wormald, C.J.Edge and R.A.Dwek, submitted) where the NMR data is not consistent with a single conformation for each linkage and molecular dynamics simulations show the linkages to be either flexible or adopt more than one conformation. This leads to the structure for the Glc_3Man unit shown in Figure 5b. A molecular model of $Glc_3Man_9GlcNAc_2$ is shown in Figure 5a, based on the NMR structure of Glc_3Man and the NMR/molecular dynamics structure of $Man_9GlcNAc_2$.

The linkage analysis for the Glc_3Man unit in the linear saccharide $Glc_3Man_4GlcNAc_1$ in ^2H_2O gave essentially identical results (data not shown). There are no significant changes in chemical shifts for residues G1, G2, G3 and D1 between the two species (Table I). This indicates that residues 4′, A and B do not affect the structure of the Glc_3Man unit.

Linkage conformation analysis of the Glc_2Man and Glc_1Man units

An identical analysis was also carried out on the species $Glc_2Man_7GlcNAc_2$ and $Glc_1Man_7GlcNAc_2$, prepared by digestion of the parent compound with α-glucosidase I and α-glucosidase II respectively. The resonance assignments for these oligosaccharides are given in Table I.

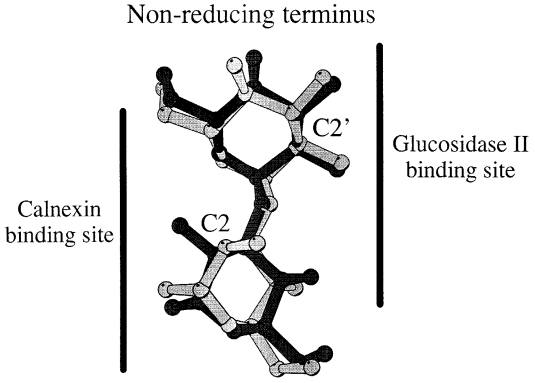

Fig. 6. Overlay of the Glcα1→3Glcα (dark grey) and Glcα1→3Manα (light grey) linkage structures derived from the NMR data for $Glc_3Man_7GlcNAc_2$. The points of epimerization for mannose relative to glucose are labelled on the non-reducing terminal (C2′) and the reducing terminal (C2) residues. The proposed binding sites for α-glucosidase II and calnexin, based on their known specificities, are indicated (see text for details). In the oligosaccharide structure, these linkages are rotated 180° with respect to each other (Figure 7b and c).

Removal of the G1 residue affects the chemical shifts of the G2 resonances, but not those of G3, D1 or C. Similarly, removal of G2 does not affect the resonances of D1 or C. The NOE analyses (data not shown) produced results both qualitatively and quantitatively very similar to those obtained from $Glc_3Man_7GlcNAc_2$ for all the remaining linkages. Thus, sequential cleavage of the Glcα1→2Glcα and Glcα1→3Glcα linkages does not alter the conformation of any of the other linkages.

Effects of the Glc_3 unit on the 1,3 arm oligomannose structure

The NOEs across the D1–C Manα1→2Manα linkage and the C–4 Manα1→3Manα linkage were also measured. The data for these linkages are given in Table II. Spectral overlap makes quantitation of the C–4 linkage NOEs very difficult (as is also observed for $Man_9GlcNAc_2$). The observed pattern of NOEs is very similar to that found for these linkages in $Man_9GlcNAc_2$ (R.J.Woods, A.Pathiaseril,

