Towards the elucidation of the 3D Structure of the N-linked glycoproteins: does rigidity play an important role?

E. C. Stanca-Kaposta, B. Liu, E. J. Cocinero, T. D. Vaden, J. Frey, J. P. Simons and L. C. Snoek
Department of Chemistry, Physical and Theoretical Chemistry Laboratory, University of Oxford, South Parks Road, Oxford, OX1 3QZ, UK

Main contact email address **Lavina@physchem.ox.ac.uk**

Glycosylation is the most common form of post-translational protein modification which results in the formation of glycoproteins. During glycosylation, carbohydrates are attached to the protein either through the amide side chain of asparagine (N-linked to N-acetylglucosamine, see Figure 1), or through the hydroxyl side chain of serine or threonine (O-linked to N-acetylgalactosamine). In N-linked glycoproteins the aspargine residue can accept the oligosaccharide *only* if it is part of a sterically accessible Asn-X-Ser or Asn-X-Thr sequence (where X is any amino acid except proline). Although there is an astounding range of possible epimers, anomers and hydroxyl linkage points that could be accessed[1], all *N*-linked glycoproteins incorporate the *same structural pentasaccharide core*, consisting of two *N*-acetylglucosamine (GlcNAc) and three mannose (Man) monosaccharide units, linked via GlcNAc to the amide side chain of asparagine, see figure 1. This observation inevitably leads to the following fascinating question: "Is there a link between a common three-dimensional structure supporting a 'key glycan scaffold' and the biological function of the glycoprotein?" [i] [2]

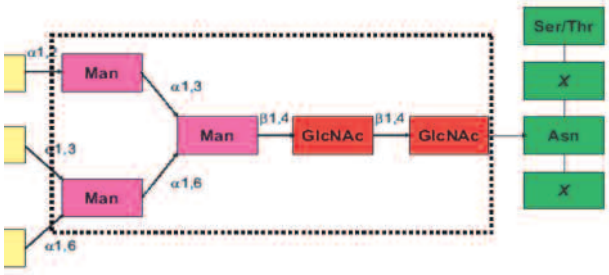

Figure 1. Typical *N*-linked glycoprotein, exhibiting the common pentasaccharide core of *N*-acetyl glucosamine (GlcNAc) and mannose (Man) monosaccharide units attached to a conserved peptide chain. The numbers and α/β notation denote the linkage points and stereochemistry of the glycosidic bonds. The hashed box highlights the core pentasaccharide.

In this report we describe our investigation of the pentasaccharide core in N-linked glycoproteins and of the conserved tripeptide chain to which the sugar attaches. The strategy we apply involves sequencing the system to its building blocks, probing their 3D structure using infrared spectroscopy and high level computations, and then using this information to help assign the larger units. Initially we focus on the monosaccharide units β-N-acetylamino glucopyranoside and β-mannopyranoside. The information obtained from these smaller units is used to elucidate the structures of the disaccharides GlcNAc(β1,4)GlcNAc and Man(β1,4)GlcNAc, as well as the branched trisaccharide Man(α1,6)Man(α1,3)Man (see Figure 1) for which calculations are very expensive. In addition, we present recent preliminary investigations of the conserved peptide sequences which are important in glycosylation. For this purpose we selected three systems: capped AsnGlySer, GlnGlySer and AsnProSer.

The results described in this report represent a significant extension to the previous studies which focused only on the conformational landscape of the mannose subunits of the conserved sugar core.[3,4] These studies bring us very close to our final aims: the elucidation of the core pentasaccharide structure and understanding the effect of glycosylation on peptides.

The experiments have been carried out in the gas phase, in the low temperature conditions of a molecular beam expansion. All molecules have been synthetically tagged with a UV chromophore at the anomeric carbon position for spectroscopic detection. The conformational landscape is investigated by double resonance infrared ion dip (IRID) and UV-UV spectroscopic techniques. Assignments of the spectroscopic signatures are aided by the knowledge gained from spectroscopy of the smaller units in combination with electronic structure calculations at DFT (B3LYP) and MP2 levels.

