Hypothesis

Do Asparagine-Linked Carbohydrate Chains in Glycoproteins Have a Preference for β-Bends?

Jaap J. Beintema

Received August 18, 1986

KEYWORDS: glycoprotein; conformation; β-bend.

X-ray structures of the conformation of carbohydrate moieties and connected regions of glycoproteins are summarized. Evidence is presented that there is some preference for carbohydrate attachment at β-bends. Evolution may have favored glycosylation to occur at bends to ensure free mobility of the carbohydrate moieties.

INTRODUCTION

Although current nucleic acid sequencing generally leads more rapidly to knowledge of the primary structure of a protein than classical protein sequencing, the latter remains essential to determine posttranslational modifications of the primary sequences. One of the most frequently occurring modifications of proteins is N-glycosylation of asparagine residues in Asn-X-Ser/Thr sequences. During an extensive study of the molecular evolution of pancreatic ribonuclease a large number of observations have been made on its glycosylation characteristics (1–4). These studies provided support for the current understanding of glycoprotein structure and biosynthesis (5).

Only asparagine residues in Asn-X-Ser/Thr sequences may act as attachment sites for carbohydrate (6). The transfer of an already assembled oligosaccharide moiety to a nascent polypeptide occurs in the lumen of the rough endoplasmic reticulum before
final folding of the protein (5). Carbohydrate attachment is thus a co-translational event, as could have been concluded from the buried position of the threonine side chain in the Asn-Leu-Thr sequence of bovine pancreatic ribonuclease (1, 7), a minor component of which, the B form, has carbohydrate attached.

Although carbohydrate is always attached to Asn-X-Ser/Thr sequences, not all such sequences in proteins synthesized in rough endoplasmic reticulum bear carbohydrate. Probably, other structural features of the protein determine the extent of glycosylation of a given site and further processing of the carbohydrate moiety. Investigations with synthetic peptides (8) and studies of species variants of glycoproteins (1–4) indicate that amino acid replacements near the Asn-X-Ser/Thr sequence, and also residue X, affect glycosylation. Also, the capability of the glycosylating system has a decisive influence. This not only varies with species (1–4), but also with tissue, as was shown by studies of the extent of glycosylation of deoxyribonuclease from bovine pancreas and parotid gland (9) and of ribonuclease from human pancreas, urine and semen (10).

X-ray studies of a number of glycoproteins have been published (Fc fragments of immunoglobulin molecules, Refs. 11, 12; haemagglutinin of influenza virus, Refs. 13, 14; neuraminidase of influenza virus, Ref. 15; hemocyanin from Panulirus interruptus, Ref. 16; bovine pancreatic deoxyribonuclease I, Ref. 17). In addition, three-dimensional structures are known of non-glycosylated proteins that also occur in a glycosylated form, or glycosylated close relatives of which occur (e.g. ribonuclease, Refs. 18–20; Japanese quail ovomucoid third domain, Ref. 21). Their structures may be compared to derive any rules which the conformation of carbohydrate moieties and connected regions of glycoproteins must obey.

It has been proposed that asparagine-linked carbohydrate occurs only at external β-bends in protein structures (22, 23). However, X-ray structures show a variety of conformations: carbohydrate is attached to homologous positions in a β-bend in the Fc fragments of human and rabbit immunoglobulin (11, 12), and to an asparagine in the middle of a β-strand structure in hemocyanin (16), and two of the seven carbohydrate chains in influenza virus haemagglutinin are attached to helical-type structures, two to β-sheets and two to β-bends, while the conformation of the seventh site is unknown because of invisibility of the polypeptide chain at that site on the electron density map (14).

**FUNCTION OF ATTACHED CARBOHYDRATE**

Although our knowledge of structure and biosynthesis of glycoproteins has increased considerably during recent years, functional aspects of the attachment of carbohydrate remain more obscure. Possibly, carbohydrate moieties may have several functions. One should even consider the possibility that they sometimes have no function at all, as may be the case with the short oligosaccharide chain in bovine ribonuclease B, the minor glycosylated component. Thus, if an Asn-X-Ser/Thr sequence passes the glycosylating enzyme system during biosynthesis of a polypeptide chain on the endoplasmic reticulum, the protein will usually be glycosylated. This may be an evolutionarily neutral property or a functionally or structurally advantageous one. Evolutionary selection against the occurrence of Asn-X-Ser/Thr sequences should
occur if glycosylation is disadvantageous for the stability or the function of the protein. Evidence for this has been presented by Sinohara and Maruyama (24), who observed that Asn-X-Ser/Thr sequences occur less frequently than random in extracellular proteins from eukaryotes.