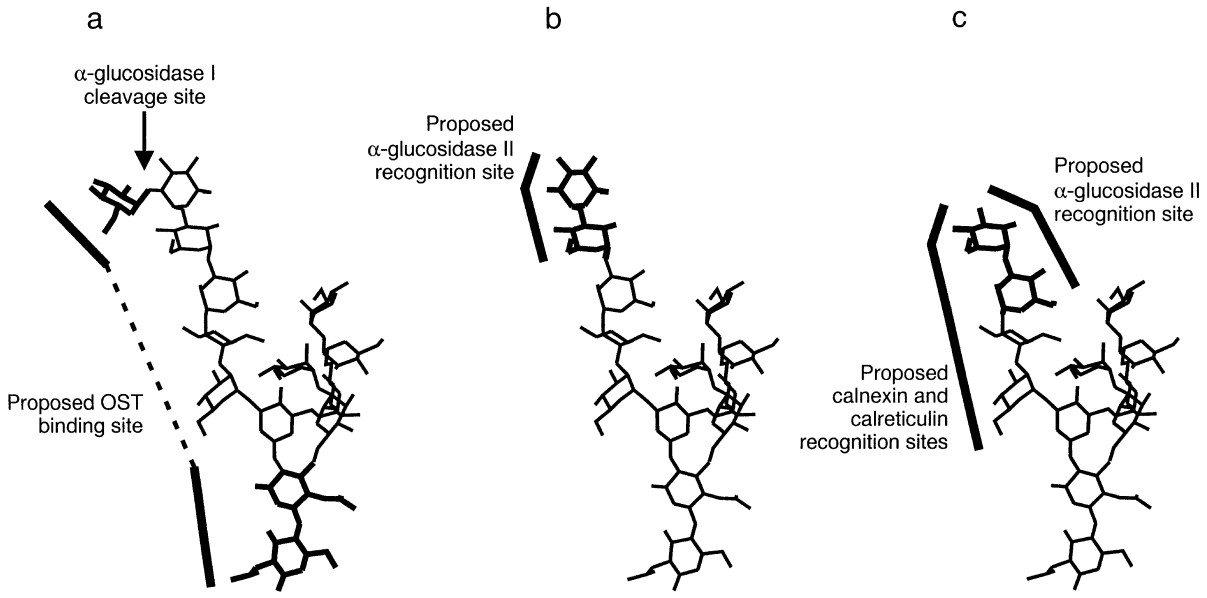

Fig. 7. The proposed recognition sites of (**a**) OST and α-glucosidase I for $Glc_3Man_9GlcNAc_2$, (**b**) α-glucosidase II for $Glc_2Man_9GlcNAc_2$ and (**c**) calnexin, calreticulin and α-glucosidase II for $Glc_1Man_9GlcNAc_2$.

M.R.Wormald, C.J.Edge and R.A.Dwek, submitted), indicating that the addition of the Glc_3 unit does not significantly alter the structure of the 1,3 arm to which it is attached. Again, this is consistent with the similarities in chemical shift of the resonances from residues C and 4 in the two species. It is interesting to note that the change in chemical shift of the D1:C2H resonance on addition of residue G3 allows an extra weak NOE between D1:C2H and C:C2H to be observed. In $Man_9GlcNAc_2$ this peak would lie too close to the diagonal to be resolved.

Discussion

We previously have reported on the solution conformation of the $Man_9GlcNAc_2$ oligosaccharide (R.J.Woods, A.Pathiaseril, M.R.Wormald, C.J.Edge and R.A.Dwek, submitted). Comparison of this structure with the NMR data on $Glc_3Man_7GlcNAc_2$ show the 1,3 arm to be unaffected by the presence of the terminal glucose residues. Furthermore, removal of the 1,6 arm mannose residues (4′, A and B) from $Glc_3Man_7GlcNAc_2$ does not alter the conformation of the triglucosyl cap. Thus, the oligomannose core and the glucosylated cap of $Glc_3Man_7GlcNAc_2$ can be considered as independent structural units. The structure of $Glc_3Man_9GlcNAc_2$ obtained by combining these data (Figure 5a) shows the Glc_3 cap to be remote from the rest of the molecule, with no interactions between the glucosyl cap and any of the mannose residues, except D1.

The structure of the triglucosyl cap (Figure 5b) is quite compact, forming a tight turn. Molecular modelling shows that the distance between G1:C1H and G3:O4 is 1.6 Å, indicating a potential hydrogen bond which would stabilize this structure. The Glc_3Man unit must also be quite rigid (particularly the Glcα1→2Glcα linkage), since the permitted areas of conformational space for the linkages, based on the negative NOE constraints (Figure 4), are small.

