INTRAMOLECULAR CARBOHYDRATE-PROTEIN INTERACTION[1]

Johannes F.G. Vliegenthart

Bijvoet Center, Department Bio-Organic Chemistry, Utrecht University, Utrecht, The Netherlands.

The widespread occurrence of protein glycosylation has stimulated the thorough investigation of the structure and function of this posttranslational protein modification [1,2]. Significant progress has been made with respect to the determination of the primary structures of glycans. Numerous structures were derived with regard to glycans as well as to glycosylation sites in glycoproteins. Although the N and O linked chains are still the most abundant types of protein decoration, a number of rare types have been discovered as well. Even today, it should be realised that the occurrence of multiple glycosylation sites in one protein, in conjunction with (micro)heterogeneity make it extremely difficult to define the complete structure of such a protein. In fact, most of the glycoproteins represent ensembles of glycoforms.

On the important question on the function of glycoprotein glycans, it is impossible to give an unambiguous answer. A large array of glycosylation effects have been described. However, it is not feasible to interpret these effects in terms of unique glycan structures being exclusively responsible for well-defined features.

In general, it is considered that glycosylation effects vary from physico-chemical to biological. Parts of the influence of glycans on the macromolecular behaviour are

[1]The author is much indebted to Prof. Dr. J.P. Kamerling, Drs. J.P.M. Lommerse, J.A. van Kuik, B.R. Leeflang, A.L.J. de Beer, C.W.E.M. Thijssen-van Zuylen, F. Casset, P.A.J. Erbel and Y. Karimi-Nejad, of the Bijvoet Center; and Drs J.Hofsteenge and J. Löffler of the Friedrich Miescher Institute, Basel, Switzerland for fruitful collaboration. The studies in the author's laboratory were supported by the Council for Chemical Sciences of the Netherlands Organisation for Scientific Research and by the European Commission.

The Molecular Immunology of Complex Carbohydrates–2.
Edited by Albert M. Wu, Kluwer Academic/Plenum Publishers, 2001.

obvious, such as molecular mass, solubility, charge, stability and resistance against proteolytic attack. However, some could be subtler and exhibit effects that represent phenomena occurring at the atomic level.

Huge is the diversity in claims for involvement of the glycans in biological processes, ranging from cell-cell, cell-pathogen, and cell-molecule interactions to signal transduction and protein folding. It is remarkable that sometimes in the same glycoprotein identical types of glycan structure are postulated to be responsible for different effects. However, sometimes the opposite is found. Different glycan structures are held responsible for similar kinds of effects. Apparently, the local molecular environment of glycans in a glycoprotein is a decisive factor in determining the functional behaviour. The presentation of glycans on the outside of glycoproteins and in membrane-bound glycoproteins on the membrane surface is crucial for recognition and interaction processes.

To gain insight into the molecular events that control the glycan presentation in space and time, it is necessary to investigate the intramolecular carbohydrate-protein interactions. Some pertinent information in this respect is available from X-ray studies on N-glycoproteins [3]. In a number of cases, details could be obtained for the structure of the core parts of the glycans. It was shown that the Asn-linked GlcNAc could occupy various positions. The α- or β-face of this residue may be oriented towards the protein or the O5 edge. Alternatively, the residue can be solvent exposed. Another variable is the local protein structure in which the Asn is located. So far turns, extended structures, β-strands, and ends of α-helices have been found as locations for N-glycosylation in the protein. These data show clearly that already at the level of the first GlcNAc the presentation of the glycan to the outside world can be completely different. It is reasonable to assume that upon extension of the carbohydrate chain beyond this first residue these differences in local structure will exert their effects until the outer parts.

It is most relevant to see how this works out in solution, because many interactions that take place under physiological conditions occur in an aqueous environment in a solution-like state. Advanced NMR techniques in conjunction with molecular dynamics calculations are the methods of choice to obtain such results. In this report, our studies on a few glycoproteins will be summarised.

I. Pineapple stem bromelain

Pineapple stem bromelain [E.C. 3.4.22.4] is a cysteine proteinase with a broad specificity for the hydrolysis of peptide bonds. It has a molecular mass of 33 kDa and contains one N-glycosylation site. Interestingly, the N-glycan exhibits no heterogeneity and consists of: Manα1-6[Xylβ1-2]Manβ1-4GlcNAcβ1-4[Fucα1-3] GlcNAcβ1-R {R represents peptide or protein}.

To investigate the influence of a folded polypeptide chain on the conformation of the glycan, studies were carried out on the carbohydrate chain in a glycotetrapeptide containing the intact glycan [4], and on the carbohydrate chain in the integral protein [5]. This should allow a comparison of the glycan conformations. Standard NMR measurements were performed providing chemical shift and NOESY data. Theoretical NOESY cross peak intensities were calculated on the basis of MD simulations, by using the CROSREL program [6]. A number of calculation methods were applied, each accounting differently for internal flexibility. The methods use generalised order parameters and/or individual rotational correlation times for every constituting monosaccharide.

