Pure & Appl. Chem., Vol. 69, No. 9, pp. 1875–1878, 1997.
Printed in Great Britain.
© 1997 IUPAC

Structural glycobiology: How to gain insight into the structure–function relationship

Johannes F.G. Vliegenthart

Bijvoet Center, Department of Bio-Organic Chemistry
Utrecht University, P.O.Box 80.075, NL-3508 TB Utrecht, The Netherlands

Abstract:: Glycosylation of proteins is a widespread feature in nature. Important is the question concerning the function of the carbohydrate chains of proteins in biological systems. The proposed funtions deal on the one hand with the physico-chemical properties, and on the other hand with direct or indirect involvement in biological processes. However, before any biological effect can be achieved, interaction with complementary molecules is a prerequisite. To gain insight into the interaction features it is essential to have a profound knowledge of the 3-dimensional structure of at least that part of the molecule that is involved. Various aspects of the structure are considered with respect to uniqueness.

Glycoproteins consist of a polypeptide backbone, bearing one or more covalently attached carbohydrate chains. Such compounds are widespread in nature and the majority of the secreted and cell-surface associated proteins of higher organisms are glycoproteins (ref. 1, 2). Given the location of the carbohydrate chains at the outer surface of the proteins, they are well suited for biological recognition. This is the more reasonable, because from a relatively small number of monosaccharides a large number of oligomers can be constructed. There is good evidence that such glycans play a role in cell-cell, cell-pathogen and cell-biomolecule interactions.

From a structural point of view, glycoproteins are rather complex, because in such a compound more than one glycosylation site may occur (ref. 3). Furthermore, the glycans at each glycosylation site can exhibit a considerable (micro)heterogeneity. By consequence a glycoprotein is rarely a single component and consists mostly of a family of glycoforms. Glycoproteins are classified as N- or O-linked depending on their linkage to the protein. The N-linked glycans are bound via GlcNAc to the amide nitrogen of Asn, and O-linked glycans of the mucin type are attached via GalNAc to the hydroxyl function of Ser or Thr. In addition there are many other types of O-glycans. The N-linked carbohydrates have in common a trimannosyl -N, N^{I}-diacetylchitobiose core, that can be extended to the i) oligomannose type, ii) the *N*-acetyllactosamine type, iii) the hybrid type, iv) the xylose-containing type. For the O-linked glycans, there is a large diversification with respect to i) the amino acid involved, ii) the connecting monosaccharide and iii) the extension.

Recently, the repertoire of glycans was extended with the discovery of a new type of linkage involving the indole nucleus of Trp. It was found that a α-mannopyranosyl unit was C-linked to C-2 in the indole side chain (ref. 4, 5).

To define the structure of a glycoprotein, the amino acid sequence has to be established and in addition the structural parameters of the glycans have to be determined. The latter comprise i) the primary structure, ii) type and location of the non-carbohydrate substituents, iii) the linkage to the protein, iv) the microheterogeneity at each glycosylation site. For gaining insight into structure-function relations it is important to have determined the 3-dimensional structure in solution, preferably for the whole glycoprotein.

Although a large variation in structure of glycoprotein glycans is found, the number of theoretically possible structures is by no means realized in nature. Apparently, the enzymes involved in the biosynthesis of these glycans are restricted in i) number, ii) donor and acceptor specificity (comprising carbohydrate and protein), iii)