Subunits of the pentasaccharide core: mono- and disaccharides

The IRID spectra for the lowest conformers of the mono- and disaccharides described in this report together with the computed vibrational frequencies for the disaccharides are shown in Figure 2. The predicted structures are shown on the right side of the figure. The labeling in the figure is the same as described in our previous studies:[5] the carbon atoms are counted in a clockwise direction starting with the anomeric carbon and the OH groups receive the same number as the carbon atoms to which they are attached. Figure 2c shows the IRID spectrum for the lowest energy conformer of βGlcNAc together with the assignments and structure based on previous calculations.[i] The spectrum exhibits a broad, intense feature, more shifted to lower wavenumber than the other OH bands; an indication of strong hydrogen bonding. From analysis of the calculations this feature can be attributed to the OH3 of the sugar unit which is strongly hydrogen bonded to the C=O of the N-acetyl group, thus making this molecule more inflexible than glucose (Glc).[ii]

[i] Two additional conformers (not shown here) are observed in the experiment.
[ii] Glucose lacks the N-acetylamino group and has instead an OH group.

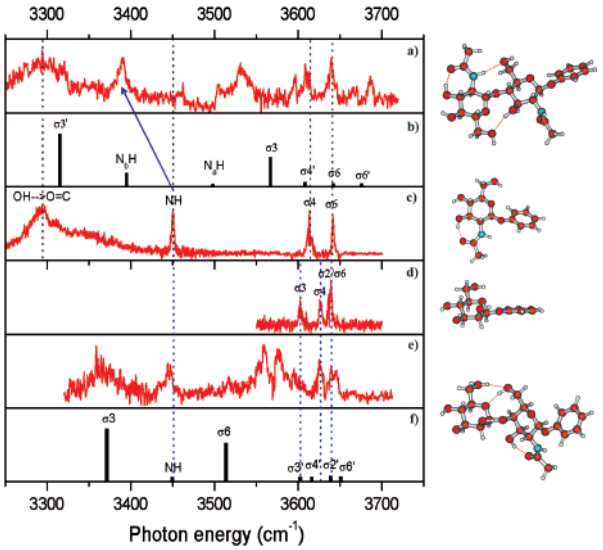

Figure 2. Experimental IRID spectra for a) GlcNAC(β1,4)GlcNAcOPh, c) βGlcNAcOPh, d) βManOPh and e) Man(β1,4)GlcNAcOPh together with assignments based on previous calculations. b) and f) predicted B3LYP/6-31+G* vibrational frequencies for GlcNAc(β1,4)GlcNAcOPh and Man(β1,4)GlcNAcOPh. The OH and NH vibrations were scaled by 0.9734 and 0.96, factors which proved suitable for carbohydrates. Corresponding structures are shown on the right. The labels with (') in the predicted spectra of the dimers refer to the left sugar rings. N_aH and N_bH represent the NH stretching bands belonging to the right and left sugar rings, respectively. The black and blue dotted lines indicate the positions of the experimental peaks in GlcNAc and Man monomers.

If *rigidity* is an important factor for the biological function of glycoproteins, then perhaps this can be the reason why nature chooses GlcNAc instead of Glc as a building block in the core pentasaccharide.

Our knowledge of the structure of the GlcNAc monosaccharide can now help us understand the conformation of the next building block, the disaccharide GlcNAc(β1,4)GlcNAc (diGlcNAc). Analogous features are observed in the spectra of βGlcNAc and GlcNAc(β1,4)GlcNAc [iii], indicated by the black dotted lines in Figure 2. The band at 3450 cm^{-1}, attributed to the NH stretching vibration in GlcNAc is shifted to 3390 cm^{-1} in GlcNAc(β1,4)GlcNAc (see blue arrow in Figure 2), suggesting that the NH group is involved in a hydrogen bond. The similarities and differences between the IRID spectra of GlcNAc and GlcNAc(β1,4)GlcNAc can now be used as a guide to generate starting structures for the latter, also reducing computational time. The final result of the simulations is shown in Figure 2b, together with the predicted structure for the most stable conformer. The agreement between theory and experiment is very good. The predicted structure presents two strong hydrogen bonds across the glycosidic linkage, making it very inflexible. Similar rigidity has been observed in lactose [6], where the glucose and galactose subunits are connected with the same (β1,4) linkage. The sub-units Man(α1,3)Man and Man(α1,6)Man have different linkages which present a higher flexibility especially in the case of the (α1,6) where there is an additional CH_2 spacer.[i] Similar structures for GlcNAc (β1,4) GlcNAc with a "trans" orientation of the N-acetylamino groups, as observed in gas phase, have been found in solution (see Figure 3) as well as in proteins. [7,8] These discoveries are very encouraging, suggesting that the gas phase experiments help offer an insight into the true ligand structure in a biological interaction.

