A statistical analysis of N- and O-glycan linkage conformations from crystallographic data

Andrei J.Petrescu[1,2], Stefana M.Petrescu[1,2], Raymond A.Dwek[1] and Mark R.Wormald[1,3]

[1]Oxford Glycobiology Institute, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK and [2]Institute of Biochemistry, Romanian Academy, PO Box 37, 296 Spl. Independentei, 77700 Bucharest 17, Romania

[3]To whom correspondence should be addressed at: Oxford Glycobiology Institute, Department of Biochemistry, South Parks Road, Oxford OX1 3QU, UK

We have generated a database of 639 glycosidic linkage structures by an exhaustive survey of the available crystallographic data for isolated oligosaccharides, glycoproteins, and glycan-binding proteins. For isolated oligosaccharides there is relatively little crystallographic data available. A much larger number of glycoprotein and glycan-binding protein structures have now been solved in which two or more linked monosaccharides can be resolved. In the majority of these cases, only a few residues can be seen. Using the 639 glycosidic linkage structures, we have identified one or more distinct conformers for all the linkages. The O5-C1-O-C(x)′ torsion angles for all these distinct conformers appear to be determined chiefly by the exo-anomeric effect. The Manα1–6Man linkage appears to be less restrained than the others, showing a wide degree of dispersion outside the ranges of the defined conformers. The identification of distinct conformers for glycosidic linkages allows "average" glycan structures to be modeled and also allows the easy identification of distorted glycosidic linkages. Such an analysis shows that the interactions between IgG Fc and its own N-linked glycan result in severe distortion of the terminal Galβ1–4GlcNAc linkage only, indicating the strong interactions that must be present between the Gal residue and the protein surface. The applicability of this crystallographic based analysis to glycan structures in solution is discussed. This database of linkage structures should be a very useful reference tool in three-dimensional structure determinations.

Key words: oligosaccharide/linkage conformers/linkage distortion

Introduction

Oligosaccharide structures are typically determined by characterizing the glycosidic linkages between the rigid monosaccharide units. Oligosaccharides and the glycan components of glycoproteins are notoriously difficult to study by crystallography, either because they do not crystallize or because crystallographic disorder leads to a lack of identifiable electron density. A lack of crystallographic data has meant that nuclear magnetic resonance (NMR) spectroscopy and theoretical calculations have been the only available techniques to provide structural information on glycosidic linkages to date (Peters and Pinto, 1996; Imberty, 1997). NMR techniques are often unable to fully define a linkage structure due to a lack of experimental data (Wooten *et al*., 1990), while theoretical calculations are limited by the accuracy of the theory used. A second problem occurs with the variable flexibility of oligosaccharides, again making linkage conformational analysis more difficult (Xu *et al*., 1996). As a result of this, relatively few oligosaccharide structures have been determined, and so general conformational rules governing glycosidic linkages have not been established. Databases of disaccharide linkage conformations have been compiled based on theoretical calculations (Imberty *et al*., 1990, 1991), but the accuracy of these is limited by the available experimental data with which to check the theory.

Crystallographic analysis has the potential advantage over NMR and theoretical methods that it can provide a complete oligosaccharide conformation from experimental data. The major limitation of this approach is that it will only work on static structures and so will give no direct information on the dynamic nature of a glycosidic linkage. It is also likely that a single static linkage conformation observed in a crystal will not correspond to the average solution conformation. However, the average conformation for a given linkage within a large enough sample of static structures is likely to correspond well to the average solution conformation and the distribution of static structures will give an indication of the flexibility of the linkage, as long as the packing forces in the available crystals do not impose systematic changes in the linkage conformations.

In this article, we survey the available (and surprisingly large) crystallographic data on oligosaccharide structures which contain linkages found in N- and O-linked glycans and use simple statistical analysis to identify distinct conformers for glycosidic linkages. As well as enabling "average" oligosaccharide conformations to be determined, this allows easy identification of distorted glycosidic linkages in specific structures. This also provides a much larger body of experimental data on which to test theoretical models.

Results and discussion

The available crystallographic data on synthetic and isolated oligosaccharides are still very limited (Table I). In the last few years, many structures have become available of glycoproteins and glycans bound to proteins in which linked monosaccharides can be resolved (Table I). There is a lack of standard nomenclature for glycan entries in PDB files (in some cases entire glycan chains are entered as a single residue), making selective searching for structures more difficult. The quality of the glycan regions of glycoprotein structures is generally far more variable than for the peptide regions. Approximately 20% of all reported glycosidic linkages involve distorted/incorrect monosaccharides or incorrect linkages (see *Materials and methods* for definitions), all of them occurring in protein linked N-glycans (Table I). The following analysis is based on the 639 glycosidic linkage structures that do not involve severely distorted monosaccharides.

Table I. Public domain crystal structures at a resolution better than 3.0 Å containing oligosaccharides with glycosidic linkages found in N- or O-linked glycans

Type of structure	Crystal structures	Glycan structures	Linkages between undistorted saccharides[a]	Incorrect linkages or linkages between distorted saccharides
Unmodified oligosaccharides	9	10	11 (8)	0
Glycoproteins with N-glycans	110	208	441 (441)	134
Glycoproteins with O-glycans	2	2	2 (0)	0
Proteins with glycan ligands	23	64	185 (184)	0

[a]The number in brackets is the number of linkages which meet the criteria for inclusion in Table II. The other six linkage structures, of which there is only one example each, are Arabα1–2Man, Fucα1–4GlcNAc, Galα1–4Gal, Galβ1–4Man, Glcα1–3Glc, and Manα1–4Man.

Table II. Average torsion angles and standard deviations for all distinct conformers (see text for details) of glycosidic linkages found in either N- or O-linked glycans. For definitions of φ, ψ and ω see *Materials and methods*

