CHEMOENZYMATIC SYNTHESIS AND X-RAY CRYSTALLOGRAPHIC INVESTIGATION OF N-GLYCOPROTEIN LINKAGE REGION MODELS AND ANALOGS

DURAIKKANNU LOGANATHAN*, UDAYANATH AICH AND THRUNEELAKANTAN LAKSHMANAN

Department of Chemistry, Indian Institute of Technology Madras, Chennai – 600036 (India)
(Received 9 May 2005; Accepted 14 July, 2005)

Protein-linked glycans play key roles in many important biological processes as both recognition determinants and modulators of the intrinsic properties including folding of proteins. In view of their inherent structural complexity and microheterogeneity, structure-function correlation of glycoproteins remains a challenging problem in glycobiology. The linkage region constituents, GlcNAc and Asn, are conserved in the N-glycoproteins of all the eukaryotes. As the presentation of the sugar on the cell surface can be greatly altered by the rotation about the Asn-GlcNAc linkage, elucidation of the structure and conformation of the linkage region of glycoproteins is fundamental to a better understanding of the inter- and intramolecular carbohydrate-protein interactions. Our contributions on the development of a chemoenzymatic methodology for the synthesis of simple models and analogs of N-glycoprotein linkage region and their X-Ray crystallographic investigation are highlighted in this review. Results of our work suggest that the linkage region constituents may play a structural role meeting three essential requirements: rigidity, planarity and linearity.

Key Words: Carbohydrates; Structural glycobiology; N-glycoproteins; Linkage region; Chemoenzymatic synthesis; Glycosidase, X-ray diffractometry; Conformation; Hydrogen bonds

1 Introduction

The soldiers of Alexander the Great (356 - 323 BC) took from India to Macedonia a plant that produced “honey without bees” thereby introducing sugarcane and hence sugar to the European continent. The ancient process of obtaining sugar consisted of boiling the sugarcane juice until solids formed as the syrup cooled. The product looked like gravel and the Sanskrit word for gravel “sakkara” gave origin to the term “saccharide”. By the fourth century AD, lime was used in Egypt as a purifying agent and for crystallization, which is still a main step in refining. The process of crystallization of sugars and their derivatives continues, till date, to be a daunting task.

The naturally occurring mono-, di-, oligo- and polysaccharides have been serving the basic needs viz., food, clothing and shelter of human beings since time immemorial. Apart from their importance as energy sources and structural materials, carbohydrates were not thought of playing any major roles in biology until 1960’s. The understanding of the roles of sugars in various biological processes has been rather slow and strenuous path, as compared to that of nucleic acids and proteins, due mainly to their inherent structural complexity and diversity. This is evident from the fact that four different amino acids can be linked to form 24 distinct tetrapeptides whereas four different monosaccharides (considering only hexopyranose forms) can be combined to afford a staggering number of 35,650 unique tetrascarabides. Technological breakthroughs in chromatographic and spectroscopic methods in the 1970’s and 1980’s paved the way for the exciting discoveries on the roles of oligosaccharide (glycan) components of cell surface glycoconjugates that have led to the accelerated growth of carbohydrate science witnessed over the past two decades.

Glycan parts of glycoproteins have been shown to play two principal physiological roles. Firstly, glycans serve as recognition determinants in such important biological processes as cellular adhesion, differentiation and proliferation of cells, fertilization
and viral & bacterial infections. The first step in many of these processes is often mediated by protein-carbohydrate interactions. Secondly, oligosaccharides modify the intrinsic properties of the proteins to which they are conjugated. These properties include catalytic activity, resistance to proteolytic attack, solubility and conformational aspects including folding. Biosynthesis of glycoproteins is a complex process, which is ordered and non-random. As glycan chains of glycoconjugates are secondary gene products, unlike proteins, glycosylation is species- and cell-specific and is also dependent on the protein sequence, cellular phenotype and physiological environment. Formation of the sugar-amino acid linkage is a crucial event in the biosynthesis as it sets into motion a complex series of posttranslational enzymatic steps leading to the formation of a plethora of protein-bound oligosaccharides with diverse biological functions.

There are two general types of protein glycosylations that are known: N- and O-glycosylations (Fig. 1). In addition to these two common types of glycosylations found on cell surface proteins, a third type of glycosylation of proteins occurs in the nucleus and cytoplasm and has been called the O-GlcNAc modification. N-Glycosylation has been examined more thoroughly than O-glycosylation. The N-linked glycans are found covalently attached through the 2-acetamido-2-deoxy-β-D-glucopyranosyl moiety to the amide nitrogen of asparagine occurring in the consensus sequence Asn-Xaa-Ser/Thr, where Xaa is any amino acid other than Pro. The GlcNAcβ-Asn linkage is most widely observed in glycoproteins isolated from not only eukaryotes but also from archaea and eubacteria. Exceptions to GlcNAc as the attachment sugar have been found and these are Glcβ-Asn found in Halobacter halobium S layer glycoprotein and in mammalian laminin, GalNAc-Asn in Halobacter halobium S layer glycoprotein and L-Rha-Asn in the cell wall of Streptococcus sanguis. In spite of the other potential N-glycosylation sites present on the side chains of several amino acids (Fig. 2), exceptions to Asn as the attachment amino acid are not known. It is intriguing to note that Gln, a homolog of Asn has not been observed to be glycosylated in Nature.

Understanding the correlations between the myriad structures and varied functions of glycan components of glycoproteins is indeed a challenging problem in structural glycobiochemistry. The structural complexity of the glycan chains, their microheterogeneity, flexibility and the non-availability of these compounds in sufficient quantities pose considerable difficulties to progress in structure-function correlations. As the rotation about the Asn-GlcNAc linkage can profoundly alter the presentation of the sugar on the cell surface, elucidation of the structure and conformation of the linkage region of glycoproteins is fundamental to a better understanding of the inter- and intramolecular interactions.

\[
\begin{align*}
N\text{-Glycoprotein linkage region} & \\
O\text{-Glycoprotein linkage region} & \\
O\text{-GlcNAc linkage region}
\end{align*}
\]

Fig. 1
carbohydrate-protein interactions, the former influencing the role of glycans as recognition determinants whereas the latter largely governing their intrinsic properties. Even though crystal structures of several glycoproteins have been reported, most often part or all of the glycan is not observed in the high-resolution electron density map. Thus structural investigation using model compounds would be a valuable approach to understand the effect of structural variations on the linkage region conformation. Prior to the beginning of our work, only two isolated reports on the crystal structures of the linkage region model compounds, GlcNAcβAsn\(^\text{11}\) and GlcβAsn\(^\text{10}\) were available in the literature.

A major research program of our laboratory initiated in 1997 is aimed at determining the three-dimensional structure of the glycoprotein linkage region based on X-ray crystallography. As part of this program, we have developed a chemoenzymatic approach to the synthesis of many mono- and disaccharide models and analogs of the N-glycoprotein linkage region. β-1-N-amidoglycopyranoses that are mimics of GlcNAcβ-Asn linkage have been demonstrated for the first time as acceptors in glycosidase catalyzed oligosaccharide synthesis both under transglycosylation\(^\text{12}\) and reversed hydrolysis.\(^\text{12c}\) Several of these structurally homogeneous models and analogs have been prepared in single crystal form often as hydrates. We have performed X-ray crystallographic investigation of its first kind on the systematic changes in the linkage sugar as well as its aglycon structure on the N-glycosidic torsion. An analysis of packing reveals that the molecular assembly of these compounds is driven by unique patterns of finite and infinite chains of hydrogen bonds. Herein we present an account of our contributions highlighting the applications of glycosidase-based chemoenzymatic synthesis and X-ray crystallography as two powerful tools for gaining deep insights in structural glyobiology.