The structure presented here for the Glc_3Man unit in water is similar to that reported for the synthetic Glc_3Man tetrasaccharide in dimethylsulfoxide (DMSO) (Alvarado *et al.*, 1991) except that the authors report a hydrogen bond between G1:OH2 and G3:OH4 for the latter. This hydrogen bond gives rise to a conformation for the Glcα1→2Glcα linkage that would lead to a strong G1:C1H to G2:C2H NOE and no G1:C1H to G2:C1H NOE, in contrast to the data presented here (see Figure 3 and Table II). This difference may be due to different hydrogen bonding patterns being favoured by different solvents. Large oligosaccharides, such as $Glc_3Man_7GlcNAc_2$, are usually insoluble in DMSO, and so the smaller $Glc_3Man_4GlcNAc_1$ was prepared. However, this species is still not sufficiently soluble in DMSO to make conformational analysis possible.

The NMR data show that the triglucosyl cap adopts a single conformation, which is situated at the end of a relatively rigid and extended 1,3 arm (R.J.Woods, A.Pathiaseril, M.R.Wormald, C.J.Edge and R.A.Dwek, submitted). This results in an accessible triglucosyl cap, ~30 Å from the linkage point to the protein (Figure 5). The OST complex recognizes both the Glc_3 cap and the chitobiose core during transfer of the oligosaccharide from the dolichol intermediate to the protein. *In vitro* experiments using purified OST have shown that the dolichol-linked chitobiose core alone is a sufficient substrate for OST (Breuer and Bause, 1995). The Glc_3 residues contribute to the transferase–dolichol intermediate binding reaction by producing a 10-fold increase in apparent affinity constants (Breuer and Bause, 1995). Thus, the OST must have a secondary lectin site involved in recognition of the triglucosyl cap ~30 Å from the active site (Figure 7a). The presence of a single conformation for the triglucosyl cap is consistent with its primary role as a recognition element.

Almost immediately after *en bloc* transfer of the oligosaccharide, the terminal glucose residue is hydrolysed by α-glucosidase I, a type II transmembrane protein with a lumenal C-terminal catalytic domain (Kalz-Fuller *et al.*,

1995). The structure places this linkage in a sterically unhindered position accessible for hydrolysis (Figure 7a). This step may be involved in rapid uncoupling of the OST–donor oligosaccharide–peptide complex.

α-Glucosidase II, an enzyme that is only loosely associated with the ER membrane (Brada and Dubach, 1984), removes the next two α1→3-linked glucose residues. Little is known about the structure of glucosidase II (Trombetta *et al.*, 1996). Kinetic analysis (Alonso *et al.*, 1991) has suggested that the protein has two binding sites which are close together and differ in the value of their K_ms by a factor of 600. The authors suggest that the two glucose residues of the $Glc_2Man_9GlcNAc_2$–glycoprotein could bind to both these sites simultaneously, leading to cleavage of both linkages before release of $Man_9GlcNAc_2$–glycoprotein. This would not allow recognition of the newly synthesized glycoprotein by the chaperones calnexin and calreticulin to take place until after subsequent reglucosylation by the UDP-Glc:glycoprotein glucosyltransferase (Sousa *et al.*, 1992). In trypanosomatid protozoa, this is the only route available for calnexin/calreticulin recognition (Parodi, 1993). However, the fact that large quantities of the mono-glucosylated glycan can be produced by partial hydrolysis of $Glc_2Man_7GlcNAc_2$ with catalytic amounts of α-glucosidase II demonstrates that, for this substrate, glycan release must occur after removal of the first α1→3-linked glucose residue.

The conformations of the Glcα1→3Glcα and Glcα1→3Manα linkages hydrolysed by α-glucosidase II are very similar, as shown by their superposition in Figure 6. The only major structural difference is in the epimerization at C2 of the inner residue. The two linkages have a common structural epitope, extending from C6 of the inner residue to C3 of the outer residue. This could be a potential epitope for α-glucosidase II recognition, requiring only a single active site on the enzyme for hydrolysis of both linkages. α-Glucosidase II does not hydrolyse the Manα1→3Man linkage. This linkage only differs from the Glcα1→3Man linkage by epimerization at C2 of the outer residue (labelled C2′ in Figure 6) which is within this proposed recognition epitope.