| Glycosidic linkage	No. of structures	Avg. linkage torsion angles for distinct conformers			Conformer population
		φ	ψ	ω	
Linkages for which there are at least 10 examples from at least five different crystal structures:					
Fuc α1–3 GlcNAc[a]	21	-70.7 ± 6.9	-101.7 ± 8.1	—	19
Fuc α1–6 GlcNAc	16	-68.2 ± 9.6	204.1 ± 22.4	66.1 ± 14.0	13
Gal β1–4 GlcNAc[b]	32	-70.4 ± 9.1	129.5 ± 7.1	—	23
GlcNAc β1–4 GlcNAc	163	-73.7 ± 8.4	116.8 ± 15.6	—	146
GlcNAc β1–2 Man	47	-80.2 ± 9.7	-97.2 ± 22.3	—	36
		58.3 ± 9.4	-87.2 ± 15.2	—	8
Man β1–4 GlcNAc	103	-88.0 ± 10.8	107.9 ± 20.3	—	89
Man α1–2 Man	48	62.2 ± 8.3	-175.0 ± 10.3	—	13
		71.9 ± 13.1	-104.4 ± 15.4	—	34
Man α1–3 Man	91	72.5±11.0	-112.3 ± 22.5	—	84
Man α1–6 Man	69	65.4 ± 9.0	182.6 ± 5.1	66.4 ± 10.2	23
		66.5 ± 10.8	180.7 ± 15.1	185.0 ± 11.2	18
		67.4 ± 14.4	109.1 ± 13.7	203.0 ± 22.7	12
Others (for which there are examples from at least two different crystal structures):					
Fuc α1–2 Gal	4	-66.5 ± 2.2	137.5 ± 1.1	—	2
		-92.7 ± 0.1	64.8 ± 0.4	—	2
Gal β1–3 GlcNAc	12	-74.3 ± 10.0	-131.5 ± 18.3	—	12
GlcNAc β1–4 Man	2	-170.0 ± 10.7	94.7 ± 6.1	—	2
NeuAc α2–3 Gal	14	68.7 ± 13.6	-125.1 ± 15.5	—	14
NeuAc α2–6 Gal	7	144.3 ± 2.5	188.6 ± 1.9	51.3 ± 4.8	3
		36.5 ± 2.4	153.0 ± 12.4	179.3 ± 6.4	2
		148.7	130.4	158.5	1
		294.5	122.2	30.1	1
Xyl β1–2 Man	4	-91.5 ± 6.6	-105.8 ± 3.9	—	4

[a]Includes both core and outer arm fucose linkages.
[b]Includes four structures with sulfated galactose at the 3- or 4- position. These all have conformations within the distinct conformer region.

In most crystal structures, only part of the glycan is seen in the electron density maps. About 50% of N-linked glycan structures contain only two or three resolved residues, indicating a general high degree of glycan mobility or disorder. Thus, there are a large number of structures for the GlcNAcβ1–4GlcNAc and Manβ1–4GlcNAc linkages and far fewer for linkages which occur closer to the nonreducing terminus of glycans such as Galβ1–4GlcNAc (Table II). There are only six glycoproteins in which seven or more residues of a single glycan can be resolved (Figure 1) and only five glycan binding proteins in which five or more residues of a single glycan can be resolved (Figure 2). In these cases, large regions of the glycan structure are seen because either the glycan lies along the protein surface or the glycan bridges between two adjacent protein molecules, in both cases leading to its immobilization. Comparison of the glycan structures found in these eleven crystal structures (Figure 3) shows a relatively wide range of conformations for a given type of glycan. This is probably due to both the inherent range of structures that
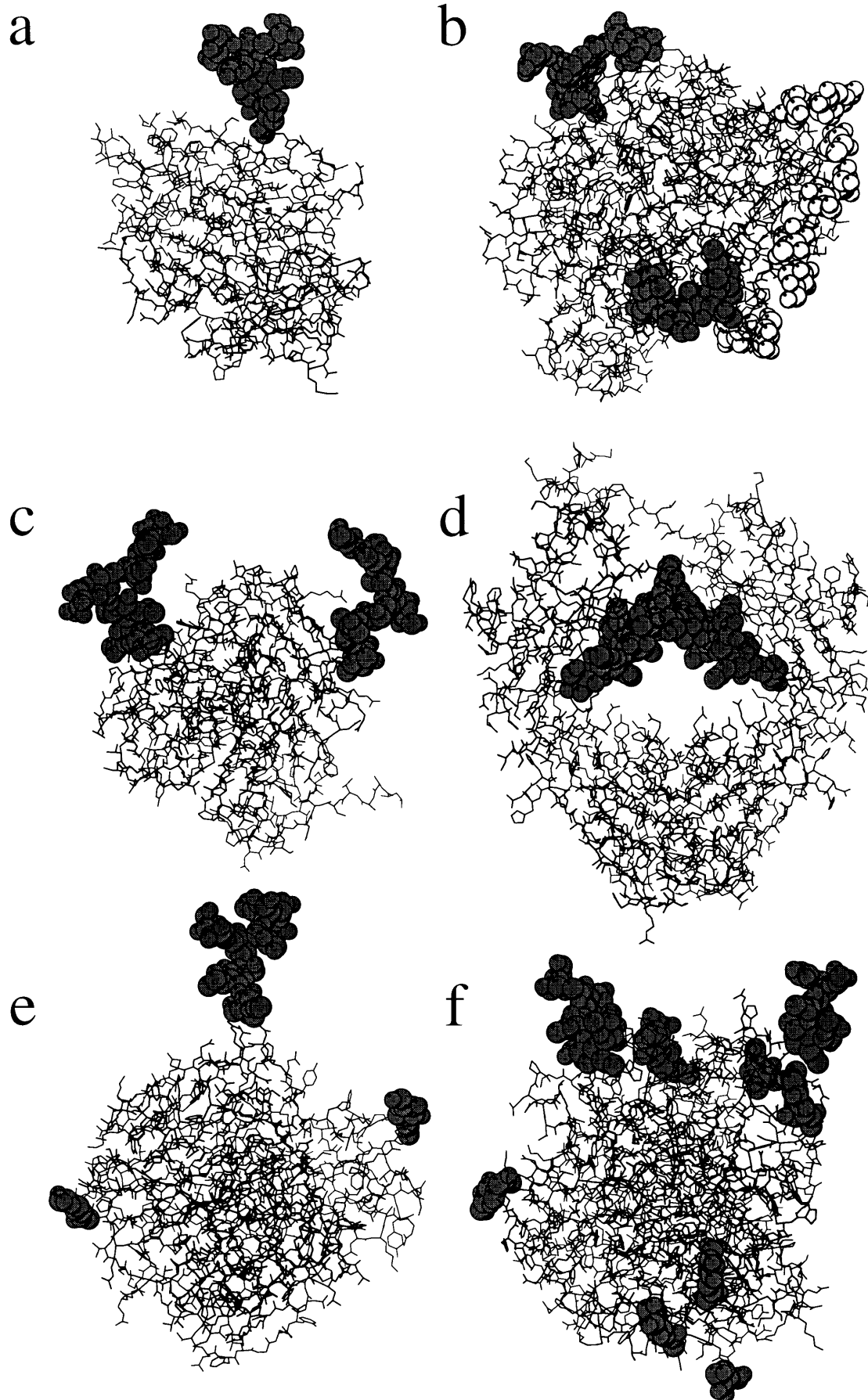
Fig. 1. All available crystal structures of glycoproteins containing N-linked glycans with seven or more linked residues resolved in the structure. (**a**) *Erythrina corallodendron* lectin (Shaanan *et al.*, 1991). The structure contains a single plant type N-glycan which makes contact with a neighboring protein molecule in the crystal. (**b**) Glucoamylase (Aleshin *et al.*, 1994). The structure contains two oligomannose type N-glycans which lie along the protein surface. There are also 10 O-linked monosaccharides shown in lighter gray. (**c**) Human leukocyte elastase (Bode *et al.*, 1989). The structure contains two complex type N-glycans which make contacts with neighboring protein molecules in the crystal. (**d**) Fc region of an intact IgG2a monoclonal antibody (Harris *et al.*, 1997). A separate structure is also available for the Fc domain (Deisenhofer, 1981). The Fc region contains two complex type N-glycans situated between the protein domains and lying along the protein surface. Although most of the glycans are not distorted, the terminal galactose residue in each glycan is in a boat conformation instead of a chair, the glycosidic linkage between the core N-acetyl-glucosamine residues in each glycan is α1–4 and the core fucose in each glycan is linked β1–6. (**e**) Influenza neuraminidase (White *et al.*, 1995). The structure contains three N-glycans but only one of these (oligomannose type) is resolved beyond the first residue. (**f**) Myrosinase (Burmeister *et al.*, 1997). The structure contains nine N-glycans but only two of these (plant type) are resolved beyond the first or second residue.