2 Chemoenzymatic Synthesis

2.1 Chemical Synthesis of β-1-N-Amido Sugars

The design of β-1-N-amidoglycopyranoses as models and as analogs of the N-glycoprotein linkage region is a novel approach in that while these compounds are structurally simple, they do represent the linkage region in a holistic fashion. A simplified retrosynthetic analysis of the N-glycoprotein linkage region (Scheme 1) led us to identify glycosylamines as synths for the preparation of various β-1-N-amidoglycopyranoses. As glycosylamines have a tendency to adopt β-anomeric configuration, these are ideally suited for the chosen purpose. β-D-Glycosylamines, prepared from the parent sugars by reaction with either ammonia in methanol\(^\text{13}\) or saturated ammonium bicarbonate in aqueous medium,\(^\text{13b}\) were transformed in reasonable yields to the corresponding glycosyl amides (2 – 13, Fig. 3) either by the selective N-acylation\(^\text{14a}\) in dry methanol (Scheme 2) or peracylation followed by selective de-O-acylation\(^\text{14b}\) (Scheme 3). The latter procedure was followed wherever selective N-acylation did not afford readily isolable solid product.

2.2 Glycosidase Catalysed Disaccharide Synthesis using Amidosugar Acceptors

Having prepared simple models and analogs by chemical method, efforts were directed toward
developing an enzymatic methodology for synthesizing larger di- and trisaccharide mimetics of the N-glycoprotein linkage region. Though there are several elegant chemical methods available for the synthesis of oligosaccharides, these involve extensive protection and deprotection steps leading to low overall yields and these also employ hazardous heavy metal salts like AgClO₄ and Hg(CN)₂ as
promoters in stoichiometric amounts. Enzymatic synthesis is an attractive alternative as (a) it involves single-step glycosylation allowing an “one-pot” synthesis with no protection or deprotection steps, (b) reactions are highly stereo- and regiospecific and (c) it is an environment friendly process and the reactions are performed typically in aqueous media at or near room temperature.
Glycosidases are employed by Nature for the degradation or hydrolysis of glycosidic linkages and are most often involved in the catabolism of oligosaccharides in vivo. Despite the fact that glycosidases are degradative in nature, they can be used to catalyse the construction of glycosidic linkages by appropriate modification of the reaction conditions. Glycosidases are very attractive catalysts for oligosaccharide synthesis as these hydrolyses are readily available, inexpensive, relatively stable, and easy to handle and require no co-factors unlike the glycosyl transferases. Glycosidase catalysed oligosaccharide synthesis is performed under two modes of reaction conditions, namely the kinetically controlled condition also known as transglycosylation (path A) or the equilibrium controlled condition also known as reversed hydrolysis (path C)\(^\text{15}\) (Scheme 4).

### 2.2.1 Transglycosylation

In this approach, typically an activated glycosyl donor (Gly-X) is used where X is a good leaving group such as \(p\)-nitrophenoxide (Scheme 4). Hydrolysis of this donor, catalyzed by the enzyme, proceeds via an enzyme-bound intermediate. In the hydrolysis reaction, this intermediate is transferred to water but it can also be transferred to an acceptor molecule such as another sugar (ROH) to generate a new oligosaccharide. Even though the activity of water in this reaction mixture is usually one or two orders higher than that of the acceptor to be glycosylated, the yield of the oligosaccharide obtained are much higher that what would be expected from thermodynamic calculations. This is explained on the basis of a higher affinity of the acceptor substrates to the enzyme bound donor complex rather than water.\(^\text{16}\) The various methods employed to decrease the extent of hydrolysis and hence improve glycosylation yields include variation of the aglycon moiety of the acceptor and also the lowering of water content in the reaction medium by the use of organic co-solvents.\(^\text{15}\) In addition, the variation of the acceptor to donor ratio, reaction temperature and reaction medium have also been demonstrated to have a defining effect on the glycosylation yields. The regioselectivity of transglycosylation has been demonstrated to depend on the source of the enzyme as well as the nature and anomeric orientation of the aglycon group of the acceptor sugars. Prior to our work, only free sugars or O-/S-glycosides have been used as acceptors.

### 2.2.2 Reversed Hydrolysis

The normal equilibrium position in an aqueous environment is in the direction of hydrolysis. This can, however, be reversed by incubating the enzyme in the presence of high concentrations of sugars and low concentrations of water. The high concentrations of the sugars are easily achieved owing to the fact that free sugars are employed unlike in transglycosylation where derivatized donors, which are typically apolar, are used. In addition to employing high concentrations of the substrates, Nilsson\(^\text{17}\) has demonstrated that performing the reaction at elevated temperatures has led to higher yields. The high concentration of the sugar prevents the enzyme from possible heat denaturation. Ajisaka\(^\text{18}\) has used activated carbon columns as molecular traps to enhance the yields in reversed hydrolysis reactions. The advantages with this procedure include: (a) unreacted substrates can be recovered and reused several times and hence the net yield of the reaction is higher and (b) the immobilized glycosidase column and the activated carbon column can be used repeatedly for the same as well as different reactions.

### 2.2.3 \(\beta\)-N-Acetilhexosaminidase Catalysed Reactions

With the objective of synthesizing disaccharide models and analogs of the chitobiosyl core structure of the linkage region (Scheme 1), transglycosylation was explored using \(p\)-nitrophenyl 2-deoxy-2-acetamido-\(\beta\)-D-glucopyranoside as the donor, the
monosaccharide model of GlcNAcβ-Asn, GlcNAcβNHAc (2), as the acceptor and the commercially available title enzyme from Aspergillus oryzae (Scheme 5). The reaction was very successful in affording the desired (1,4)-linked disaccharide essentially as the single regioisomer in 17% yield. This product, which could be readily purified from the reaction mixture by flash column chromatography, was thoroughly characterized based on spectroscopic techniques including two-dimensional homonuclear COSY and heteronuclear 1H-13C HMQC and HMBC. Attempted transglycosylation to GlcbNHAc (3) using the same enzyme under similar conditions did not afford any disaccharide product revealing the high specificity of the enzyme for the acceptor.

Earlier reports on glycosidase catalysed reversed hydrolysis have invariably utilized only free sugars as reactants. The product oligosaccharides synthesised in this manner consist of several regioisomers each existing as anomic mixtures rendering their separation extremely difficult. An innovative modification involving the use of GlcNacbNHAc (2) with a fixed aglycon group and GlcNAc, the free sugar taken in large excess, as co-reactants in reversed hydrolysis catalysed by the title enzyme from Aspergillus oryzae has led to formation of the (1,6)-linked disaccharide as the only isolable product in 13% yield (Scheme 6).