The next sub-unit of the pentasaccharide core to be studied is the disaccharide Man(β1,4)GlcNAc, and its experimental IRID spectrum can be compared with those of its monosaccharide parts. Figures 2c and 2d show the IRID spectra, assignments and structures for the lowest conformers of GlcNAc and Man, respectively. Figures 2e and 2f show the experimental IRID spectrum for Man(β1,4)GlcNAc, and the predicted frequencies and structure that best fit the experiment, respectively. The blue dotted lines indicate similar features in the spectra of

iii A second, minor conformer has been observed which is not discussed here.

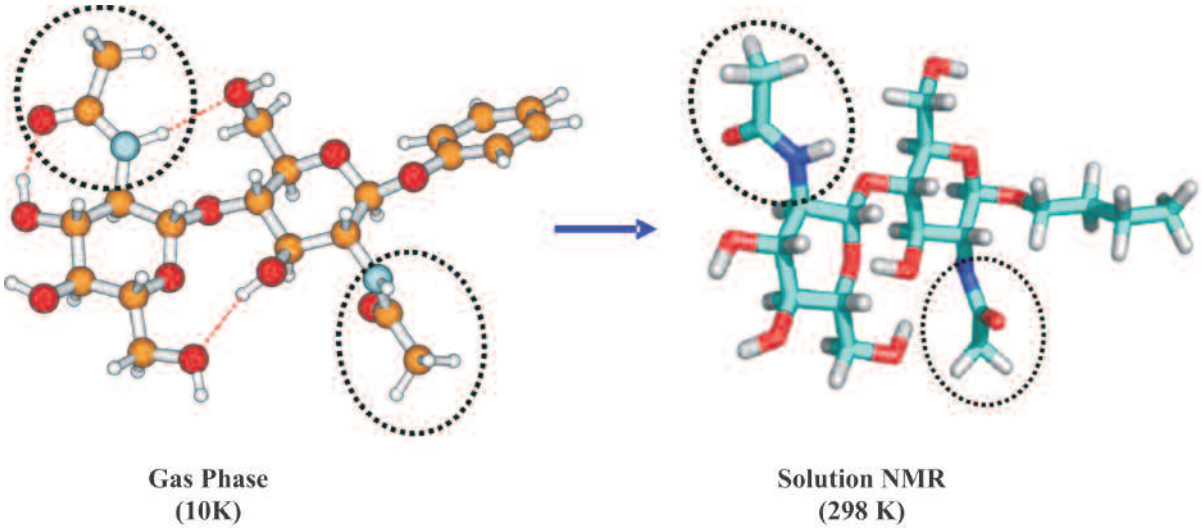

Figure 3. Gas phase and solution phase NMR structures for GlcNAc(β1,4)GlcNAcOPh. Note the 'trans' orientation of the N-acetylamino groups in both cases.

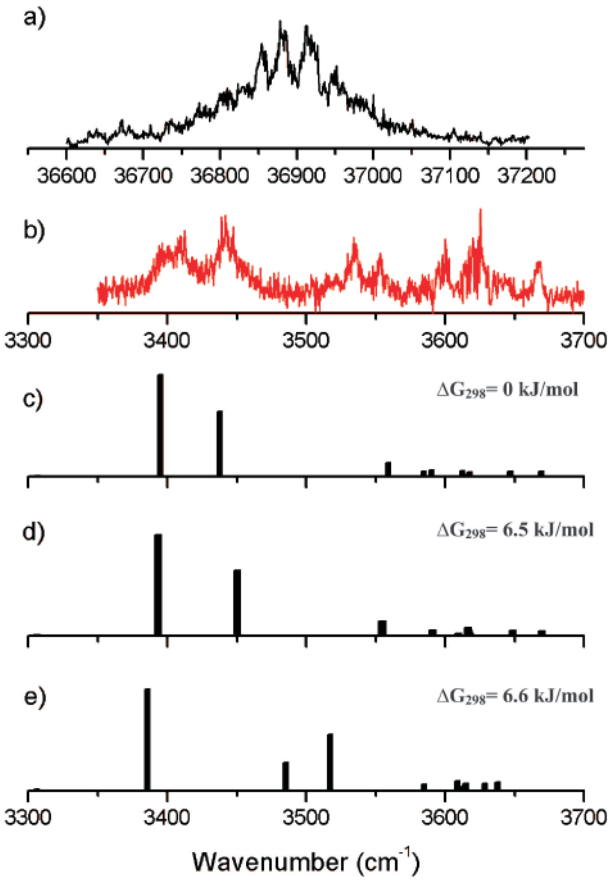

Figure 4. Experimental and theoretical results on Man(α1,3)Man(α1,6)ManOPh. a) REMPI spectrum b) IR depletion spectrum. c)-e) theoretically predicted spectra for the three most stable conformers using the B3LYP/6-31+G*//HF/6-31G ONIOM approach. Hereby the mannose rings were treated using B3LYP/6-31+G*, while the phenyl ring was calculated with help of HF/6-31G. All frequencies are scaled by 0.9734.