---

a given isolated glycan can adopt and the additional variations caused by the various extensive interactions between the glycan and its own, or adjacent, protein.

The Brookhaven database also includes 19 structures of enzymes (all lysozyme) containing in their active sites glycan substrates with N- or O-type glycosidic linkages. All of these are GlcNAcβ1–4GlcNAc linkages. Because a major role of enzyme binding sites is to distort the substrate, these have not been included in any of the analyses. However, it is worth noting that out of the 40 examples of the GlcNAcβ1–4GlcNAc linkage from these structures 38 of them fall within the range identified as the only distinct conformer for this linkage (see below).

The φ/ψ plots and histogram plots for the nine glycosidic linkages for which there are at least ten examples from at least five different crystal structures are shown in Figures 4–6. The data for linkages that do not meet these criteria but where there are examples from at least two different crystal structures are summarized in Table II. As can be seen, the different linkages show different degrees of conformational dispersion. Little dispersion is seen for the three α1–2/3 linkages (Figure 5). The Manβ1–4GlcNAc and GlcNAcβ1–4GlcNAc linkages appear to show considerable dispersion on the basis of the φ/ψ plot (Figures 4c,g), but it is clear from the histogram plots (Figures 4d,h) that about 90% of the structures fall within well defined φ/ψ regions. In contrast, considerable scatter is seen for the Galβ1–4GlcNAc, Manα1–6Man and Fucα1–6GlcNAc linkages (Figures 4a, 6a,b).

The identification of distinct conformers for any given glycosidic linkage must be somewhat subjective at this stage. The rules that we have chosen to use to identify distinct conformers are that there must be at least two peaks with a clear minimum between them in the histogram plot of at least one of the torsion angles, adjacent peaks in the histogram plot must be separated by at least 60° and each conformer must be represented by at least 10% of the total sample population for that linkage. The ranges of torsion angles associated with each conformer can be judged from the width of the peaks in the histogram plots and the dispersion of the peaks in the torsion angle plots. In cases of doubt we have generally included rather than excluded structures from the conformer regions, to give the largest possible populations for statistical analysis. Figures 4–6 show the torsion angle regions identified as distinct conformers for these glycosidic linkages. Table II gives the average torsion angles and standard deviations for all structures within the distinct conformer regions identified for each linkage, together with the conformer populations. All glycosidic linkages give at least one distinct conformer using these criteria and ~88% of linkage structures fall within their distinct conformer regions. This method of identifying distinct conformers is obviously not applicable in cases such as NeuAc-α2–6Gal where there are very few structures, and so all the following statistical analyses are limited to those linkages listed

in Table II for which there are at least 10 examples from at least five different crystals.

Using these criteria, the Manβ1–4GlcNAc linkage (Figures 4c,d) is identified as having a single conformer. Although there are three separate peaks in the histogram plot for the ψ torsion angle, adjacent peaks are only separated by 20°. This meets the criteria for a single conformer and leads to a large standard deviation for the average ψ value (Table II). The Manα1–2Man linkage (Figure 5c,d) is identified as having two distinct conformers. In this case, there are two separate peaks in the histogram plot for the ψ torsion angle separated by 80°.

All the distinct conformers identified by these criteria for α-linkages show similar φ values, +69.9° ± 11.6° for d-Man (184 structures) and -69.7° ± 8.1° for l-Fuc (32 structures). The Fucα1–2Gal and NeuAcα2–3Gal linkages also give similar φ values but considerable variation is seen for the NeuAcα2–6Gal linkage φ angle (Table II). However, there are very few examples of these linkages. Theoretical Hartree–Fock calculations on a simple model system with an α-linkage show a single minimum energy for this torsion angle at +60° for d-saccharides (Woods *et al.*, 1995). More variation is seen in the φ values for the β-linkages. Most β-linkage distinct conformers have similar φ values, -78.5° ± 11.6° for d-Gal, d-GlcNAc and d-Man (258 structures, not including GlcNAcβ1–2Man linkage). However, the GlcNAc-β1–2Man linkage has two well-populated conformers with average φ values of -80.2° and +58.3° (Table II). Model Hartree–Fock calculations show two minima for a β-linkage, a lower energy minimum at -60° and a higher energy minimum at +60° (Woods *et al.*, 1995). The Galβ1–3GlcNAc and Xyl-β1–2Man linkages show similar φ values but the GlcNAc-β1–4Man linkage shows yet another φ value of -170.0°. However, again there are very few examples of these linkages. Thus, the φ-angles of the identified conformers for both α- and β-linkages fit the molecular orbital calculations remarkably well, suggesting that rotation about the C1-O bond is dominated by the exo-anomeric effect in all the distinct conformers.

It is interesting to note that for the β1–4 linkages the nonreducing terminal residue appears to have very little effect on the observed linkage conformers, Galβ1–4GlcNAc including sulfation of the galactose at the 3- and 4-positions (Figure 4a), Manβ1–4GlcNAc (Figure 4c) and GlcNAcβ1–4GlcNAc (Figure 4g) having distinct conformers with very similar φ and ψ values.