For the glycopeptide, wherein R corresponds to Asn-Glu-Ser-Ser, we found that the Xylβ1-2Manβ1-4GlcNAcβ1-4GlcNAcβ part is rather rigid and has mainly one conformation. The Fucα1-3GlcNAc and Manα1-6Man linkages are relatively flexible. It is interesting to note that the characteristic times for conformational transitions of glycosidic linkages were calculated to be in the order of 1 –100 ns.

NMR measurements on the intact glycoprotein could easily be performed if, in view of the autodigestion of the protein, the enzyme was irreversibly inhibited. The NMR spectra contained two sets of signals. One corresponds to a denatured form of the protein, the other to the native form. The data of the protein in its denatured form fit the information obtained for the glycopeptide excellently. The narrow lines for the glycan in the denatured form indicate a considerable flexibility for this part of the molecule.

However, for the native intact form the situation is different from the glycopeptide with regard to the glycan conformation and flexibility. The Fucα1-3GlcNAc linkage has an altered distribution over the two minima ϕ,ψ = -90,-100 and ϕ,ψ = -140,-135. In comparison to the glycopeptide, the former conformation is more abundant in the glycoprotein. An obvious explanation for this observation is an interaction with the

protein. In addition, the behaviour of the Manα1-6Man linkage is different. In the glycoprotein the Man *gt* (ω = 60) rotamer is more abundant.

The mean rotation correlation time of the glycan in the intact glycoprotein is larger than that in the glycopeptide ($\tau_0 \pm 9$ ns versus $\tau_0 \pm 0.7$ns), reflecting the large difference in mass of the protein and the peptide moiety. In the glycoprotein the mobilities in the glycosidic linkages of the Xylβ1-2Manβ1-4GlcNAcβ1-4GlcNAcβ part are rather similar, but much less mobile are the linkages in the Fucα1-3GlcNAc and Manα1-6Man elements. The reduced mobilities have to be ascribed to interactions with the protein. It is noteworthy that the terminal Man and Fuc residues are located on the same side of the carbohydrate chain. The unaffected mobility of the Xyl residue indicates that it is exposed to the bulk solution.

This study has led to the important conclusion that conformational data on oligosaccharides or glycopeptides in solution cannot simply be transferred to glycans in glycoproteins.

II Ribonuclease-2

Non-secretory ribonuclease 2 (RNase2) is a heat and acid stable enzyme involved in the digestion of RNA. It is predominantly found in kidney, liver, lung, spleen and leukocytes. It has five N-glycosylation sites. The amino acid sequence, comprising 134 amino acids, has a high homology to other RNases. Surprisingly, Trp at position 7 bears an α-D-mannopyranosyl residue via a C-C bond [7,8]. This represents a novel type of glycosidic linkage in eukaryotic glycoproteins. Site directed mutagenesis has revealed that the sequence Trp-X-X-Trp is required to achieve C-mannosylation of the first Trp residue [9]. This sequence motif occurs in 336 mammalian proteins suggesting that this new form of glycosylation is not necessarily a rare feature. From a conformational point of view, the behaviour of the C-Man residue is interesting. For the glycopeptide Phe-Thr-[Manα]Trp-Ala-Gln-Trp, we concluded that, based on vicinal ^{1}H-^{1}H coupling constants and rotating frame nuclear Overhauser enhancements, the mannopyranosyl ring adopts an ensemble of conformations [8]. The usual $^4C_1$ conformation gives only a small, if any, contribution to the ensemble.

By means of NMR spectroscopy, we showed that the situation in the protein is even more complicated. In a denatured form, the Man residue behaves rather similar as in

the peptide. However, in the native form the Man residue has a different orientation around the C-C linkage. Obviously, this must be the result of interaction with the protein. To gain further insight into the conformational behaviour of the Manα1-Trp linkage dynamic simulations were carried out in vacuum starting from the $^{1}C_{4}$, $^{4}C_{1}$, and $^{1}S_{3}$ conformations with various force fields. It appeared that the $^{1}C_{4}$ conformation is preferred. Simulations in water with the GROMOS programme starting with the $^{1}S_{3}$ conformation showed the transition to the $^{1}C_{4}$ conformation after about 40 ps. The NMR data exclude almost completely the $^{4}C_{1}$ conformation. The contribution of the $^{1}S_{3}$ conformation can only be small, because a NOE between H1 and H4 of Man, which is characteristic for this conformation, can hardly be detected.