The proposed Glcα1→3Glc and Glcα1→3Man linkage epitopes are on opposite sides of the glucosylated oligomannose structure (Figure 7b and c). The implication of this is that release and reorientation of the glycoprotein would be required before hydrolysis of the second linkage, thus allowing calnexin/calreticulin recognition. The use of a single enzyme to cleave both these linkages in a process involving release, reorientation and rebinding would also lead to a longer lifetime for the mono-glucosylated glycans.

In contrast, calnexin and calreticulin specifically recognize the Glcα1→3Man linkage, not Glcα1→3Glc, indicating that C2 of the mannose residue is involved in the oligosaccharide–protein interactions (Figure 6). More recent studies have shown the Glcα1→3Manα1→2Manα-1→2Man tetrasaccharide to be a 100-fold more potent competitive inhibitor of calnexin than the Glcα1→3Man disaccharide (A.Vassilakos, M.Michalak, M.A.Lehrman and D.B.Williams, submitted), suggesting that calnexin recognizes the whole 1,3 arm. These four residues, including the proposed recognition epitope in Figure 6, form a continuous molecular surface on $Glc_1Man_9GlcNAc_2$ (Figure 7c).

On the basis of this model, we propose that calreticulin/calnexin and glucosidase II approach their common substrate from different sides of the molecule (Figure 7c). This would allow glucosidase II to hydrolyse the G3–D1 linkage whilst the glycoprotein is bound to calreticulin/calnexin and thus to promote dissociation of the complex. This is consistent with the recent observation that glucosidase II causes dissociation of the complexes formed between mono-glucosylated ribonuclease B and calreticulin or calnexin (Rodan *et al.*, 1996).

Both calnexin (A.Vassilakos, M.Michalak, M.A. Lehrman and D.B.Williams, submitted) and calreticulin (Spiro *et al.*, 1996) also require the presence of the 4′ mannose residue (see Figure 1 for notation) for maximum binding. In $Glc_xMan_9GlcNAc_2$, this residue is protected from recognition by the oligomannose 1,6 arms (Figure 5). A single continuous molecular surface cannot be formed which involves both the 1,3 arm residues and the 4′ mannose. One possible explanation for this is that the presence of the 4′ residue alters the conformation and/or flexibility of the 1,3 arm, thus indirectly affecting the chaperone binding. Molecular modelling of the $Man_9GlcNAc_2$ structure does suggest the presence of a water-mediated hydrogen bond between residues 4 and 4′ (R.J.Woods, personal communication). We are currently investigating the solution structures of the glucosylated $Man_5GlcNAc_2$ and $Man_4GlcNAc_2$ saccharides.

Materials and methods

Reagents

NB-DNJ was a gift of G.D.Searle & Co. ConA–Sepharose was obtained from Sigma; Chelex 100, Dowex AG50 X-12 (H^+ form), Dowex AG3 X-A4 (OH^- form) and Bio-Gel P4 (–400 mesh) were purchased from Bio-Rad Laboratories Ltd. QAE-Sephadex A25 was from Pharmacia Ltd. α-Glucosidase I and α-glucosidase II were prepared as previously described (Karlsson *et al.*, 1993); EndoH was purchased from New England Biolabs; Jack bean α-mannosidase was prepared as previously reported (Li and Li, 1972).

Cell culture

CHO cells, transfected with pEE6HCMVgp120GS, secreting recombinant HIV-IIIB gp120 were obtained from Dr P.Stevens (MRC AIDS Directed Programme Research Project). Cells were cultured in Glasgow MEM culture medium (Gibco Ltd, Uxbridge, UK) supplemented with 10% fetal calf serum (FCS, Techgen), 50 U/ml penicillin and 50 μg/ml streptomycin (Gibco) in the presence of 3 mM NB-DNJ and maintained at 37°C with 5% CO_2 for 1 week.