For the two α1–6 linkages, three conformers can be identified for Manα1–6Man whereas only a single distinct conformer can be identified for Fucα1–6GlcNAc (Figure 6). This may simply be due to lack of data for the Fucα1–6GlcNAc linkage. Alternatively, modeling of the Fucα1–6GlcNAc linkage in a complex-type N-glycan shows that a ω torsion angle of ~190° (as is observed in the other Manα1–6Man conformers) would bring the fucose ring very close to the N-acetyl group of the second GlcNAc residue. The observed ω value of ~66° orients the fucose more
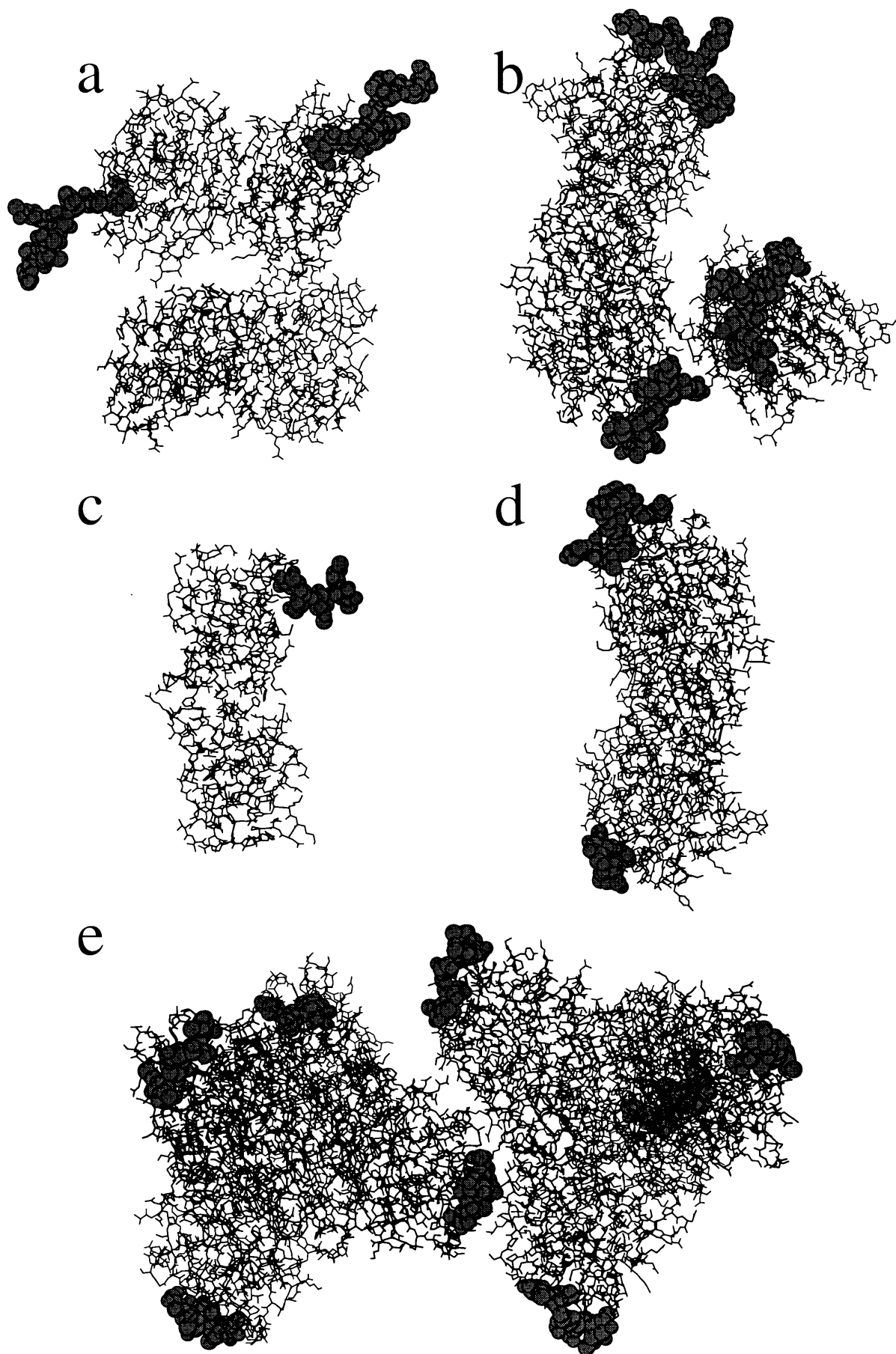
Fig. 2. All available crystal structures of glycan binding proteins complexed with N-linked type glycans with five or more linked residues resolved in the glycan chain. (**a**) Galectin 1 complexed with a complex type oligosaccharide (Bourne *et al*., 1994a). Each protein monomer binds to a terminal oligosaccharide residue. Each oligosaccharide forms a bridge between two monomers in adjacent unit cells. (**b**) Legume isolectin II complexed with a complex type glycopeptide (Bourne *et al*., 1994b). Each protein monomer binds to a single glycan, involving a large complementary surface area between the glycan and the lectin. (**c**) Mannose binding protein complexed with an oligomannose type glycopeptide. (Weis *et al*., 1992). Each protein monomer binds to a terminal oligosaccharide residue. Each glycan forms a bridge between two monomers in adjacent unit cells. (**d**) Legume isolectin I complexed with a complex type glycan (Bourne *et al*., 1992). Each protein monomer binds to a single glycan, involving a large complementary surface area between the glycan and the lectin. (**e**) Concanavalin A complexed with a complex type glycan (Moothoo and Naismith, 1998). Each protein monomer binds to a single glycan, using a continuous cleft on the protein surface and interacting with both nonreducing terminal residues of the glycan.

towards the protein surface. Thus, the single conformer observed for the Fucα1–6GlcNAc linkage may well be a function of the GlcNAcβ1–4(Fucα1–6)GlcNAc structural unit, rather than an inherent property of the glycosidic linkage. The average torsion angles for the GlcNAcβ1–4GlcNAc linkage conformer are virtually unchanged by the presence or absence of a 6-linked (or 3-linked) fucose residue.

Most glycosidic linkages show a relatively small degree of dispersion outside the ranges of these distinct conformers, the exceptions being the Galβ1–4GlcNAc and Manα1–6Man linkages, with 28% and 23% of the linkage structures falling outside the identified distinct conformer regions, respectively (Table II). For the Manα1–6Man linkage, this probably reflects the greater flexibility of the 1–6 linkage (there being three variable torsion angles rather than two) and thus its greater susceptibility to distortion by other interactions of the glycan within the crystal (such as with the protein surface or with neighboring protein molecules). The relatively large dispersion seen for the Galβ1–4GlcNAc linkage (Figures 4a,b) is slightly misleading because of the small number in the sample (32 structures). A few specific cases of altered linkage conformations will have a large effect on the percentage dispersion.