For modelling experiments, the X-ray structural data of recombinant eosinophil-derived neurotoxin (EDN) were used. EDN has the same amino acid sequence as RNase 2 and is also C-mannosylated. In conjunction with NMR data of native RNase, the modelling results demonstrate that the mannose residue has interactions with loop residues 115-123, the end of β strand Met105-Arg114 and with the beginning of β strand Pro124-Ile134. By these interactions, Trp7 is kept in a specific conformation [10]. For RNase2 the carbohydrate-protein interaction is important to realise a well-defined protein conformation.

III. Human Chorionic Gonadotropin

Human chorionic gonadotropin (hCG) belongs to a family of glycoprotein hormones. It is crucial for the maintenance of the corps luteum in early pregnancy. HCG consists of a non-covalently associated α,β heterodimer. The X-ray structure of the chemically deglycosylated hCG has been determined [11,12]. The subunits can be dissociated and the bioactivity can completely be restored upon recombination. This important experiment leads to the conclusion that the dissociation is reversible. Apparently, the subunits are not irreversibly denatured.

The α-subunit is of special interest, because in addition to the hormonal action when associated with the β-subunit, it has its own bioactivity [13]. The α-subunit consists of 92 amino acids and has two N-glycosylation sites at Asn52 and Asn78, respectively. The α-subunit is a good starting compound for analysing the solution structure of the hormone. Its size and the occurrence of two N-glycosylation sites make it an attractive

example of a glycoprotein and it is suited for the study of the mutual structural influences of intramolecular carbohydrate and protein chains.

Three probes are available for the determination of the solution structure of the α-subunit by NMR spectroscopy: intact αhCG, PNGase -F- deglycosylated native αhCG carrying only the glycan at Asn78 (pdαhCG), and endo-B-treated αhCG possessing only GlcNAc (fdαhCG) at Asn52 and Asn78. All NMR experiments were performed at natural abundance [14-18]. By application of various 2D NMR techniques, the ^{1}H signals of the amino acids and carbohydrates of the different NMR probes were assigned. Based on the NOE data, the protein structure of fdαhCG was modelled. Comparison with the crystallographic data for the heterodimer reveals some interesting differences in structure. Not surprisingly, there is a larger overall flexibility. This is in particular evident in the loop of the residues 33 – 57, which exhibit structural disorder, and also by conformational heterogeneity in the hairpin loop α70 – 74 and in the tight turn α20 – 23. The backbone conformation near Val76 and Glu77 is also altered. Further studies on pdαhCG and intact free αhCG showed that the local protein structure around Asn78 is influenced by glycosylation. This feature can be demonstrated by removal of the carbohydrate chain beyond the Asn linked GlcNAc, leading to a decreased structural order. The conformation of the glycan at Asn78 is in good accordance with the energy minima, present in rigid geometry maps of free model disaccharides. Furthermore, the protein sterically hinders the glycosidic linkages GlcNAcβ1-4GlcNAc and GlcNAcβ1-4Man. In addition to macroscopic effects like solubility and increased protease stability, the Asn78 linked GlcNAc functions as a stabiliser of the three-dimensional structure. This residue shields the protein surface from the environment through interactions with hydrophobic amino acid residues.

The glycan at Asn52 behaves quite differently and is far more flexible and solvent exposed. The first indication for this phenomenon stems from the accessibility of this glycosylation site in native αhCG for PNGase F. In sharp contrast with this observation is the resistance of this glycosylation site towards PNGase F in the not denatured heterodimer. Further indications are obtained from the NMR data. These first comprise the small number of NOE cross peaks between glycan and protein in comparison to the glycan at Asn78 and secondly the line widths.

Most of the differences in three-dimensional structure and flexibility between the free α-subunit and the α-subunit as part of the heterodimer have to be ascribed to

interactions with the β-subunit. Currently, studies are carried out to define these interactions in more detail.

SUMMARY

In this report, an overview is presented of studies on the three-dimensional structure in solution of glycoproteins. The investigations were carried out by using high resolution NMR spectroscopy in conjunction with molecular dynamics calculations and molecular modelling. For pine apple stem bromelain, ribonuclease 2 and the free α-subunit of human chorionic gonadotropin it could be shown that intramolecular interactions between carbohydrate and protein exist. Glycosylation sites are rather unique structural entities in glycoproteins, because of the local spatial arrangements. This is reflected in different intramolecular interactions between carbohydrate and protein.