Sugar release

Glycoproteins containing oligomannose glycan structures were separated from the cell culture medium by affinity chromatography on ConA–Sepharose. The material obtained was lyophilized and glycan release achieved by large scale manual hydrazinolysis (Wing *et al.*, 1992; Patel *et al.*, 1993). The sugars were desalted on a four resin column consisting of 200–300 μl each of Chelex 100, Dowex AG50 X-12 (H^+ form), Dowex AG3 X4A (OH^- form) and QAE-Sephadex.

Gel permeation chromatography

Hydrazinolysis released sugars were first subjected to high resolution gel permeation chromatography using two Bio-Gel P4 (1.5×100 cm) columns in series. Columns were maintained at 55°C and eluted with distilled water. Fractions containing hyperglucosylated sugars ($Glc_3Man_{7-9}GlcNAc_2$) were pooled and lyophilized.

Normal phase HPLC

Co-eluting sugars from the P4 chromatography step were separated further by normal phase HPLC on a Waters HPLC system with a

GlycoSep-N column (size 4.6×250 mm, from Oxford GlycoSciences) using the low salt solvent system previously described (Guile *et al.*, 1996). Briefly, solvent A was 50 mM ammonium formate at pH 4.4 and solvent B was acetonitrile. A linear gradient of 35–53% A over 100 min at a flow rate of 0.4 ml/min was followed by an increase to 100% A over 3 min and a 17 min wash, before re-equilibration in 35% A. The same conditions were used in both analytical and preparative runs. For analytical runs, the sugars were fluorescently labelled with 2-AB (Bigge *et al.*, 1995) using a Signal™ labelling kit from Oxford Glycosciences and monitored using a Jasco FP-920 fluorescence detector. For preparative runs, quantities of oligosaccharide injected were 6–7 orders of magnitude larger than those normally used for the analytical runs and the sugars were not fluorescently labelled. Fractions were analysed subsequently using matrix-assisted laser desorption/ionization.

Mass spectrometry

Matrix-assisted laser desorption/ionization mass spectrometry was performed on a time-of-flight Finnigan LaserMat instrument. Samples were prepared using a 2,5-dihydroxybenzoic acid matrix as previously described (Harvey *et al.*, 1994).

Preparation of $Glc_2Man_7GlcNAc_2$ and $Glc_1Man_7GlcNAc_2$

$Glc_2Man_7GlcNAc_2$ was obtained from $Glc_3Man_7GlcNAc_2$ by digestion with α-glucosidase I at 37°C until completion. The processed oligosaccharide was separated from the enzyme and free glucose by Bio-Gel P4 permeation chromatography, as described above. $Glc_1Man_7GlcNAc_2$ was prepared from $Glc_2Man_7GlcNAc_2$ by partial digestion with α-glucosidase II. A digestion time of 4 h 30 min was found to give optimal yield of $Glc_1Man_7GlcNAc_2$. The resulting mixture of $Glc_2Man_7GlcNAc_2$, $Glc_1Man_7GlcNAc_2$ and $Man_7GlcNAc_2$ species was separated using Bio-Gel P4 permeation chromatography, as described above. The $Glc_2Man_7GlcNAc_2$ fraction obtained in this first run was subjected to a second, identical run. The $Glc_1Man_7GlcNAc_2$ fractions from the two runs were pooled and used for NMR studies.

Preparation of $Glc_3Man_4GlcNAc_1$

$Glc_3Man_4GlcNAc_1$ was prepared by enzymatic digestion of $Glc_3Man_7GlcNAc_2$ with a mixture of EndoH (2 U) and Jack bean α-mannosidase (6 U) for 72 h at 37°C. $Glc_3Man_4GlcNAc_1$ subsequently was purified by gel permeation chromatography on Bio-Gel P4, as described above.