Having identified distinct conformers for linkages, it is very easy to identify specific distorted structures. The three Galβ1–4GlcNAc linkage structures (at -177°, 59°) that differ most from the single identified conformer are from IgG Fc molecules, free and complexed with a fragment of protein A (Deisenhofer, 1981), in which the glycan lies along the protein surface. The Galβ1–4GlcNAc linkage appears to be distorted considerably by this interaction. All the other linkage structures in the IgG Fc glycans fall within the conformer ranges for their respective linkages, consistent with the major glycan–protein interactions involving the terminal galactose residue. The presence of strong interactions specifically between the terminal 6-arm galactose residue and the protein surface have been shown by loss of this galactose residue leading to release of the glycan from the protein surface (Wormald *et al*., 1997).

When considering the range of conformations that we might expect free glycans to adopt in solution, two further points need to be considered. As most protein crystal structures are obtained from crystals at around 100 K, the linkages will give a much narrower distribution about a minimum energy linkage conformer than would be found in solution at 300 K. However, additional crystal packing forces may lead to glycosidic linkage distortions

a   b   c
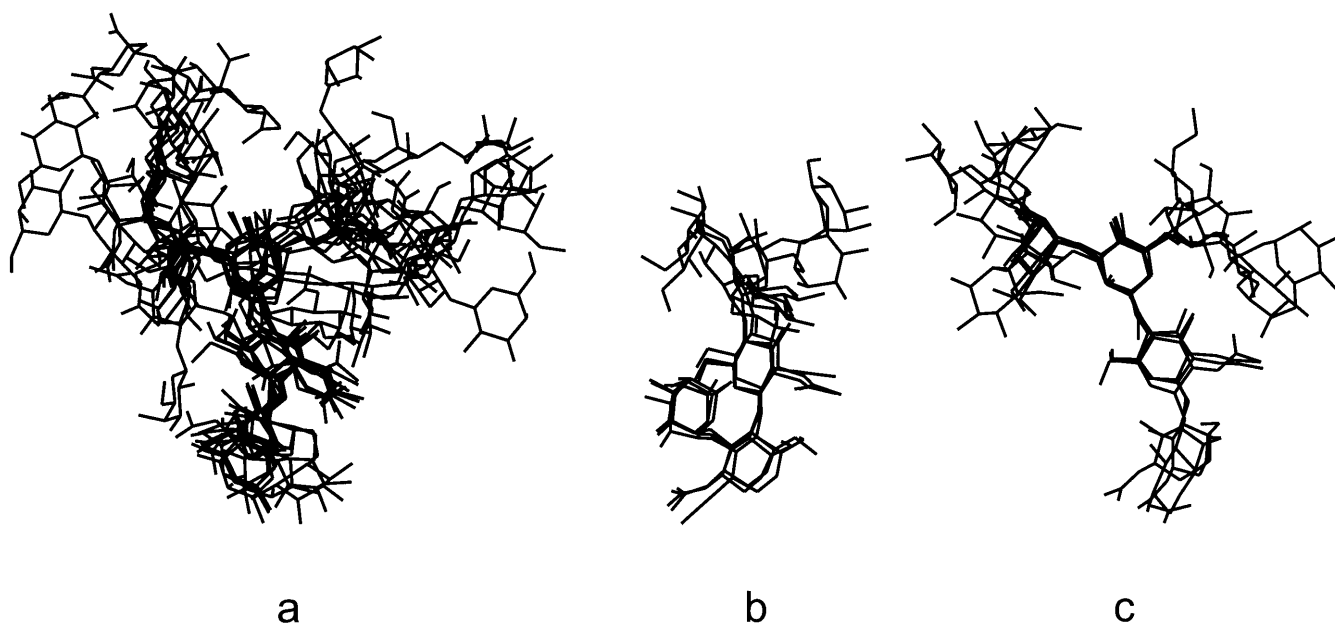
Fig. 3. Overlays of the oligosaccharide structures from the selected glycoproteins shown in Figures 1 and 2. (**a**) Complex type biantennary oligosaccharides—two N-glycans from elastase (Figure 1c), two N-glycans from IgG2a Fc (Figure 1d), two substrate chains from galectin 1 (Figure 2a), three substrate chains from legume isolectin II (Figure 2b), the octasaccharide substrate from legume isolectin I (Figure 2d), and the eight substrate chains from Concanavalin A (Figure 2e). (**b**) Plant type oligosaccharides containing xylose and three-linked core fucose—one N-glycan from *erythrina* lectin (Figure 1a) and two N-glycans from myrosinase (Figure 1f). (**c**) Oligomannose type oligosaccharides—two N-glycans from glucoamylase (Figure 1b), one N-glycan from influenza neuraminidase (Figure 1e), and one substrate chain from mannose binding protein (Figure 2c).