REFERENCES

1. Montreuil, J., Vliegenthart, J.F.G. and Schachter, H. (1995) Glycoproteins, New Comprehensive Biochem. Vol 29a, Elsevier, Amsterdam.
2. Montreuil, J., Vliegenthart, J.F.G. and Schachter, H. (1997) Glycoproteins, New Comprehensive Biochem. Vol 29b, Elsevier, Amsterdam
3. Protein Brookhaven Databank.
4. Lommerse, J.P.M., Kroon-Batenburg, L.M.J., Kroon, J., Kamerling, J.P. and Vliegenthart, J.F.G. (1995) Conformations and internal mobility of a glycopeptide derived from bromelain using molecular dynamics simulations and NOESY analysis. J. Biomol. NMR, 5:79-94.
5. Lommerse, J.P.M., Kroon-Batenburg, L.M.J., Kamerling, J.P. and Vliegenthart, J.F.G. (1995) Conformational analysis of the xylose-containing N-glycan of pineapple stem bromelain as part of the intact glycoprotein. Biochemistry, 34:8196-8206.
6. Kroon-Batenburg, L.M.J., Kroon, J., Leeflang, B.R. and Vliegenthart, J.F.G. (1993) Conformational analysis of methyl beta-cellobioside by ROESY NMR spectroscopy and MD simulations in combination with the CROSREL method. Carbohydr. Res., 245:21- 42.
7. Hofsteenge, J., Müller, D.R., De Beer, T., Löffler, A., Richter W.J. and Vliegenthart, J.F.G. (1994) New type of linkage between a carbohydrate and a protein: C-glycosylation of a specific tryptophan residue in human RNase Us. Biochemistry, 33:13524-13530.
8. De Beer, T., Vliegenthart, J.F.G., Löffler, A. and Hofsteenge, J. (1995) The hexopyranosyl residue that is C-glycosidically linked to the side chain of tryptophan-

7 in human RNase Us is alpha-mannopyranose. Biochemistry, 34:11785-11789.
9. Krieg, J., Hartmann, S., Vicentini, A., Gläsner, W., Hess, D. and Hofsteenge, J. (1998) Recognition signal for C-mannosylation of Trp-7 in RNase 2 consists of sequence Trp-x-x-Trp. Mo.Biol.Cell, 9:301-309.
10. Vliegenthart, J.F.G. and Casset, F. (1998) Novel forms of protein glycosylation. Curr. Opin. Struct. Biol., 8:565-571.
11. Lapthorn, A.J., Harris, D.C., Littlejohn, Lustbader, J.W., Canfield, R.E., Machin, K.J., Morgan, F.J. and Isaacs, N.W. (1994) Crystal structure of human chorionic gonadotropin, Nature, 369:455-461.
12. Wu, H., Lustbader, J.W., Liu, Y., Canfield, R.E. and Hendrickson, W.A. (1994) Structure of human chorionic gonadotropin at 2.6 Å resolution from MAD analysis of the selenomethionyl protein, Structure, 2:545-558.
13. Blithe, D.L., Richards, R.G. and Skarulis, M.C. (1991) Free alpha human chorionic gonadotropin molecules from pregnancy stimulate secretion of prolactin from human decidual cells: a novel function or free alpha hcg in pregnancy, Endocrinology, 129:2257-2259.
14. Van Zuylen, C.W.E.M., De Beer, T., Rademakers, G.J., Haverkamp, J., Thomas-Oates, J.E., Hård, K., Kamerling, J.P. and Vliegenthart, J.F.G. (1995) Site-specific and complete enzymatic deglycosylation of the native human chorionic gonadotropin alpha-hcg subunit. Eur. J. Biochem., 231:754-760.
15. De Beer, T., van Zuylen, C.W.E.M., Leeflang, B.R., Hård, K., Boelens, R., Kaptein, R., Kamerling, J.P. and Vliegenthart, J.F.G. (1996) NMR studies of the free alpha-subunit of human chorionic gonadotropin; structural influences of N-glycosylation and the beta-subunit on the conformation of the alpha-subunit. Eur. J. Biochem., 241:229-242.
16. Van Zuylen, C.W.E.M., Kamerling, J.P. and Vliegenthart, J.F.G. (1997) Glycosylation beyond the Asn78-linked GlcNAc residue has a significant enhancing effect on the stability of the α subunit of human chorionic gonadotropin. Biochem. Biophys. Res. Commun., 232:117-120.
17. Thijssen-Van Zuylen, C.W.E.M., De Beer, T., Leeflang, B.R., Boelens, R., Kaptein, R., Kamerling, J.P. and Vliegenthart, J.F.G. (1998) Mobilities of the inner three core residues and the Man(alpha1-6) branch of the glycan at Asn78 of the alpha-subunit of human chorionic gonadotropin are restricted by the protein. Biochemistry, 37:1933-1940.
18. Erbel, P.J.A., Karimi-Nejad, Y., De Beer, T., Boelens, R., Kamerling, J.P. and Vliegenthart, J.F.G. (1999) Solution structure of the α-subunit of human chorionic gonadotropin. Eur. J. Biochem., 260:490-498.