Possible functional roles of carbohydrate moieties are as intracellular or extracellular recognition signals for other proteins, as specific building elements in the three-dimensional structures of proteins and as a more diffuse general protection of protein structures. Depending on the type of function, the mobility of carbohydrate moieties relative to the protein core and, thereby their visibility on electron density maps of crystals, may vary. Visible carbohydrate moieties occur in the crystal structures of both immunoglobulins (11,12), of hemocyanin (16) and of deoxyribonuclease (17). Only part of them are visible in both influenza virus proteins (13–15), the other ones and the carbohydrate moiety in kallikrein (25) are invisible. Invisibility may be caused either by a high free mobility relative to the core, even in the crystalline state, or by adoption of a number of different orientations relative to the core (26).

**STRUCTURE OF ATTACHED CARBOHYDRATE**

Crystal structures, both of glycoproteins and of oligosaccharides with covalent structures as encountered in glycoproteins (27), and Nuclear Magnetic Resonance (NMR) studies (28), indicate that the two internal N-acetylglycosamines and the innermost mannose linked to them form a hydrogen-bonded rigid coplanar “cellulose-like” structure. Thus, the visibility of carbohydrate moieties in crystal structures should be determined primarily by the freedom of rotation of the side-chain of asparagine and of the internal trisaccharide core relative to the protein.

In immunoglobulin Fc fragments (11, 12) and in the influenza virus glycoproteins (13–15), carbohydrate moieties participate extensively in inter-subunit contacts. They may have an important structural role and, therefore, have a discrete position in the conformation. One of the other roles of attached carbohydrate in influenza virus proteins may be masking of certain regions of the protein surface from interaction with the immune system of the host (14).

The carbohydrate chains in Fc fragments fold back along the surface of the subunits to which they are connected. This is possible because of the presence of a cis-amide bond between asparagine and carbohydrate. A contribution to the immobilization of a carbohydrate chain connected by a trans-amide bond to virus haemagglutinin is provided by a hydrogen bond between the C-6 hydroxyl of the innermost N-acetylglicosamine and the hydroxyl group of a threonine in the β-strand, located two residues from the asparagine to which the carbohydrate is linked (14). In hemocyanin, the visible carbohydrate chain is also linked to an asparagine residue in a β-sheet structure. This chain has very little other contact with the protein surface. However, the density around the asparagine-side chain suggests that its carbonyl oxygen is hydrogen-bonded to the side-chain of an arginine residue in the opposite strand of the β-sheet (Volbeda, A. and Hol, W. G. J., personal communication). Since steric hindrance restricts rotation around the bond between the innermost N-acetylglicosamine and the nitrogen of the amide group of asparagine
rather severely, and since the amide group itself is planar, mobility of the planar carbohydrate core region relative to the polypeptide backbone may be largely dependent upon rotation around the two bonds between the C-atoms of the asparagine side-chain. Thus, hydrogen bonding of the carbonyl group of asparagine should restrict the mobility of the carbohydrate linked to it.

Recently, Montreuil (29) has reviewed evidence on spatial structures of glycan chains of glycoproteins and proposed an umbrella-like conformation with the rigid internal trisaccharide core extending more or less perpendicular to the protein surface and the branches (antennae) parallel to the surface. The latter may be mobile or make specific contacts with the protein surface. There is evidence, for instance, that the \( \alpha(1 \rightarrow 6) \)-linked chain may fold back along the core region (29), enlarging the protein-carbohydrate contact area. However, simple-type oligosaccharide chains, such as occur in bovine ribonuclease B, cannot form an additional interaction site.