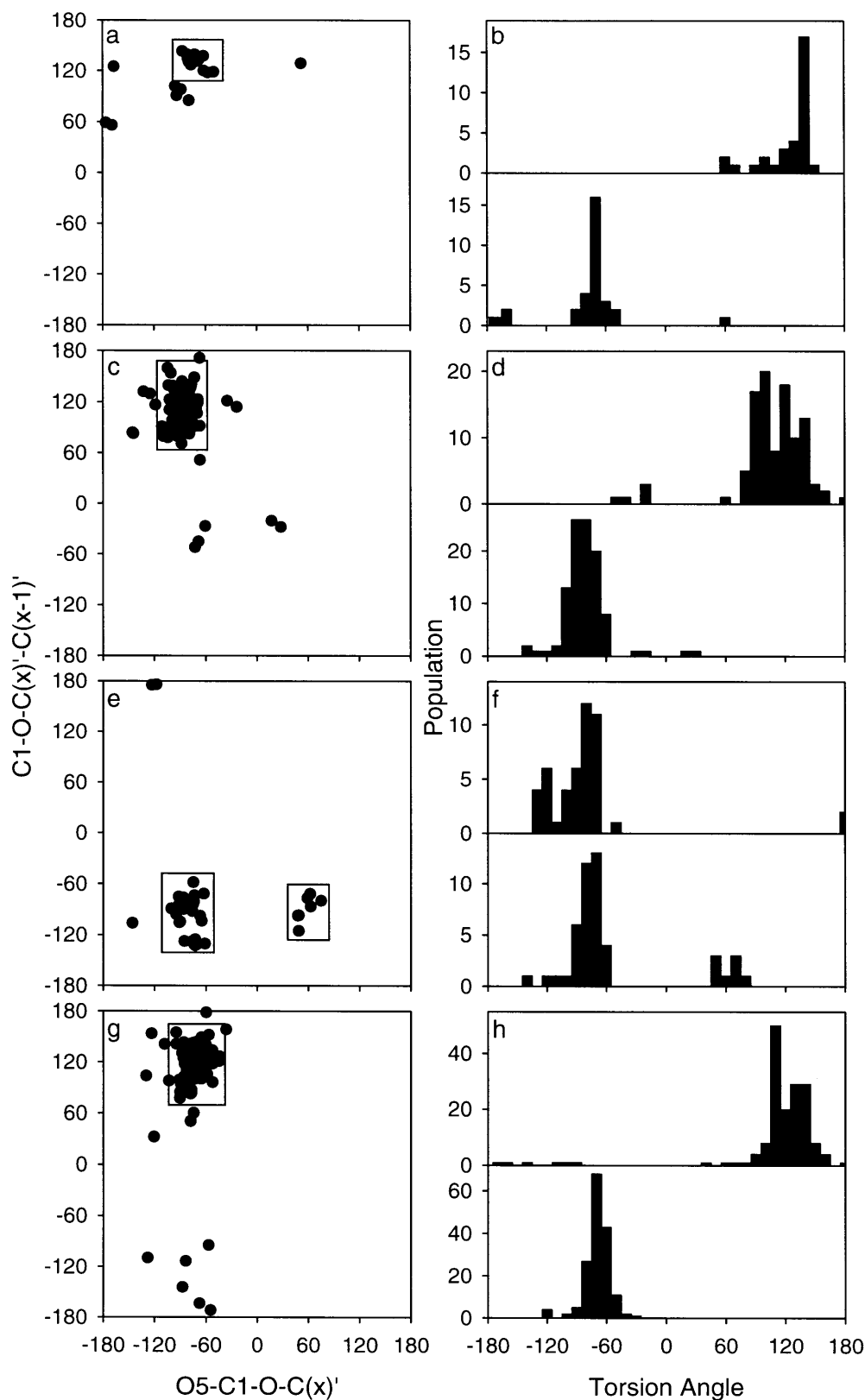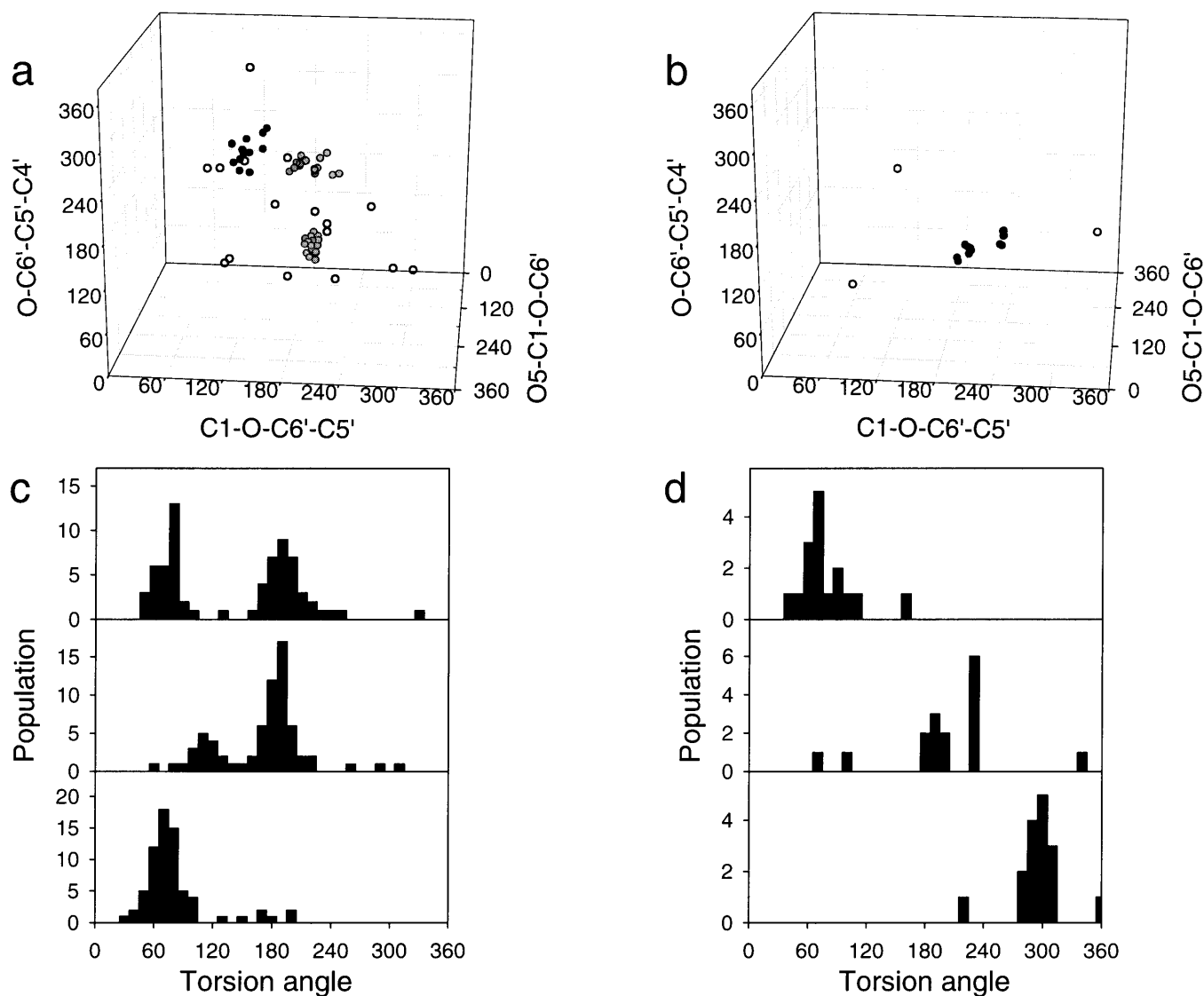
Fig. 4. φ/ψ torsion angle plots and population histograms for β-glycosidic linkages for which there are at least 10 examples from at least five different crystal structures. (**a**), (**c**), (**e**), and (**g**) are plots of O5-C1-O-C(x)′ versus C1-O-Cx′-C(x-1)′ for a β1-x linkage. The boxed regions show the areas identified as distinct conformers (see text for details). (**b**), (**d**), (**f**), and (**h**) are plots of histogram population (using a 10° window) versus torsion angle, the lower panel of each giving the O5-C1-O-C(x)′ histogram and the upper panel giving the C1-O-Cx′-C(x-1)′ histogram. (**a**) and (**b**), Galβ1–4GlcNAc linkage. This also includes four structures with sulfated Gal at either the 3- or 4-position. These all fall within the boxed region. (**c**) and (**d**), Manβ1–4GlcNAc linkage. (**e**) and (**f**), GlcNAcβ1–2Man linkage. (**g**) and (**h**), GlcNAcβ1–4GlcNAc linkage.
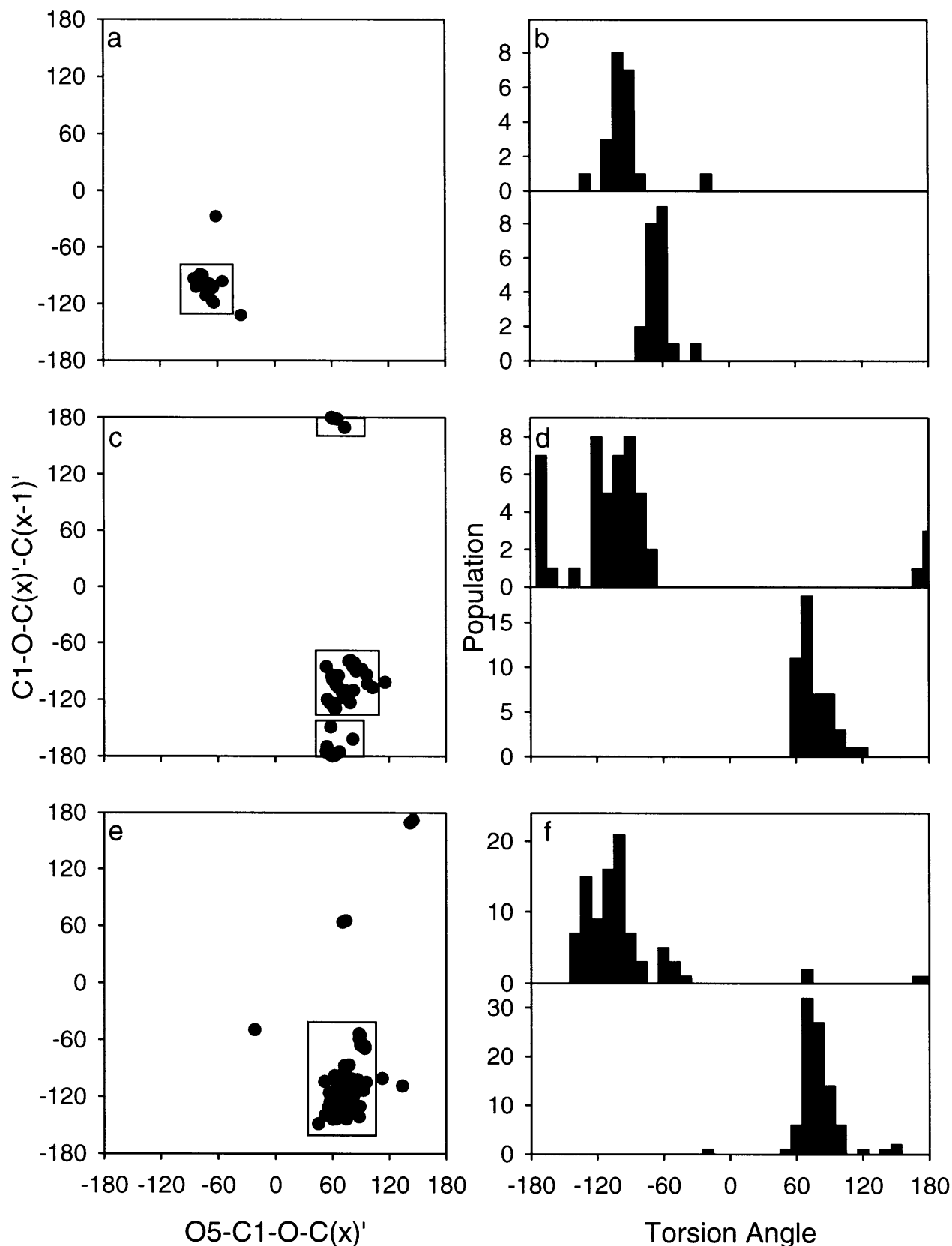
Fig. 5. φ/ψ torsion angle plots and population histograms for α-glycosidic linkages (except α1–6, see Figure 6) for which there are at least 10 examples from at least five different crystal structures. (**a**), (**c**), and (**e**) are plots of O5-C1-O-C(x)′ versus C1-O-Cx′-C(x-1)′ for a α1-x linkage. The boxed regions show the areas identified as distinct conformers (see text for details). (**b**), (**d**), and (**f**) are plots of histogram population (using a 10° window) versus torsion angle, the lower panel of each giving the O5-C1-O-C(x)′ histogram and the upper panel giving the C1-O-Cx′-C(x-1)′ histogram. (**a**) and (**b**), Fucα1–3GlcNAc linkage. (**c**) and (**d**), Manα1–2Man linkage. (**e**) and (**f**), Manα1–3Man linkage.