sequence of actions, and iv) competition for the same site. Compartmentalization, leading to a spatial organization of the glycosyltransferases in assembly lines is an important factor in this respect.
Glycans at individual glycosylation sites in the same or in other glycoproteins and having identical primary structures may be different in their effects on the physico-chemical and/or biological properties. This may be due to the local environment leading to different overall structures at these glycodomains. The 3-dimensional structures of the glycans can be different due to stabilization by parts of the protein backbone, differing in amino acid sequence at each glycosylation site. The accessibility of the glycans may be different due to steric factors or charge caused by the protein or by neighbouring glycans. Alternatively, for specific properties of individual glycans it may be essential that in addition to the glycan, part of the protein is recognized. A dual commitment of glycan and protein could be required for a productive interaction. An other aspect in recognition processes may be the masking or demasking of protein epitopes by glycans.
For N-glycans exist interesting indications for variation in the 3-dimensional structure at the level of the first GlcNAc residue. First there are X-ray structural data. Although glycoproteins are notoriously difficult to crystallize, the Brookhaven Protein Databank contains quite some information on such compounds. It shows that the 4 dihedral angles, defining the position of the GlcNAc residue with respect to the C_α of Asn, can vary for individual glycosylation sites. The orientation of GlcNAc towards the protein can be: α-face, β-face, 05-edge or solvent exposed. The local structure in which Asn is located can be: turn, extended, β-strand or end of an α-helix.
NMR spectroscopy has also provided evidence for considerable differences in interaction between this GlcNAc residue and the protein. For example it could be demonstrated that in the isolated α-subunit of human chorionic gonadotropin, GlcNAc at glycosylation site Asn-52 has hardly any interaction with the protein, whereas the residue attached to Asn-78 has numerous contacts with the amino acid residues in the environment (ref. 6). The variation in strength of the interaction between the first GlcNAc residue and the protein has consequences for the mobility and flexibility of the glycans and thereby for the exploration of the conformational space. It should be realized that several factors contribute to the glycan flexibility comprising i) motional disorder of the atoms within the constituting monosaccharides, ii) conformational alterations, and eventually transitions within the ring structures, iii) torsional oscillations, iv) conformational transitions around the glycosidic bonds. Each process takes place at its own timescale. Flexibility in one residue of a glycan is transmitted through the chain. In fact the flexibilities are additive ultimately leading to a larger flexibility at the peripheral parts. As an outcome, it can be concluded that glycans with the same primary structure behave differently when attached to different glycosylation sites. This makes it reasonable that non-unique glycan structures are involved in unique activities, thanks to an interactive influence of the local environment in the glycoprotein.

The search for primary stuctures that are unique for a specific glycoprotein with a specific activity, usually ends in a disappointing result. In most of the cases, sooner or later, the same structure will be found in other glycoproteins. A nice example is formed by the glycoprotein hormone lutropin (LH), produced in the anterior lobe of the pituitary. As demonstrated (ref. 7), the Asn-linked diantennary glycans in LH have one or two branches terminated with the sequence SO_3-4GalNAcβ1-4GlcNAcβ1-2Manα1-R. Although claimed to be exclusive, we found that this element occurs also in Tamm-Horsfall glycoprotein (ref. 8). Later on we established that in urokinase many structures occur that end in this element as shown in Fig. 1 (ref. 9). In the meantime in several other glycoproteins, this structural element could be demonstrated. In relation to the finding of this sequence in LH, Baenziger et al. (ref. 10) have shown that β-GalNAc (1-4) linked is added to the glycan in a

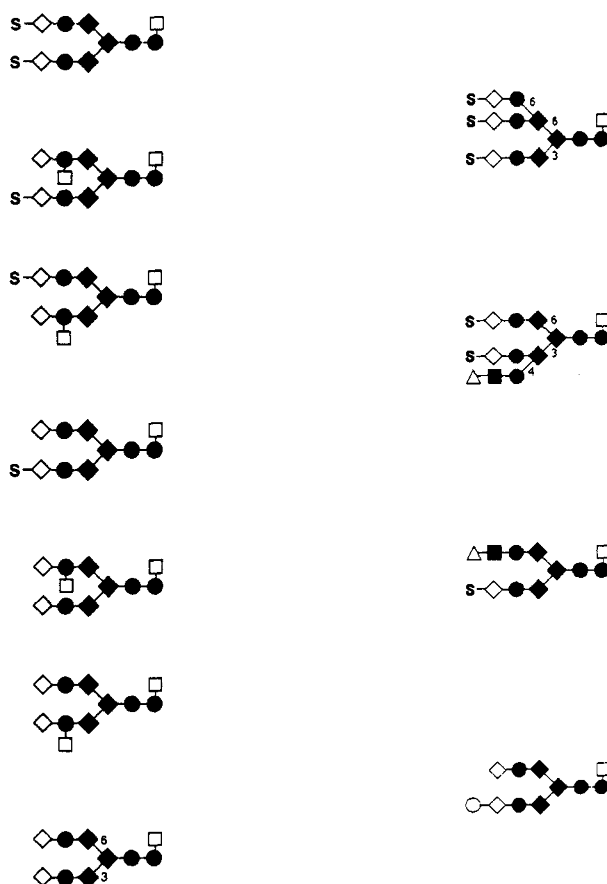

Fig. 1 The major N-linked carbohydrate chains from human urokinase
Compounds are represented by short-hand symbolic notation: (□), L-Fuc; (■), D-Gal; (●), D-GlcNAc; (◆), D-Man; (○), Neu5Acα2-6; (△), Neu5Acα2-3; (◇), D-GalNAc; (S-◇), (SO_4^-)-4GalNAc.