Replacement of 2 with GlcNacbNHPt (5), a model
of the hitherto unknown GlcNAcβ-Gln linkage, as a co-reactant in this reaction also proved successful in furnishing the corresponding (1,6)-linked disaccharide in 8% yield (Scheme 6). The excellent and mutually complementary (1,4) and (1,6) regioselectivities observed under transglycosylation and reversed hydrolysis conditions, respectively, would be very valuable in synthesizing various extended glycopeptide analogs of the linkage region.

Considering the limited number of commercially available β-N-acetylhexosaminidases, there is a need to identify newer and relatively inexpensive sources. Our efforts in this direction have led to partial purification of β-N-acetylhexosaminidase from mung beans (Vigna radiata) and successful demonstration of its synthetic utility. Transglycosylation catalysed by this enzyme employing p-nitrophenyl 2-deoxy-2-acetamido-β-D-glucopyranoside as the donor and GlcNAcβNHAc (2), as the acceptor afforded again a single disaccharide product, characterized as the (1,6)-linked regioisomer, in 12% isolated yield (Scheme 7). Interestingly, β-N-acetylhexosaminidases obtained from the microbial (Aspergillus oryzae) and the plant (mung bean) sources exhibit complementary regioselectivities, (1,4) and (1,6) respectively, under transglycosylation mode of oligosaccharide synthesis.

2.2.4 β-Galactosidase Catalysed Reactions

β-Galactosidase from Bacillus circulans, a commercially available and relatively inexpensive crude enzyme preparation, has earlier been shown to have a high degree of tolerance with respect to acceptor sugars in transglycosylation reactions. Enzymatic β-galactosyl transfer to GlcNβNHAc (3), we reasoned, would generate interesting structural variants of Glcβ-Asn linkage for crystalllographic investigation. Hence, transglycosylation catalysed by this enzyme was explored using α-nitrophenyl β-D-galactopyranoside as the donor and 3 as the acceptor (Scheme 8). A systematic variation of reaction conditions including the donor to acceptor ratio, water content of the medium (by using varying concentration of organic co-solvent), pH, temperature and amount of enzyme used led to improvement in the yield of disaccharides from 5 to 42%. The composition of the product mixture obtained in 41% yield turned out to be 35:47:16 of the (1,3)-, (1,4)- and (1,6)-linked regioisomers with a tiny amount (2%) of trisaccharides.

Similar distribution of regioisomeric disaccharides has also been observed in transglycosylation catalysed by the same enzyme using methyl β-D-glucopyranoside as the acceptor. However, use of ethylthio β-D-glucopyranoside as the acceptor under similar conditions did not afford any (1,3)-linked disaccharide at all. Hydrogen bonding interactions between the glycosidic atom (O/N vs S) and the acceptor binding site of the enzyme appear to be important for the formation of (1,3)-linked disaccharide.

2.2.5 β-Glucosidase Catalysed Reactions

O’Connor and Imperiai performed NOE spectroscopic study of a series of glycopeptides based on the hemagglutinin (a glycoprotein) from

![Scheme 7](image-url)
influenza virus has demonstrated the role of the N-acetyl groups in facilitating the formation of the compact β-turn conformation for the peptide. These authors attribute this effect to the steric interactions of the saccharide, predominantly by those of the C-2 acetamido group of the GlcNAc of the glycan chain bringing about such dramatic ramifications of the glycopeptide conformation. Replacement of chitobiosyl portion containing C2 and C2’ N-acetyl groups by celllobiosyl moiety having the corresponding hydroxyl groups resulted in a less ordered conformation of the conjugated peptide. Having developed enzymatic routes for assembling (1,4)- and (1,6)-linked GlcNAc-GlcNAc disaccharide mimetics, it was of interest to synthesise differently linked Glc-Glc analogs. Hence, transglucosylation was explored using p-nitrophenyl β-D-glucopyranoside as the donor and GlcβNHAc (3) as the acceptor under two different conditions employing almond β-glucosidase and cloned Pichia etchellsii β-glucosidase II in each case.12b Though the yield of disaccharides obtained from these two reactions is rather low (3 % in each case), the almond β-glucosidase catalysed reaction gave the (1,6)-linked disaccharide as the major product while the other afforded predominantly the (1,3)-linked regioisomer (Scheme 9). Efforts to synthesise Glcβ-GlcNAcβNHAc and Glcβ-ManβNHAc analogs by transglucosylation involving GlcNAcβNHAc (2) and ManβNHAc (8) as acceptors, respectively, in two separate almond β-glucosidase catalysed reactions proved to be of little use thereby pointing out the high acceptor specificity of the enzyme.

3 X-Ray Crystallographic Investigation

As mentioned earlier, crystal structures of only two linkage region model compounds - GlcNAcβAsn11 and GlcβAsn11b - were available in the Cambridge Crystallographic Database prior to the beginning of our work in 1997. Since then we have added over a dozen structures and an equal number of them are currently being investigated. Among the various sugar amides examined, GlcNAcβNHAc (2)9b represents the simplest model of the highly conserved linkage region in N-glycoproteins of all eukaryotes. The benzamido analog of 2, namely GlcNAcβNHAc (4)9b was chosen for investigation in order to understand the possible effects of the aromatic group in the linkage region conformation. GlcβNHAc (3)21 and RhaβNHAc (12)23 are models of GlcβAsn and RhaβAsn, respectively, of the unusual N-glycosidic linkages found in glycoproteins of certain bacteria, while LacβNHAc (13)25 serves as a disaccharide analog of 3. GlcNAcβNHPr (5)22 and GlcβNHPr (6)22, which possess a propionamido aglycan moiety, are analogs of the respective sugar conjugates of glutamine hitherto unknown in Nature. The influence of chlorine, an electronegative substituent, in place of methyl group present in the above propionamido analogs, on the linkage conformation was sought to be explored by examining the crystal structures.
of the chloroacetamido analogs. Compounds 8, 9 and 10 are the acetamido derivatives of Man, Xyl and Gal which are commonly found attached to Ser/Thr in O-glycoproteins. These compounds, 8-10, were included in the structural investigation in an effort to evaluate the conformational preferences of these sugars when they are N-linked.

The single crystals of all these compounds were obtained from aqueous methanol by the slow evaporation method at room temperature. Compounds 2, 4, 5 and 8 were obtained as monohydrates, 13 as a dihydrate while the others in anhydrous form. Their molecular structures are shown in Fig. 4 along with the atom numbering. All the ORTEP drawings are drawn at 50% probability. In compound 4, two molecules (A and B) are present in the unit cell with their aromatic rings stacked in two different orientations in close proximity to each other. The b anomeric configuration of 2-13 is evident from the ORTEP (Fig. 4). As expected, the pyranose ring(s) in all of them adopt $^{1}C_{4}$ conformation excepting that of the L-sugar, 12, which is in $^{1}C_{4}$ conformation.

3.1 Bond Lengths

A list of selected bond lengths of all the compounds (2-13) is given in Table I. The C-C bond lengths in all these compounds are close to 1.54 Å, which is observed in most sugar derivatives. Among the sugar ring C-O bond distances, namely C1-O5 and C5-O5, the latter are generally found to be longer than the former in glycosides and this shortening of the C1-O5 bond has been attributed to a stabilizing effect arising from the delocalization of the glycosidic oxygen lone pair of electrons into the anti-bonding C1-O5 orbital. Such a trend has been noted in the crystal structures of GlcNAcβAsn and GlcβAsn. The same trend is noticed in the crystal structures of 7-13. However, a reversal in these bond lengths is observed in the case of benzamido (4) and propionamido (5 & 6) derivatives (Table I).