NMR spectroscopy

All NMR experiments were performed on a Varian UNITY 500 spectrometer, at a probe temperature of 30°C. For $Glc_3Man_7GlcNAc_2$, ^1H resonance assignments were obtained from two-dimensional phase-sensitive COSY, RELAY and TOCSY spectra. For all other samples, resonance assignments were obtained from TOCSY spectra with 80 ms mixing time. Spectra were multiplied by unshifted sine- or cosine-bell functions in both dimensions, as appropriate. Sequence-specific assignments were made by comparison with reported assignments for oligomannose-type oligosaccharides (Vliegenthart *et al.*, 1983; Wormald *et al.*, 1991) and analysis of the cross-linkage NOE patterns. 2D NOESY spectra were recorded at mixing times of 100, 200 and 400 ms without any random variation in mixing time. Absolute peak volumes were measured from the phase-sensitive data set. Any contributions to the cross-peaks from scalar coupling are anti-phase and thus, whilst distorting the peak shape, will not contribute to the peak volume. Spectral noise was estimated by measuring the volume integrals of regions of the baseline around the cross-peaks. The NOE build-up curves for specific proton pairs were obtained and fitted using a second order polynomial function to give the initial linear build-up rates (Table I). Inter-proton distance constraints were obtained from these initial rates using the two-spin approximation. The intra-residue glucose C1H–C2H distance [obtained from the crystal structure of Glcα1→3Glcα-OMe (Neuman *et al.*, 1980)] was used as the internal calibration for calculating distances from the NOE build-up rates.

Molecular modelling

Torsion angle maps were generated as previously described (Wooten *et al.*, 1990), making use of distance constraints from both the presence and absence of NOEs (Wormald and Edge, 1993). Standard torsion angle nomenclature was used throughout, i.e. for a 1-*x* linkage, ϕ = H1-C1-O-Cx' and ψ = C1-O-Cx'-Hx'. Disaccharide X-ray crystal structures were obtained from searching the Cambridge Crystallographic Database (Allen and Kennard, 1993) at the Chemical Database Service at Daresbury (Fletcher *et al.*, 1996). Molecular modelling was performed on a Silicon Graphics 4D/35 workstation using the programs INSIGHT II and DISCOVER (Biosym Tech. Inc.).

Acknowledgements

We would like to thank Brian Matthews for performing the large scale hydrazinolysis and Geoffrey Guile for help with the normal-phase HPLC system. A.J.P. is supported by a visiting Research Fellowship from The Royal Society. S.P. is a Collaborative Biomedical Research Fellow supported by The Wellcome Trust. F.M.P. is a Lister Institute Research Fellow. T.D.B. and F.M.P. are supported by Searle/Monsanto. We wish to acknowledge the use of the EPSRC's Chemical Database Service at Daresbury.

References

Allen,F.H. and Kennard,O. (1993) 3D search and research using the Cambridge Structural Database. *Chemical Design Automation News*, **8**, 1, 31–37.

Alonso,J.M., Santa Cecilia,A. and Calvo,P. (1991) Glucosidase II from rat liver microsomes. Kinetic model for binding and hydrolysis. *Biochem. J.*, **278**, 721–727.

Alvarado,E., Nukada,T., Ogawa,T. and Ballou,C.E. (1991) Conformation of the glucotriose unit in the lipid-linked oligosaccharide precursor for protein glycosylation. *Biochemistry*, **30**, 881–886.

Arunachalam,B. and Cresswell,P. (1995) Molecular requirements for the interaction of class II major histocompatibility complex molecules and invariant chain with calnexin. *J. Biol. Chem.*, **270**, 2784–2790.

Bause,E., Breuer,W., Schweden,J., Roeser,R. and Geyer,R. (1992) Effect of substrate structure on the activity of Man9-mannosidase from pig liver involved in *N*-linked oligosaccharide processing. *Eur. J. Biochem.*, **208**, 451–457.

Bergeron,J.J., Brenner,M.B., Thomas,D.Y. and Williams,D.B. (1994) Calnexin: a membrane-bound chaperone of the endoplasmic reticulum. *Trends Biochem. Sci.*, **19**, 124–128.