Fig. 6. φ/ψ/ω torsion angle plots and population histograms for α1-6 glycosidic linkages for which there are at least ten examples from at least five different crystal structures. (a) and (b) are plots of O5–C1–O–C6' versus C1–O–C6'–C5' versus O–C6'–C5'–C4'. Solid circles, structures belonging to distinct conformers (Table II); open circles, other structures. (c) and (d) are plots of histogram population (using a 10° window) versus torsion angle, the lower panel of each giving the O5–C1–O–C6' histogram, the middle panel giving the C1–O–C6'–C5' histogram and the upper panel giving the O–C6'–C5'–C4' histogram. (a) and (c), Manα1-6Man linkage. (b) and (d), Fucα1-6GlcNAc linkage. The O5–C1–O–C6' axis in (b) is reversed relative to (a) to enable easy comparison between the linkage conformations for D–Man and L–Fuc.

and thus a larger range of linkage conformations than would be observed for free glycans. In the majority of crystal structures used, only small regions of the glycans give resolvable electron density, the rest being too mobile or disordered. In these cases, additional crystal interactions are likely to be small. In the other cases, as long as the additional interactions present in protein crystals do not cause systematic changes in the glycosidic linkage conformations (i.e., they can alter torsion angles in either direction) they will not effect the calculated average torsion angles. Thus, the average torsion angles for the linkage conformers in the crystalline and solution state are likely to be similar but the distributions around these averages will be different.

The glycosidic linkage conformers given in Table II can be used to construct "average" structures for common N-glycans. Such structures are likely to be a good representation of the overall shape and topology of the glycan. Individual structures will still have to be determined on a case by case basis where accurate atomic level information is required. The statistical analysis of linkage conformers also allows easy and rapid identification of distorted linkages in individual glycans. This can be used both as a quality control measure during structure refinement and as an indication of the degree of specific interactions between a monosaccharide residue and its immediate environment (often the protein surface or binding site).