The anomeric and the exo-anomeric effects arise as a result of an additional bonding between the anomeric carbon and acetal oxygens of free sugars or O-glycosides through the participation of the unshared pair of electrons of the oxygen when present in certain orientations relative to neighboring bonds. The anomeric carbon is special in that it is bonded to two electronegative atoms – the O5 of the ring and the glycosidic oxygen. This not only results in the preferred axial orientation of the glycosidic group but also manifests itself in the form of shortening of certain bond lengths, angles and torsion angles. The first observation on such bond shortening was reported by Altona et. al., who observed a significant difference in the C-O bond distances in the crystal structure of trans-2,5-dichloro-1,4-dioxane.

The exocyclic bond distance, C6-O6, falls in the range 1.41 Å to 1.42 Å and is in good agreement with the normal C-O bond distance. The bridge C-O bond distance in 13 is slightly greater (1.43 Å) and is comparable to those observed for similar accurately determined disaccharides. The C-N
bond distance of 1.41 to 1.44 Å for all the compounds is comparable to the value of 1.44 Å reported for the linkage region model compound, namely, GlcNAc-Asn. The extent of electron delocalization from the N1 to the carbonyl oxygen, O1', varies slightly among the compounds 2-13 as seen from the values of C1'-N1 and C1'-O1' and this could be attributed to differences in the strengths and patterns of hydrogen bonds involving these two atoms (cf. Section 3.4).

### 3.2 Bond Angles

All the angles involving carbon are close to the tetrahedron value of 109.5° (Table II), whereas those involving nitrogen are in the trigonal range.

The bond angles of all the model compounds are
comparable to those of GlcNAcβAsn. The exocyclic bond angles, namely C4-C5-C6 and O5-C5-C6 have a mean value of 114.0 and 107.3°, respectively, with the latter being consistently smaller than the former. This is attributed to the effect of the lone pair of electrons causing a bend at the oxygen on the O5-C5-C6 bond angle. Such a difference in these bond angles has been observed for GlcNAcβAsn and in several N-linked pyranoses as well.

In general, the valence angle at the ring oxygen atom, namely C5-O5-C1 deviates from the tetrahedral angle by about 1.8° brought about by the presence of the lone pair of electrons. This variation has been attributed to the anomeric effect, which tends to impart a partial trigonal character at O5. This deviation in all the compounds, 2-13 including the Gal residue of lactosyl acetamide [C5B-O5B-C1B = 113.5°(1)], is in the range of 1.8-4° on the higher side. Such deviation is in fact true for all the angles involving the ring oxygen. Thus, the O5-C1-N1 valence angle varies from 105.9 to 108.1° and is consistent with the less than tetrahedral angle (by about 4°) observed in other β-pyranosides. In all the cases, the angle, O5-C1-N1, is also less than the angle C2-C1-N1.

### 3.3 Torsion Angles

As the valence electrons of O5 produce a substantial torsion around C1-O5 and the O5-C5 bonds, the puckering of the ring from a perfect chair conformation (as compared to the value of 55° for cyclohexane) is more pronounced about these bonds than that around C1-C2, C2-C3, C3-C4 and C4-C5 bonds in all the compounds. The notations used in the ensuing discussion for the key torsion angles are depicted in Fig. 5.

#### 3.3.1 Exocyclic Hydroxymethyl Group Conformation

The orientation of the primary alcohol group in the hexopyranoses is of great interest for determining the three-dimensional structures of oligosaccharides and glycoconjugates. The conformational preferences around the C5-C6 bond have been investigated using both experimental and theoretical methods. In general, the exocyclic hydroxymethyl group is known to assume two of the three possible staggered orientations, namely gauche-gauche (gg), gauche-trans (gt) and trans-gauche (tg) (Fig. 6) depending on the orientation of the hydroxyl group at C-4. When the hydroxyl group at C-4 is equatorial as in C1-,D-gluco configuration, gg and gt orientations are favoured. The tg conformer is destabilized due to 1,3-diaxial repulsions between O6 and O4. On the other hand, gt and tg are the preferred orientations when the C-4 hydroxyl group is axial as in C1-,D-galacto configuration.

The torsion angles, C4-C5-C6-O6 and O5-C5-C6-O6 (ω), observed for 2 – 8 and 13 are both close to -60° and 60° respectively (Table III). The values of ω indicate that the primary hydroxyl group of all the Glc and GlcNAc and Man residues in these compounds adopts a gg conformation. This conformational preference is also observed in the solution phase where the J of the H-5 with H-6a and H-6b turn out to be < 1 Hz and ~ 4 Hz respectively, for all these compounds, indicating the gauche and gauche conformation (gg) of the

### Table II

<table>
<thead>
<tr>
<th>Compound Code &amp; No</th>
<th>C4-C5-C6</th>
<th>O5-C5-C6</th>
<th>C5-C6-O6</th>
<th>C5-O5-C1</th>
<th>O5-C1-N1</th>
<th>C2-C1-N1</th>
<th>N1-C1'-O1'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAcβNHAc (2)</td>
<td>114.7(18)</td>
<td>107.3(17)</td>
<td>112.9(18)</td>
<td>113.1(15)</td>
<td>107.5(16)</td>
<td>112.7(16)</td>
<td>122.4(2)</td>
</tr>
<tr>
<td>GlcβNHAc (3)</td>
<td>113.7(4)</td>
<td>108.0(3)</td>
<td>110.2(3)</td>
<td>111.3(2)</td>
<td>108.0(3)</td>
<td>111.4(3)</td>
<td>121.4(3)</td>
</tr>
<tr>
<td>GlcNAcβNHBz (4A)</td>
<td>114.3(1)</td>
<td>107.4(3)</td>
<td>113.3(2)</td>
<td>112.6(3)</td>
<td>107.7(3)</td>
<td>114.0(3)</td>
<td>121.5(3)</td>
</tr>
<tr>
<td>GlcNAcβNHPr (5)</td>
<td>113.9(3)</td>
<td>106.4(4)</td>
<td>115.1(4)</td>
<td>112.8(3)</td>
<td>107.4(3)</td>
<td>111.8(3)</td>
<td>123.7(5)</td>
</tr>
<tr>
<td>GlcβNHPr (6)</td>
<td>113.8(4)</td>
<td>108.7(4)</td>
<td>112.2(4)</td>
<td>111.7(3)</td>
<td>107.0(3)</td>
<td>111.1(3)</td>
<td>123.1(4)</td>
</tr>
<tr>
<td>GlcβNHOCH2-Cl (7)</td>
<td>114.5(3)</td>
<td>108.2(3)</td>
<td>111.5(3)</td>
<td>112.1(3)</td>
<td>106.3(3)</td>
<td>110.5(3)</td>
<td>124.4(4)</td>
</tr>
<tr>
<td>ManβNHAc (8)</td>
<td>113.2(2)</td>
<td>106.5(2)</td>
<td>113.5(2)</td>
<td>113.1(2)</td>
<td>105.9(2)</td>
<td>112.5(2)</td>
<td>122.6(2)</td>
</tr>
<tr>
<td>XylβNHAc (9)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>110.9(16)</td>
<td>107.9(16)</td>
<td>111.8(18)</td>
</tr>
<tr>
<td>GalβNHAc (10)</td>
<td>113.8(3)</td>
<td>106.7(3)</td>
<td>111.7(3)</td>
<td>111.4(2)</td>
<td>107.6(3)</td>
<td>110.9(3)</td>
<td>122.0(3)</td>
</tr>
<tr>
<td>GalβNHOCH2-Cl (11)</td>
<td>114.1(2)</td>
<td>106.9(2)</td>
<td>111.9(2)</td>
<td>112.3(2)</td>
<td>106.7(2)</td>
<td>110.6(2)</td>
<td>124.5(3)</td>
</tr>
<tr>
<td>L-RhaβNHAc (12)</td>
<td>113.7(2)</td>
<td>107.0(2)</td>
<td>-</td>
<td>111.9(1)</td>
<td>108.1(1)</td>
<td>112.1(2)</td>
<td>121.0(2)</td>
</tr>
<tr>
<td>LacβNHAc (13)</td>
<td>113.5(2)</td>
<td>107.0(2)</td>
<td>110.0(2)</td>
<td>111.8(2)</td>
<td>106.9(2)</td>
<td>111.9(2)</td>
<td>122.2(2)</td>
</tr>
<tr>
<td>Mean value</td>
<td>113.9</td>
<td>107.3</td>
<td>112.2</td>
<td>112.1</td>
<td>107.3</td>
<td>111.8</td>
<td>122.4</td>
</tr>
</tbody>
</table>
hydroxyl group (Fig. 6). For the Gal residue in 10 and 11 where the C-4 OH group is in axial position, the primary hydroxyl group assumes a tg conformation with the torsion angle, C4-C5-C6-O6, close to ~180°. These results are in agreement with earlier observations made for carbohydrates in the s2t state. The conformational preferences of the primary hydroxyl group are also influenced by hydrogen bonding (cf. Section 3.4).