protein specific process. This may not be an exceptional feature, but a good example of the individuality of a protein including its glycans, even in the denatured state.
Another example of a more or less unique structure is the human natural killer cell -1 (HNK-1) carbohydrate epitope SO_3-3GlcNAcβ1-3Galβ1-4GlcNAc. However, this epitope has also been found on myelin associated glycoprotein and on other glycoconjugates in myelin such as P0 and the HNK-1 glycolipids. This HNK-1 carbohydrate has now been recognized as an important mediator of molecular recognition in normal development (ref. 11).

In conclusion, the complete characterization of the 3-dimensional structure of a glycan of a glycoprotein requires that the protein structure is fully taken into account in order to describe the situation at a glycosylation site. The situation may even be more complex in the case other glycosylation sites are nearby. In general the latter aspect will be more relevant for mucin type structures than for N-linked glycans. For gaining insight into the structure-function relationships, above all adequate and specific assays have to be available to probe the function. Very often, however, functions are tested in complex systems that do not permit discrimination according to all structural parameters. To study the effect of the fine structure of the carbohydrate chains on the properties, it is important to have the tools to explore the structure in all its details. Furthermore, it is essential to get the

possibilities to change the structure of a glycodomain either along biosynthetic routes or via chemical and enzymic routes. For the biosynthetic approach, engineering of the protein part of the glycodomain is quite feasible by molecular biology techniques. This may be used to investigate the importance of certain amino acid residues or to create a complete knock-out of a glycosylation sited. Glyco-engineering is much more difficult, and is an important challenge for future research.

ACKNOWLEDGEMENTS

These investigations were supported by the Netherlanda Foundation for Chemical Research (SON) with financial aid from the Netherlands Organization of Scientific Research (NWO). The contributions of Drs. T. de Beer, A.A. Bergwerff, K. Hård, J.P. Kamerling, and C.W.E.M. van Zuylen are gratefully acknowledged.

REFERENCES

1. H. Lis and N. Sharon. *Eur. J. Biochem.* **218**, 1-27 (1993).
2. A. Varki. *Glycobiology* **3**, 97-130 (1993).
3. J.F.G. Vliegenthart and J. Montreuil. In *Glycoproteins, New Comprehensive Biochemistry 29A* (J. Montreuil, J.F.G. Vliegenthart and H. Schachter, eds.), pp. 13-28. Elsevier, Amsterdam (1995).
4. J. Hofsteenge, D.R. Müller, T. de Beer, A. Löffler, W.J. Richter and J.F.G. Vliegenthart. *Biochemistry* **33**, 13524-13530 (1994).
5. T. de Beer, J.F.G. Vliegenthart, A. Löffler and J. Hofsteenge. *Biochemistry* **34**, 11785-11789 (1995).
6. T. de Beer, C.W.E.M. van Zuylen, B.R. Leefland, K. Hård, R. Boelens, R. Kaptein, J.P. Kamerling and J.F.G. Vliegenthart. Submitted for publication (1996).
7. E.D. Green, H. van Halbeek, I. Boime and J.U. Baenziger. *J. Biol. Chem.* **260**, 15623-15630 (1985).
8. K. Hård, G. van Zadelhoff, P. Moonen, J.P. Kamerling and J.F.G. Vliegenthart. *Eur. J. Biochem.* **209**, 895-915 (1992).
9. A.A. Bergwerff, J. van Oostrum, J.P. Kamerling and J.F.G. Vliegenthart. *Eur. J. Biochem.* **228**, 1009-1019 (1995).
10. B.J. Mengeling, S.M. Manzella and J.U. Baenziger. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 502-506 (1995).
11. M. Schachner and R. Martini. *Trends Neurosci.* **18**, 183-191 (1995).