Bigge,J.C., Patel,T.P., Bruce,J.A., Goulding,P.N., Charles,S.M. and Parekh,R.B. (1995) Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal. Biochem.*, **230**, 229–238.

Brada,D. and Dubach,U.C. (1984) Isolation of a homogeneous glucosidase II from pig kidney microsomes. *Eur. J. Biochem.*, **141**, 149–156.

Breuer,W. and Bause,E. (1995) Oligosaccharyl transferase is a constitutive component of an oligomeric protein complex from pig liver endoplasmic reticulum. *Eur. J. Biochem.*, **228**, 689–696.

Dwek,R.A. (1996) Glycobiology: toward understanding the function of sugars. *Chem. Rev.*, **96**, 683–720.

Fischer,P.B., Karlsson,G.B., Butters,T.D., Dwek,R.A. and Platt,F.M. (1996) *N*-butyldeoxynojirimycin-mediated inhibition of human immunodeficiency virus entry correlates with changes in antibody recognition of the V1/V2 region of gp120. *J. Virol.*, **70**, 7143–7152.

Fletcher,D.A., McMeeking,R.F. and Parkin,D. (1996) The United Kingdom Chemical Database service. *J. Chem. Inf. Comput. Sci.*, **36**, 746–749.

Guile,G.R., Rudd,P.M., Wing,D.R., Prime,S.B. and Dwek,R.A. (1996) A rapid high-resolution high-performance liquid chromatographic method for separating glycan mixtures and analysing oligosaccharide profiles. *Anal. Biochem.*, **240**, 210–226.

Harvey,D.J., Rudd,P.M., Bateman,R.H., Bordoli,R.S., Howes,K., Hoyes,J.B. and Vickers,R.G. (1994) Examination of complex oligosaccharides by matrix-assisted laser desorption mass spectrometry on time-of-flight and magnetic sector instruments. *Org. Mass Spectrom.*, **29**, 753–766.

Hebert,D.N., Foellmer,B. and Helenius,A. (1995) Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell*, **81**, 425–433.

Helenius,A. (1994) How *N*-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell*, **5**, 253–265.

Kalz-Fuller,B., Bieberich,E. and Bause,E. (1995) Cloning and expression of glucosidase I from human hippocampus. *Eur. J. Biochem.*, **231**, 344–351.

Karlsson,G.B., Butters,T.D., Dwek,R.A. and Platt,F.M. (1993) Effects of the imino sugar *N*-butyldeoxynojirimycin on the *N*-glycosylation of recombinant gp120. *J. Biol. Chem.*, **268**, 570–576.

Kornfeld,S., Gregory,W. and Chapman,A. (1979) Class E Thy-1 negative mouse lymphoma cells utilize an alternate pathway of oligosaccharide processing to synthesize complex-type oligosaccharides. *J. Biol. Chem.*, **254**, 11649–11654.

Li,Y.-T. and Li,S.-C. (1972) α-Mannosidase, β-*N*-acetylhexosaminidase and β-galactosidase from Jack Bean meal. *Methods Enzymol.*, **28**, 702–713.

Mehta,A., Lu,X., Block,T.M., Blumberg,B.S. and Dwek,R.A. (1997) Hepatitis B virus (HBV) envelope glycoproteins vary drastically in their sensitivity to glycan processing: evidence that alteration of a single *N*-linked glycosylation site can regulate HBV secretion. *Proc. Natl Acad. Sci. USA*, **94**, 1822–1827.

Neuman,A., Avenel,D., Arene,F., Gillier-Pandraud,H., Pougny,J.-R. and Sinay,P. (1980) Structure cristalline du methyl-3-*O*-alpha-D-glucopyranosyl-alpha-D-glucopyranoside (methyl-alpha-nigeroside). *Carbohydr. Res.*, **80**, 15–24.

Ou,W.J., Cameron,P.H., Thomas,D.Y. and Bergeron,J.J. (1993) Association of folding intermediates of glycoproteins with calnexin during protein maturation. *Nature*, **364**, 771–776.

Parodi,A.J. (1993) *N*-glycosylation in trypanosomatid protozoa. *Glycobiology*, **3**, 193–199.