3.3.2 Linkage region conformation

Considering the hindered rotation about the amide bond (N1-C1') and the free rotation possible about the N-glycosidic linkage (C1-N1), the amido aglycon moiety can, in principle, choose any one of the four orientations viz., Z-syn, Z-anti, E-syn and E-anti (Fig. 7). The Z geometry of the amide linkage (N1-C1') in the simplest model, GlcNacβNHAc (2), as well as in all the other compounds is evident from the torsion angle ψn being close to 180° (Table III). This structural feature of the amide linkage has been observed in GlcNacβAsn. A statistical analysis of 44 different glycosylation sites belonging to 26 glycoproteins has also revealed a narrow distribution of ψn around 180°. There have been only two exceptions to this.

### Table III

<table>
<thead>
<tr>
<th>Compound</th>
<th>C4-C5-C6-O6</th>
<th>O5-C5-C6-O6</th>
<th>H1-C1-N1-NH</th>
<th>C1-N1-C1'-'C2'</th>
<th>C1-N1-C1'-'O1'</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNacβNHAc (2)</td>
<td>55.4(2)</td>
<td>-66.4(2)</td>
<td>-161.9</td>
<td>174.2(2)</td>
<td>-6.2(4)</td>
</tr>
<tr>
<td>GlcβNHAc (3)</td>
<td>50.1(4)</td>
<td>-71.8(3)</td>
<td>-154.8</td>
<td>-179.2(3)</td>
<td>0.6(5)</td>
</tr>
<tr>
<td>GlcNacβNHBz (4A)</td>
<td>55.3(5)</td>
<td>-68.2(3)</td>
<td>-166.5(4)</td>
<td>177.3(2)</td>
<td>-3.9(4)</td>
</tr>
<tr>
<td>GlcNacβNHPr (5)</td>
<td>54.5(5)</td>
<td>-66.8(4)</td>
<td>-151.8(4)</td>
<td>172.5(4)</td>
<td>-6.5(7)</td>
</tr>
<tr>
<td>GlcβNHPr (6)</td>
<td>53.6(5)</td>
<td>-68.6(5)</td>
<td>-150.3(4)</td>
<td>166.5(5)</td>
<td>-11.4(8)</td>
</tr>
<tr>
<td>GlcβNHOCHCl (7)</td>
<td>51.9(5)</td>
<td>-69.0(4)</td>
<td>-157.1</td>
<td>169.9(4)</td>
<td>-8.1(7)</td>
</tr>
<tr>
<td>ManβNHAc (8)</td>
<td>67.0(2)</td>
<td>-55.5(2)</td>
<td>177.3</td>
<td>171.3(2)</td>
<td>-8.4(3)</td>
</tr>
<tr>
<td>XylβNHAc (9)</td>
<td>-</td>
<td>-</td>
<td>-177.1</td>
<td>-177.8(3)</td>
<td>1.2(4)</td>
</tr>
<tr>
<td>GalβNHAc (10)</td>
<td>-177.1(2)</td>
<td>60.5(3)</td>
<td>-171.0</td>
<td>-178.5(3)</td>
<td>-0.5(7)</td>
</tr>
<tr>
<td>GalβNHOCHCl (11)</td>
<td>-175.8(2)</td>
<td>61.7(3)</td>
<td>174.4</td>
<td>176.6(3)</td>
<td>-4.7(5)</td>
</tr>
<tr>
<td>L-RhaNHOHAc (12)</td>
<td>-</td>
<td>-</td>
<td>165.9(1)</td>
<td>177.2(2)</td>
<td>-0.9(3)</td>
</tr>
<tr>
<td>LacβNHAc (13)</td>
<td>62.4(3)</td>
<td>-59.3(2)</td>
<td>-162.5(2)</td>
<td>175.1(2)</td>
<td>-7.6(4)</td>
</tr>
</tbody>
</table>

* For compounds 2, 4A and 5, the values of torsion angle, C2-N2-C1'-'C2', are -179.3°, 175.2° & -179.1° respectively, and those of torsion angle, H2-C2-N2-HN2, are -175.7°, -174.7° & -172.4° respectively. The interglycosidic torsion angles, O5B-C1B-O4A-C4A & C1B-O4A-C4A-CSA, observed for Compound 13 are 152.06(19) & -157.84(18).

Fig. 6 Newman projections of the staggered conformations of segment of D-glucose involving exocyclic hydroxymethyl group in viewed along C5-C6 bond.
reported in glycoprotein literature. An E geometry for the amide linkage has been observed in the crystal structures of phaseolin and the glycosylated IgG. As a result, the glycan chain in both these glycoproteins does not protrude out, as is the usual case, but rather is trapped in a pocket of the protein surface resulting in long-range carbohydrate-protein interactions. Similar analysis, reported recently, of a larger database of 1678 N-glycan structures present in 506 glycoproteins showed that only 21 (1.25%) of them have amide bond in E geometry with \( \psi_N \) varying between -30° to +30°. It has been recently proposed that the nascent glycopeptides, assembled by oligosaccharyl transferase, contain a metastable (E) amide linkage which is equilibrated to the stable Z form in the active site before release thus possibly preventing product inhibition. In view of this, the N-glycan structures observed in the database with E geometry of the amide bond may represent non-equilibrated forms that may serve their biological role through intramolecular carbohydrate-protein interactions.