the monosaccharides and the disaccharide. Two hydrogen bonds are formed across the glycosidic linkage, but the linkage seems less rigid than in diGlcNAc, resulting in a "cis" orientation of the two sugar components which favors only one side hydrogen bonding across the linkage. Furthermore, the spectra exhibit less-shifted IR bands, an indication of weaker hydrogen bonds.

Subunits of the pentasaccharide core: the tri-mannoside

NMR studies performed on the core-pentasaccharide indicated that the structure of this large molecule is well preserved in solution [9], in spite of the high conformational flexibility of carbohydrates.[ii] These investigations have also suggested that its mobility is inhibited by the presence of water molecules which form hydrogen bonds between the branches of tri-mannose.

Investigation of the tri-mannoside building block in the gas phase, with and without included water, will give further insight into the mechanisms which help stabilise its 3D structure.

The gas phase R2PI and IRID experimental spectra of Man(α1,3)Man(α1,6)Man (TriMan) and the predicted vibrational frequencies for the lowest energy conformers are shown in Figure 4. A relatively good agreement is observed for the lowest energy conformer. Although the second conformer cannot be excluded, its energy is predicted to be 6.5 kJ/mol higher which strongly favors the assignment to the lowest energy conformer. The bands around 3400 cm^{-1} indicate strong hydrogen bonding, which is in agreement with the predicted structure shown in Figure 5 where the arms of the branch are kept close together. Figure 5 also indicates that the structure of Man(α1,3)Man is conserved in one of the arms of the trisaccharide, whereas the more flexible branch folds over to facilitate hydrogen bonding and hence, rigidity. The co-operative hydrogen network incorporates only one OH group in the sugar pocket formed by the two arms of tri-mannose. The observation of this structure raises further questions: Can water break the hydrogen bonds and replace this OH group with one water OH? How does water influence the structure of the (α1,3) arm and of the trisacharide? Answers to these questions can be obtained by studying the infrared spectra of tri-mannose complexed to one and more water molecules. These can be introduced in the experiment in a controlled way through our mass-selective experimental setup, and experiments and calculations are currently underway.

The 3D structures of tripeptides

Preliminary experiments have been carried out to investigate the conformational preference of certain capped tripeptides which play a crucial role in N-linked

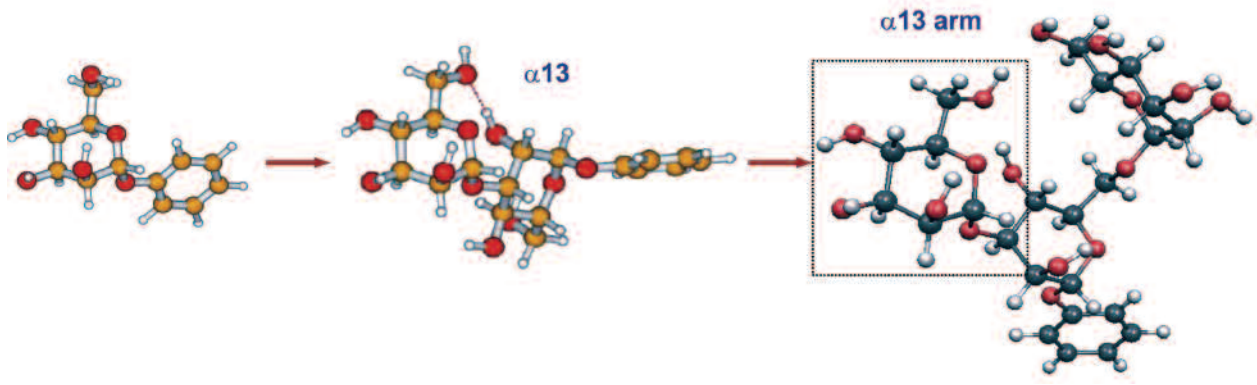

Figure 5. Building up the TriMan structure: from monosaccharide to trisaccharide. The figure highlights the geometry on Man(α1,3)Man which is conserved in one of the arms of TriMan.

glycoproteins. Figure 6 shows the R2PI spectra for two of the capped tripeptides we have studied, GlnGlySer and AsnGlySer. The UV spectra show mainly one peak indicating the presence of only one major conformer for each species. In Figure 6 the structure of capped GlnGlySer from preliminary calculations at B3LYP/6-31G* level is shown as inset. Additional calculations and IRID experiments, which are scheduled in the near future, are essential in order to understand the conformational preference of these systems.