**ATTACHMENT AT \( \beta \)-BENDS?**

The strict requirement of carbohydrate attachment for an Asn-X-Ser/Thr sequence suggests that the polypeptide substrate should have a very specific conformation—at least temporarily—in order to be recognized by the active site of the glycosylating enzyme system. A hypothetical structure of such a conformation has been described by Marshall (6). It is lost during final folding of the polypeptide chain, for glycoproteins of known structure have different conformations near their carbohydrate attachment sites (11–17). A change in polypeptide conformation after posttranslational modification has also been hypothesized for the hydroxylation of proline in collagen before formation of the collagen triple-chain fold (30).

However, there may be a preference of carbohydrate for a localization at external bends of protein structure. Carbohydrate attachment has been observed to occur at positions 21, 22, 34, 62, 76 and 88 in ribonucleases (1, 2, 3, 4, 10). All these positions are localized in or close to \( \beta \)-turns or to bends which do not strictly obey to the definition of a \( \beta \)-turn in bovine ribonuclease A (20). X-ray structures of other non-glycosylated proteins also indicate that carbohydrate is attached to asparagine residues in external \( \beta \)-turns (21). Aubert et al. (22) and Beeley (23) suggested that carbohydrate occurs only at asparagine residues in \( \beta \)-bends. They found a rather high \( \beta \)-turn potential of peptide stretches with asparagine-linked carbohydrate in a number of glycoproteins with the method of Chou and Fasman (31). (The Asn-X-Ser/Thr sequence itself has a lower potential for forming \( \beta \)-turns.) However, there was no preferential occurrence of asparagine at any of the four positions in a \( \beta \)-bend, and if asparagine occurs at the third or fourth position, the serine or threonine is outside the bend. Therefore, the \( \beta \)-bend hypothesis does not fulfill the strict specificity requirement of an enzyme-catalyzed reaction. Also, Mononen and Karjalainen (32) observed that, when they predicted secondary-structural elements in a large number of proteins, about 70% of the Asn-X-Ser/Thr sequences occurred in \( \beta \)-bends, 20% in \( \beta \)-sheets and 10% in \( \alpha \)-helices, and that there was no difference between glycosylated and non-glycosylated sequences.

Hypotheses on conformational differences between homologous proteins derived from differences in \( \beta \)-turn potential should also be considered with caution. The attachment of carbohydrate to the Asn-Gly-Ser sequence in rat \( \alpha \)-lactalbumin but not
to the homologous Asn-Gln-Ser sequence in bovine $\alpha$-lactalbumin was explained by the shift of a $\beta$-bend (33). However, a model of the sequence built into the crystal structure of lysozyme predicted a bend in bovine $\alpha$-lactalbumin exactly at the position of the Asn-Gln-Ser sequence (34).

CONCLUSIONS

Thus, there is a preference, though not an absolute one, for the location of carbohydrate moieties at external $\beta$-bends, especially in relatively small proteins. This preference may be explained not from structural requirements but from a functional one, in which attached carbohydrate acts as a diffuse protection layer around the protein. The carbohydrate layer may be described as part of the surrounding solvent of these glycoproteins, but protects them from proteolytic degradation or adsorption to cellular surfaces (35, 1). Also, the carbohydrate layer decreases the ionic strength in the vicinity of the protein surface (35, 36). The protective function of the carbohydrate chains is ensured by their free mobility relative to the protein surface, like in the umbrella conformation proposed by Montreuil (29). Figure 1 presents such a structure for a ribonuclease molecule with three attached carbohydrate moieties. However, such mobility is only possible if the asparagine side-chain can rotate freely and if there are no further interactions between the carbohydrate and the protein surface. This requirement can only be met if the attachment occurs at external bends. Therefore, selection has favored glycosylation at positions in nascent polypeptide chains that adopt this conformation during folding.

Fig. 1. Space-filling models of bovine ribonuclease (left) and of a glycosylated ribonuclease with three biantennary complex-type oligosaccharide chains attached to the surface (right). This Figure is a collage of the Figure on p. 81 in Ref. 37 (ribonuclease) and Figure 11B in Ref. 29 (spatial conformation of a biantennary glycan of the N-acetyllactosamine type in the T-conformation).
ACKNOWLEDGEMENTS

This paper results from a study together with Dr N. Borkakoti at Birbeck College (London, UK) to attach carbohydrate to bovine pancreatic ribonuclease by model building, using an Evans & Sutherland Picture II molecular graphics system. Many thanks are due to Dr R. N. Campagne for carefully reading the manuscript.

REFERENCES