Turning the attention on the N-glycosidic torsion, the conformation about the C1-N1 bond is anti in all the compounds as seen from the values of torsion angle, H1-C1-N1-N1H, which is greater than 150° (Table III). The anti conformation of the glycosidic linkage has been observed in various Asn-linked glycopeptides in solution too. From a vacuum ultraviolet circular dichroism study of a series of glycopeptides, Bush et al. have concluded that the linkage region is rigidly fixed in a conformation having the amide protons trans to the anomic proton. Studies based on the 3 J coupling data for asparaginyl oligosaccharides and NOE data obtained for a 22 amino acid IgM glycopeptide and NOE studies on hen ovomucoid glycopeptide all indicate the anti orientation of the N-glycosidic linkage. Given the similarity of the C1-N1-C1-O1’ atomic sequence of the N-linkage to that of Cα-N-C=O of the peptide bond, the preference of the former for planarity and a value of torsion angle close to 0° are worth mentioning. Furthermore, the C-2 acetamido group in the GlcNAc derivatives (2, 4 and 5) (Table III) also adopts a Z-anti conformation as is observed in the case of GlcNAcβAsn (1).

The linkage region torsion angle \( \phi_\alpha \), O5-C1-N1-C1’, is determined with better accuracy than that of H1-C1-N1-N1H. The values of this angle for all the compounds investigated are listed in Table IV.

### Table IV

<table>
<thead>
<tr>
<th>Compound</th>
<th>( \phi_\alpha )</th>
<th>Compound</th>
<th>( \phi_\alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNHPr(6)</td>
<td>-89.5(5)</td>
<td>ManNHAc (8)</td>
<td>-114.5(2)</td>
</tr>
<tr>
<td>GlcNAcNHP(5)</td>
<td>-91.0(4)</td>
<td>GlcNHCOCHCl (7)</td>
<td>-93.9 (4)</td>
</tr>
<tr>
<td>GlcNAcNHBe (4)</td>
<td>-97.0(4)</td>
<td>L-RhaNHAc (12)</td>
<td>101.3 (2)</td>
</tr>
<tr>
<td>LacNHAc (13)</td>
<td>-100.9(2)</td>
<td>GalNHCOCHCl (11)</td>
<td>-111.5 (3)</td>
</tr>
<tr>
<td>GalNHAc (10)</td>
<td>-105.2(3)</td>
<td>GlcNHAc (3)</td>
<td>-93.8(2)</td>
</tr>
<tr>
<td>XylNHAc (9)</td>
<td>-121.7(2)</td>
<td>GlcNAcNHAc (2)</td>
<td>-89.8(5)</td>
</tr>
</tbody>
</table>
As the structure of saccharide or the aglycon moiety is varied, the torsion angle $\phi_N$ sweeps a range of values deviating from that of GlcNacNHAc (2), by as much as 31.9° (Fig. 8). Among the compounds derived from GlcNac and Glc, the propionamido analogs 5 and 6 have comparable $\phi_N$ values which are also close to those of the acetamido (2 and 3), benzamidio (4) and chloroacetamido (7) derivatives. It is worth recalling here that GlcNacNHPr (5) and GlcβNHPr (6) are models of GlcNacβGln and GlcNacβGln, which have never been observed in the linkage region of N-glycoproteins. The $\phi_N$ values for the acetamido of L-rhamnose (12) as well as the disaccharide analog 13 [Galβ(1,4)GlcβNHAc] deviate from that of GlcNacNHAc (2) by more than 11°. There is larger deviation seen in the case of acetamido analogs derived from Gal, Man and Xyl as well as the chloroacetamido analog of Gal (11). The observed torsion angles (Table III and Table IV) indeed correspond to a near anti conformation about the N-glycosidic bond in these compounds. Considering that Gal, Man and Xyl are commonly found attached through an $\alpha$-glycosidic linkage to the hydroxyl group of Ser or Thr in the linkage region of O-glycoproteins, these results are very interesting.

Lastly, the variation in torsion angle $\chi_3$ (N1-C1'-C2'-X) among the propionamido (X = CH$_3$) and chloroacetamido (X = Cl) compounds was examined. The $\chi_3$ value for the GlcNac analog 5 was found to be 172.3°(3) as compared to the value of -172.2° reported$^{14}$ for GlcNacβAsn.3H$_2$O. While such a variation in $\chi_3$ values is also noted for the N-glycan chains in the crystal structures of glycoproteins, the width of variation is much smaller for the glycosylated Asn residues as compared to the unglycosylated ones.$^{32,35}$ In the case of the propionamido derivative of Glc, 6, the $\chi_3$ value turned out to be 114.7°(7) differing by about 51.8° from the value of -166.5° reported$^{19}$ for GlcβAsn.3H$_2$O and reveals the near gauche conformation about C1'-C2' bond as compared to an anti conformation adopted in 5. Replacement of CH$_3$ group in 6 by Cl results in a similar conformation (Fig. 9). Interestingly, in the crystal structure of the simplest non-sugar amide model, N-methylchloroacetamide, N and Cl are found to be in syn orientation. The $\chi_3$ values of 5 and 6 were inadvertently reported in our earlier publication as 114.7°(7) and 172.3°(3), respectively.

The above findings point out that the variations in the sugar structure significantly alter the N-glycosidic torsion ($\phi_N$). In the absence of such differences (as in the case of propionamide and acetamide derivatives), variation in aglycon structure has not resulted in any significant changes in $\phi_N$. In an effort to rationalize the variations in the linkage region conformation, a detailed examination of the molecular packing of 2-13 was undertaken.

### 3.4 Molecular Packing and Hydrogen Bonding

The structure and functions of proteins, nucleic acids and carbohydrates are greatly influenced by hydrogen bonding. Generally, these classes of biomolecules are involved in an intense network of intermolecular hydrogen bonding in aqueous environment. In proteins, there is approximately one hydrogen bonding group, namely the amide, present for every four carbon atoms while in nucleic acids it is about one for every two carbons. In contrast, the proportion is about one for every carbon atom in carbohydrates where the OH or O is the hydrogen-bonding component.$^{9}$ Not surprisingly, associated with water, sugars form the most extensive chains of hydrogen bonding network. Hydrogen bonding patterns observed in diverse carbohydrate crystal structures have been reviewed by Jeffrey.$^{9}$ Modeling of the hydrogen bonding in sugars is more difficult than for the other classes of compounds owing to their inherent orientational freedom of the C-OH groups.$^{40}$ Single crystal X-ray analysis allows one to study the daedal details of hydrogen bonding. The information thus gained would not only lead to a better understanding of the molecular recognition involved in the assembly of carbohydrates but also provide guidelines for their molecular modeling.
More than 95% of the carbohydrates crystallize in non centro-symmetric structures. Analysis of the carbohydrate structural database for the space groups in crystal structures shows that the four of them viz., P2₁2₁2₁, P2₁, P, and C₂, together account for more than 80% of the observed space groups. For the twelve compounds reported from our laboratory, the space groups are distributed among the first two, P₂₁2₁₂ (3, 6-11) and P₂₁ (2, 5 and 12) with the exception of GlcNAcβNHbz (4) and LacβNHAc (13), both of which crystallized in the P₃ system. The hydrogen bonding patterns observed (Table V) are elaborated in the following section.