Patel,T., Bruce,J., Merry,A., Bigge,C., Wormald,M., Jacques,A. and Parekh,R. (1993) Use of hydrazine to release in intact and unreduced form both *N*- and *O*-linked oligosaccharides from glycoproteins. *Biochemistry*, **32**, 679–693.

Petrescu,S., Petrescu,A.J., Dwek,R.A. and Platt,F.M. (1997) Inhibition of *N*-glycan processing in B16 melanoma cells results in inactivation of tyrosinase but does not prevent its transport to the melanosome. *J. Biol. Chem.*, in press.

Rodan,A.R., Simons,J.F., Trombetta,E.S. and Helenius,A. (1996) *N*-linked oligosaccharides are necessary and sufficient for association of glycosylated forms of bovine RNase with calnexin and calreticulin. *EMBO J.*, **15**, 6921–6930.

Sousa,M.C., Ferrero Garcia,M.A. and Parodi,A.J. (1992) Recognition of the oligosaccharide and protein moieties of glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase. *Biochemistry*, **31**, 97–105.

Spiro,R.G., Zhu,Q., Bhoyroo,V. and Soling,H.-D. (1996) Definition of the lectin-like properties of the molecular chaperone, calreticulin, and demonstration of its copurification with endomannosidase from rat liver Golgi. *J. Biol. Chem.*, **271**, 11588–11594.

Trombetta,E.S., Simons,J.F. and Helenius,A. (1996) Endoplasmic reticulum glucosidase II is composed of a catalytic subunit, conserved from yeast to mammals, and a tightly bound noncatalytic HDEL-containing subunit. *J. Biol. Chem.*, **271**, 27509–27516.

Verostek,M.F., Atkinson,P.H. and Trimble,R.B. (1993) Glycoprotein biosynthesis in the *alg3 Saccharomyces cerevisiae* mutant. I. Role of glucose in the initial glycosylation of invertase in the endoplasmic reticulum. *J. Biol. Chem.*, **268**, 12095–12103.

Vliegenthart,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.

Ware,F.E., Vassilakos,A., Peterson,P.A., Jackson,M.R., Lehrman,M.A. and Williams,D.B. (1995) The molecular chaperone calnexin binds Glc1Man9GlcNAc2 oligosaccharide as an initial step in recognizing unfolded glycoproteins. *J. Biol. Chem.*, **270**, 4697–46704.

Wing,D.R., Rademacher,T.W., Field,M.C., Dwek,R.A., Schmitz,B., Thor,G. and Schachner,M. (1992) Use of large-scale hydrazinolysis in the preparation of *N*-linked oligosaccharide libraries: application to brain tissue. *Glycoconjugate J.*, **9**, 293–301.

Woods,R.J., Dwek,R.A., Edge,C.J. and Fraser-Reid,B. (1995) Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. GLYCAM_93 parameter development. *J. Phys. Chem.*, **99**, 3832–3846.

Wooten,E.W., Edge,C.J., Bazzo,R., Dwek,R.A. and Rademacher,T.W. (1990) Uncertainties in structural determination of oligosaccharide conformation using measurements of nuclear Overhauser effects. *Carbohydr. Res.*, **203**, 13–17.

Wormald,M.R. and Edge,C.J. (1993) The systematic use of negative nuclear Overhauser constraints in the determination of oligosaccharide conformations: application to sialyl-Lewis X. *Carbohydr. Res.*, **246**, 337–344.

Wormald,M.R., Wooten,E.W., Bazzo,R., Edge,C.J., Feinstein,A., Rademacher,T.W. and Dwek,R.A. (1991) The conformational effects of *N*-glycosylation on the tailpiece from serum IgM. *Eur. J. Biochem.*, **198**, 131–139.

Zhang,Q., Tector,M. and Salter,R.D. (1995) Calnexin recognizes carbohydrate and protein determinants of class I major histocompatibility complex molecules. *J. Biol. Chem.*, **270**, 3944–3948.

Received on February 28, 1997; revised on April 21, 1997