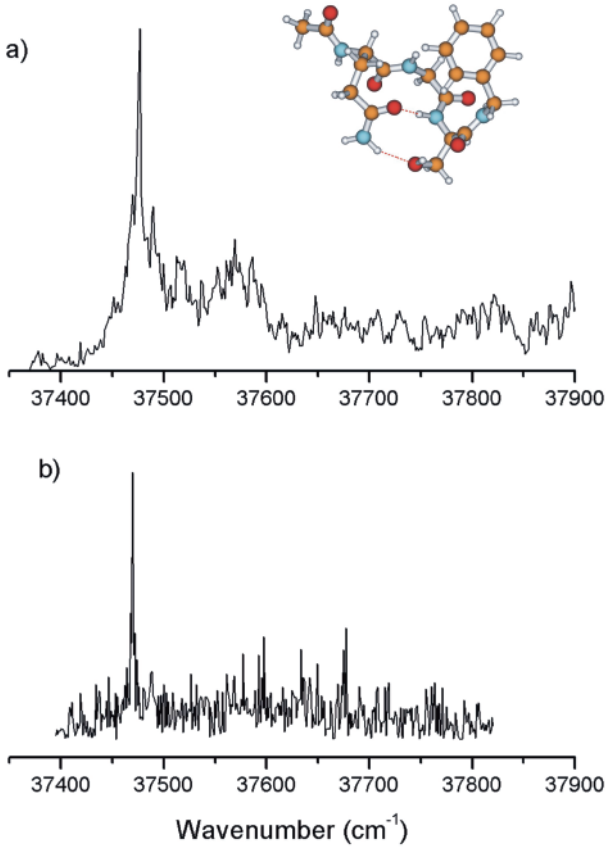

Figure 6. R2PI spectra of capped a) GlnGlySer and b) AsnGlySer. The inset shows the structure for GlnGlySer tripeptide based on preliminary B3LYP/6-31G* calculations.

Conclusions

The studies presented here are an essential progress towards our final aims: understanding the structural preference of the conserved pentasaccharide core in N-linked glycoproteins as well as the effect of glycosylation on the protein itself. The experiments have shown that the (β1,4) linkage, present in two different carbohydrate environments in the core, is very rigid and suggest that rigidity might play an important role in nature's choice for this conserved structure. The rigidity is caused by strong co-operative intramolecular hydrogen bonding across the glycosidic linkage. As the systems grow larger the conformational landscape seems to become more simple, leading to our observation in the gas phase of fewer conformers. Further experiments are planned in the near future to characterize the gas-phase structures of the second block of N-linked glycoproteins: the conserved peptide chain near the glycosylation region.

Acknowledgments

We would like to thank Prof. B. G. Davis and Dr. D. P. Gamblin for providing us with the experimental samples. Support for this research was provided by the EPSRC, the Royal Society (T.D.V., USA Research Fellowship; L.C.S., University Research Fellowship), the Leverhulme Trust (Grant No. F/08788D), the EPSRC Laser Loan Pool and the Physical and Theoretical Chemistry Laboratory at Oxford.

References

1. H.-J., Gabius, *Naturwissenschaften*, **87**, 108 (2000).
2. R.A. Dwek, *Chem. Rev.*, **96**, 683 (1996).
3. P. Çarçabal *et al., J. Am. Chem. Soc.*, **128** (6), 1976 (2006).
4. P. Çarçabal *et al., J. Am. Chem. Soc.*, **127** (23), 11414 (2005).
5. J. P. Simons *et al., Int. Rev. Phys. Chem.*, **24** (3-4), 489 (2005).
6. R. A. Jockusch *et al., J. Am. Chem. Soc.*, **126**, 5709 (2004).
7. F. Yu *et al., Biophysical J.*, **91**, 1952 (2006).
8. F. Fusetti *et al., J. Biol. Chem.*, **278**, 37753 (2003).
9. M. R. Wormald *et al., Chem. Rev.* **102**, 371 (2002).