3.4.1 Finite Chains

In all the compounds, N₁ and O₁' are connected through a finite chain of hydrogen bonds. The characteristic feature of GlcNAc derivatives (2, 4 and 5) is the direct hydrogen bonding between N₁ and O₁' accompanied by a similar one involving N₂ and O₁" leading to a double-pillared molecular packing along the crystallographic a axis as shown in Fig. 10. On the other hand, all the Glc (3, 6 and 7) derivatives exhibit a finite chain of hydrogen bonds that begins with N₁, passes through O₆, O₂ and O₃ of symmetry related molecules and ends in O₁' (Fig. 11). There is also a direct hydrogen bond between O₄-H and O₅. The Gal derivatives, 10 & 11, differ from those (3 & 7) of Glc in terms of altered sequence of oxygen atoms of symmetry related molecules connecting N₁ with O₁' as well as by the absence of the direct hydrogen bond between O₄-H and O₅.

Interestingly, the dihydrate of lactosyl derivative, 13 [Galβ(1,4)GlcβNHAc], exhibits the direct hydrogen bonding between N₁ and O₁' not seen in the acetamido derivative of Glc (3). The second finite chain starts from O₄B and ends at O₁'. Thus, O₁' is involved in a bifurcated acceptor type of hydrogen bond (Fig. 12). The third finite chain also starts from O₄B and ends at O₅B passing through the two water molecules and O₆, O₂ and O₃ of symmetry related molecules. In β-(1,4)-linked disaccharides, if the molecule were to exist in a fully extended conformation the angles O₅B-C₁₁B-O₄A-C₄A and C₁₁B-O₄A-C₄A-C₃A should be close to -110 and +110°, respectively. However, in order to accommodate the intramolecular hydrogen bond between O₃A and O₅B the disaccharide 13 undergoes a symmetrical twist about the oxygen
Fig. 10  Finite chain of hydrogen bonds observed in GlcNAcβNHaC.1H₂O (2)

Table V
Hydrogen Bonding Patterns Observed for Compounds 2 – 13

<table>
<thead>
<tr>
<th>Compound Name and No.</th>
<th>Finite Chain H-bonding</th>
<th>Infinite Chain H-bonding</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcNAcβNHaC.1H₂O (2)</td>
<td>N1-H1N...O1'</td>
<td>...O1W-H1W...O3-H3O...O6-H6O...</td>
</tr>
<tr>
<td></td>
<td>N2-H2N...O1''</td>
<td>...O1W-H2W...O4-H4O...</td>
</tr>
<tr>
<td>GlcNAcβNHaBz.1H₂O (4A)</td>
<td>N1-H1N...O1'</td>
<td>...O1W-H1W...O3-H3O...O6-H6O...</td>
</tr>
<tr>
<td></td>
<td>N2-H2N...O1''</td>
<td>...O1W-H2W...O4-H4O...</td>
</tr>
<tr>
<td>GlcNAcβNHaP.1H₂O (5)</td>
<td>N1-H1N...O1'</td>
<td>...O1W-H1W...O3-H3O...O6-H6O...</td>
</tr>
<tr>
<td></td>
<td>N2-H2N...O1''</td>
<td>...O1W-H2W...O4-H4O...</td>
</tr>
<tr>
<td>GlcβNHaC (3)</td>
<td>N1-H1N...O6-H6O...O2-H2O...O3-H3O...O1'</td>
<td>No infinite chain</td>
</tr>
<tr>
<td></td>
<td>O4-H4O...O5</td>
<td></td>
</tr>
<tr>
<td>GlcβNHaPr (6)</td>
<td>N1-H1N...O6-H6O...O2-H2O...O3-H3O...O1'</td>
<td>No infinite chain</td>
</tr>
<tr>
<td></td>
<td>O4-H4O...O5</td>
<td></td>
</tr>
<tr>
<td>GlcβNHaCCH₃Cl (7)</td>
<td>N1-H1N...O6-H6O...O2-H2O...O3-H3O...O1'</td>
<td>No infinite chain</td>
</tr>
<tr>
<td></td>
<td>O4-H4O...O5</td>
<td></td>
</tr>
<tr>
<td>GalβNHaC (10)</td>
<td>N1-H1N...O3-H3O...O6-H6O...O2-H2O...O1'</td>
<td>No infinite chain</td>
</tr>
<tr>
<td>GalβNHaCCH₃ (11)</td>
<td>N1-H1N...O3-H3O...O6-H6O...O2-H2O...O1'</td>
<td>No infinite chain</td>
</tr>
<tr>
<td>LacβNHaC.2H₂O (13)</td>
<td>N1-H1N...O1'</td>
<td>...O3B-H3OB...O6B-H6OB...</td>
</tr>
<tr>
<td></td>
<td>O4B-H4OB...O2W-H1W2...O1W-H1W1...O1'</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O4B-H4OB...O2W-H2W2...O1W-H2W1...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O6A-H6OA...O2A-H2OA...O2B-H2OB...</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O3A-H3OA...O5B</td>
<td></td>
</tr>
<tr>
<td>ManβNHaC.1H₂O (8)</td>
<td>N1-H1N...O2-H2O...O3-H3O...O6-H6O...O1'</td>
<td>...O1W-H2W...O4-H4O...</td>
</tr>
<tr>
<td></td>
<td>O1W-H1W...O1'</td>
<td></td>
</tr>
<tr>
<td>XylβNHaC (9)</td>
<td>N1-H1N...O2-H2O...O4-H4O...O3-H3O...O1'</td>
<td>No infinite chain</td>
</tr>
<tr>
<td>L-Rhaβ (12)</td>
<td>N1-H1N...O3-H3O...O1'</td>
<td>No infinite chain</td>
</tr>
<tr>
<td></td>
<td>O2-H2O...O5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>O4-H4O...O3</td>
<td></td>
</tr>
</tbody>
</table>
bridge between the residues with the two torsion angles being -89.3° (2) and +81.3° (3) respectively. This is a commonly observed phenomenon in most β(1,4) linked disaccharides. One (O2W) of the water molecules is tri-coordinated while the other (O1W) is tetra-coordinated (Fig. 9). Four-coordination is more common than three-coordination in the carbohydrate hydrates.

The monohydrate of Man derivative (8) also shows a bifurcated type of hydrogen bond involving O1', H6O and O1W-H2W (Fig. 13). As compared to GlcβNHAc (3) and GalβNHAc (10), the finite chain starting from N1 and ending at O1' differs in the sequence interconnecting oxygen atoms of symmetry related molecules (Fig. 13). Yet another variation of the same finite chain is seen in the xylosyl derivative 9, while in case of L-Rha (12) derivative a single symmetry related oxygen atom is seen to connect N1 with O1' (Table V). Two more finite chains, O2-H2O...O5 and O4-H4O...O3 are also observed for the latter compound.