Materials and methods

X-Ray crystal structures containing glycosidic linkages between unprotected monosaccharides were obtained by exhaustive searching of the Cambridge Crystallographic Database (Allen

and Kennard, 1993) at the Chemical Database Service at Daresbury (Fletcher *et al*., 1996) and the Brookhaven Protein Database (Bernstein *et al*., 1977). Only structures at a resolution of 3 Å or better were used. Where entries are available for the same crystal at different resolutions, only the best resolved structure was used. Monosaccharides were defined as incorrect if they did not have the right configuration (e.g., 5-epi-fucose instead of fucose) or as distorted if the monosaccharide rings were not in a low energy conformation (e.g., not in a chair form). Any glycosidic linkage involving a distorted/incorrect monosaccharide was not used in the data analysis. Linkages were only defined as incorrect if they occurred in glycans derived from biological sources but are biosynthetically unknown (e.g., a GlcNAc-$\alpha1$–4GlcNAc linkage in the core of an N-linked glycan) or involved impossible bond lengths or angles (e.g., a C1-O-Cx bond angle of 80°). Glycosidic linkage torsion angles were measured for every available glycosidic linkage that occurs in N-linked or O-linked glycans, regardless of the type of glycan structure in which the linkage was found. Histogram plots for each linkage were obtained by counting the number of structures with a given torsion angle within a specific 10° window (-180° to -170°, -170° to -160°, etc.). Molecular modeling was performed on a Silicon Graphics Indigo 2 workstation using InsightII software (MSI). The nomenclature used for the torsion angles are $\phi = $ O5-C1-O-C(x)′ and $\psi = $ C1-O-C(x)′-C(x-1)′ for 1–2, 1–3, and 1–4 linkages (x = 2, 3, or 4); $\phi = $ O5-C1-O-C6′, $\psi = $ C1-O-C6′-C5′, and $\omega = $ O-C6′-C5′-C4′ for 1–6 linkages; and $\phi = $ O6-C2-O-C6′, $\psi = $ C2-O-C6′-C5′, and $\omega = $ O-C6′-C5′-C4′ for 2–6 linkages.

Acknowledgments

References

Aleshin,A.E., Hoffman,C., Firsov,L.M. and Honzatko,R.B. (1994) Refined crystal structures of glucoamylase from *Aspergillus awamori* var. X100. *J. Mol. Biol*., **238**, 575–591.

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

Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.E., Jr., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol*., **112**, 535–542.

Bode,W., Meyer,E. and Powers,J.C. (1989) Human leukocyte and porcine pancreatic elastase: x-ray crystal structures, mechanism, substrate specificity and mechanism-based inhibitors. *Biochemistry*, **28**, 1951–1963.

Bourne,Y., Rouge,P. and Cambillau,C. (1992) X-Ray structure of a biantennary octasaccharide-lectin complex refined at 2.3-Å resolution. *J. Biol. Chem*., **267**, 197–203.

Bourne,Y., Bolgiano,B., Liao,D.I., Strecker,G., Cantau,P., Herzberg,O., Feizi,T. and Cambillau,C. (1994a) Crosslinking of mammalian lectin (galectin-1) by complex biantennary saccharides. *Nat. Struct. Biol*., **1**, 863–870.

Bourne,Y., Mazurier,J., Legrand,D., Rouge,P., Montreuil,J., Spik,G. and Cambillau,C. (1994b) Structures of a legume lectin complexed with the human lactotransferrin N2 fragment, and with an isolated biantennary glycopeptide: role of the fucose moiety. *Structure*, **2**, 209–219.

Burmeister,W.P., Cottaz,S., Driguez,H., Iori,R., Palmieri,S. and Henrissat,B. (1997) The crystal structures of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure*, **5**, 663–675.

Deisenhofer,J. (1981) Crystallographic refinement and atomic models of human Fc fragment and its complex with Fragment B of Protein A from *Staphylococcus aureus* at 2.9 and 2.8 Å resolution. *Biochemistry*, **20**, 2361–2370.

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

Harris,L.J., Larson,S.B., Hasel,K.W. and McPherson,A. (1997) Refined structure of an intact IgG2a monoclonal antibody. *Biochemistry*, **36**, 1581–1597.

Imberty,A. (1997) Oligosaccharide structures: theory versus experiment. *Curr. Opinion Struct. Biol*., **7**, 617–623.

Imberty,A., Gerber,S., Tran,V. and Perez,S. (1990) Data-bank Of 3-dimensional structures of disaccharides, a tool to build 3-D structures of oligosaccharides. 1. Oligo-mannose type N- glycans. *Glycoconjugate J*., **7**, 27–54.

Imberty,A., Delage,M.M., Bourne,Y., Cambillau,C. and Perez,S. (1991) Data bank of three-dimensional structures of disaccharides. II. N-Acetyllactosaminic type N-glycans. Comparison with the crystal structure of a biantennary octasaccharide. *Glycoconj. J*., **8**, 456–483.

Moothoo,D.N. and Naismith,J.H. (1998) Concanavalin A distorts the beta-GlcNAc-(1→2)-Man linkage of beta-GlcNAc-(1→2)-alpha-Man-(1→3)-[beta-GlcNAc-(1→2)-alpha- Man- (1→6)]-Man upon binding. *Glycobiology*, **8**, 173–181.

Peters,T. and Pinto,B.M. (1996) Structure and dynamics of oligosaccharides— NMR and modeling studies. *Curr. Opinion Struct. Biol*., **6**, 710–720.

Shaanan,B., Lis,H. and Sharon,N. (1991) Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*, **254**, 862–866.

Weis,W.I., Drickamer,K. and Hendrickson,W.A. (1992) Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature*, **360**, 127–134.

White,C.L., Janakiraman,M.N., Laver,W.G., Philippon,C., Vasella,A., Air,G.M. and Luo,M. (1995) A sialic acid-derived phosphonate analog inhibits different strains of influenza virus neuraminidase with different efficiencies. *J. Mol. Biol*., **245**, 623–634.

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. *Carb. Res*., **203**, 13–17.

Wormald,M.R., Rudd,P.M., Harvey,D.J., Chang,S.C., Scragg,I.G. and Dwek,R.A. (1997) Variations of oligosaccharide-protein interactions in immunoglobulin G determine the site-specific glycosylation profiles and modulate the dynamic motion of the Fc oligosaccharides. *Biochemistry*, **36**, 1370–1380.

Xu,Q., Gitti,R. and Bush,C.A. (1996) Comparison of NMR and molecular modeling results for a rigid and a flexible oligosaccharide. *Glycobiology*, **6**, 281–288.