3.4.2 Infinite Chains

These are observed in all the hydrated compounds namely 2, 4, 5, 8 and 13. The packing of all the three GlcNHAc derivatives (2, 3 and 5) is stabilized by the same set of two infinite chains of hydrogen bonds. The first one involving the water molecule and the O3 and O6, propagates along the b-axis in the form of a helix as shown in the Fig. 14. In the second, the water molecule, acting as a bifurcated donor, forms hydrogen bonds with its neighboring O4 facilitating the formation of a
homodromic cycle of hydrogen bonded network (Fig. 15). Such hydrogen bonds are common in cyclodextrins where they have been classified as homodromic, heterodromic and antidromic of which homodromic is more stable.\textsuperscript{39} Similar homodromic cycle is also seen in the mannosyl derivative, 8, involving ...O1W-H2W...O4-H4O... the difference being that the water molecule is also acting as a donor to O1' in a bifurcated fashion. In contrast, the Gal residue of the disaccharide LacβNHAc.2H$_2$O (13) exhibits a zig-zag type infinite chain propagating along the β-axis (Table V).

3.4.3 Aromatic π-π Stacking

Intermolecular interactions involving aromatic rings are key events in many biological recognition phenomena.\textsuperscript{42} In compound 4 (GlcNAcβNHbz. H$_2$O), two independent molecules present in the asymmetric unit show a remarkable difference in the orientation of the phenyl ring as seen in
Fig. 16. The two phenyl rings are co-planar within ~5° and the closest separation between them [C2' (molecule A) and C4' (molecule B)] is 3.29 Å. This distance is comparable to the value of 3.35 Å observed in graphite indicating the strong π-π interaction between the aromatic rings of the molecules A and B of 4. This hydrophobic interaction provides an additional mode of stabilization complementing the hydrogen bonding for the molecular packing.

4 Conclusions

Our efforts aimed at elucidating the structure and conformation of the linkage region of glycoproteins have led to the development of (a) a chemoenzymatic approach to the synthesis of N-glycoprotein linkage regions models and analogs and (b) their detailed structural investigation by X-ray diffractometry. β-1-N-Amido-D-glyco-pyranoses, which are structural mimics of GlcNAcβ-Asn, have been successfully utilized as acceptors in transglycosylation catalysed by three different glycosidases, viz., β-N-acetylhexo-saminidase, β-galactosidase and β-glucosidase. Several regioisomeric disaccharide models and analogs of the linkage region have been synthesized in reasonable yields. Their structural design also permits their ready preparation in crystalline form suitable for single crystal X-ray diffractometry.

The structural investigation of the model compounds prepared was undertaken to understand the effect of structural variations in the linkage sugar as well as its aglycon moiety on the linkage region conformation. The primary focus of the conformational analysis was on the glycosidic torsion $\phi_{\text{N}}$ (O5-C1-N1-C1') which represents one of the most important parameters both in the presentation of recognition determinant saccharides on the cell surface and also in the folding of nascent polypeptide chains. Among the different amido sugars examined, $\phi_{\text{N}}$ deviates by as much as 24.6 and 31.1° for the Xyl and Man analogs, respectively, as compared to that of GlcNAcβNHAc revealing that the N-glycosidic torsion is influenced by basic structural
features of the sugar. These are important findings considering the fact that Man and Xyl are the linkage sugars commonly found in O-glycoproteins. On the other hand, the acetamide and the propionamide derivatives of Glc have an N-glycosidic torsion which is very comparable to that of the model compound, GlcNAcβNHAc, implying that Glc could as well serve the role of GlcNAc at the linkage region.

There is also a good agreement between the ϕ_N value of the propionamide derivative GlcNAc and that of the corresponding acetamide derivative. These results indicate that the changes in the aglycon moiety have a minimal effect on the N-glycosidic torsion suggesting GlcN could serve as well as Asn as the linkage amino acid. The extension of the Asn side chain by a CH_2 group (i.e. replacing Asn by Gln), however, would result in an additional rotational freedom about the C-C bond which in turn would lead to loss of rigidity of the linkage region. Given the major functions of the N-linked glycans, our detailed structural investigation of models and analogs has led us to propose that the conserved linkage region of the eukaryotic N-glycoproteins is characterized by three essential structural features: rigidity, planarity and linearity (Fig. 17).

The choice of an amide linkage with a Z-anti conformation ensures the planarity of the linkage region, whereas the rigidity is provided by Asn having a minimal rotational freedom about its side chain. Finally, the linearity that begins with the β-anomeric orientation of GlcNAc residue is continued by the chitobiosyl moiety and further by the β(1,4) linked mannosyl residue. Meticulous analysis of the molecular packing of the various models and analogs has unraveled exquisite arrays of hydrogen bonds. A systematic investigation on the molecular recognition involved in the crystal packing also on the influence of the underlying non-covalent interactions on the linkage region conformation is warranted. Efforts are currently underway to further exploit our chemoenzymatic approach for the synthesis of oligosaccharide as well as glycopeptide models and analogs of the N-glycoprotein linkage region and to subject them to a comprehensive study involving crystallographic analysis and molecular modeling.

Acknowledgement

Funding provided by the Department of Science and Technology, New Delhi as well as the Indo-French Centre for Promotion of Advanced Research (IFCPAR), New Delhi is gratefully acknowledged. The authors are thankful to their co-workers and collaborators whose names appear in the references. One of the authors (UA) thanks Council of Scientific and Industrial Research (CSIR) New Delhi for the award of Senior Research Fellowship. The authors also thank Cambridge Crystallographic Data Centre (CCDC), United Kingdom for making the program Mercury 1.2.1. available for use.

References

2 N Sharon and H Lis Chem & Eng News March 30 (1981) 21
3 A Varki Glycobiology 3 (1993) 97
4 R A Dwek Chem Rev 96 (1996) 683
5 D A Cumming Devel Biol Standard Karger Basel 76 (1992) 83
7 R D Marshal Ann Rev Biochem 41 (1972) 673
9 (a) F Weiland, R Heitzer and W Schaefer Proc Natl Acad Sci USA 80 (1983) 5470; (b) R Schreiner, E Schanbel and F


13 H S Isbell and H L Frush J Org Chem 23 (1958) 1309; (b) L M Likhosherstov, V A Novikova, V A Derevitskaya and N A Kochetkov Carbohydr Res 146 (1986) C1


18 K Ajisaka Methods Carbohydr Chem X (1994) pp 305


20 D Sriram, T Lakshmanan and D Loganathan and S Srinivasan Carbohydr Res 309 (1998) 227


22 T Lakshmanan, K Priya, D Sriram and D Loganathan Biochem Biophys Res Commun 312 (2003) 405


26 R U Lemieux, S Koto and D Voisin ACS Symposium Series (Eds D Horton and W Szarek) 87 (1979) pp 17

27 C Altona, C Knobler and C Romers Acta Crystallogr 16 (1963) 1217

28 DC Fries, S T Rao and M Sundaralingam Acta Crystallogr B27 (1971) 994


31 R H Marchessault and S Perez Biopolymers 18 (1979) 2369

32 A Imbery and S Perez Protein Eng 8 (1995) 699

33 M C Lawrence, T Izard, M Beuchat, R J Blagrove and P M Colman J Mol Biol 238 (1994) 748

34 J Deisenhofer Biochemistry 20 (1981) 2361


38 J T Davis, S Hirani, C Bartlett and B Ried J Biol Chem 269 (1994) 3331


42 E A Meyer, R K Castellano and F Diederich Angew Chem Int Ed 42 (2003) 1210