Glycosylation Demystified

Skylar Washington

Glycosylation Demystified

Edited by **Skylar Washington**

New York

Published by Callisto Reference,
106 Park Avenue, Suite 200,
New York, NY 10016, USA
www.callistoreference.com

Glycosylation Demystified
Edited by Skylar Washington

International Standard Book Number: 978-1-63239-365-4 (Hardback)

Printed in the United States of America.

Glycosylation Demystified

Preface

I am honored to present to you this unique book which encompasses the most up-to-date data in the field. I was extremely pleased to get this opportunity of editing the work of experts from across the globe. I have also written papers in this field and researched the various aspects revolving around the progress of the discipline. I have tried to unify my knowledge along with that of stalwarts from every corner of the world, to produce a text which not only benefits the readers but also facilitates the growth of the field.

Glycosylation is generally described as a form of co-translational and post-translational modification. This book emphasizes on the concepts of glycobiology, and is a collective effort of a group of recognized researchers. It addresses some of the crucial topics in this domain, presenting a broad range of theoretical and practical topics in the subject of glycobiology. This book will be of great significance for scholars and researchers involved in drug glycoengineering and biomedical research.

Finally, I would like to thank all the contributing authors for their valuable time and contributions. This book would not have been possible without their efforts. I would also like to thank my friends and family for their constant support.

Editor

Contents

Glycosylation and Structure

Beyond the Sequon: Sites of N-Glycosylation

Benjamin Luke Schulz

Additional information is available at the end of the chapter

1. Introduction

Asparagine (N-) linked protein glycosylation is a common and essential post-translational modification of proteins in eukaryotes, archaea and some bacteria. It plays crucial roles in protein folding and in regulation of protein function. Although the general principles of N-glycosylation have been long known, the precise details governing whether a particular asparagine residue will be N-glycosylated or not are not well understood. This is of broad general importance in understanding the structure and function of the immense variety of N-glycoproteins in diverse biological systems. This chapter will review the current understanding of the mechanisms that determine how asparagine residues are selected for glycosylation by the enzyme oligosaccharyltransferase.

2. Overview of N-glycosylation in the endoplasmic reticulum

The initial steps in N-glycosylation take place in the lumen of the endoplasmic reticulum (ER). The enzyme oligosaccharyltransferase (OTase) catalyzes the key step in N-glycosylation, *en bloc* transfer of mature glycan from a lipid carrier to selected asparagine residues in nascent polypeptide chains (Kelleher & Gilmore, 2006). Glycan to be transferred to protein is synthesized by sequential addition of monosaccharides linked to a dolichol pyrophosphate lipid carrier (Burda & Aebi, 1999). This process is essentially linear, and in most organisms OTase specifically recognizes the final α-1,2-linked glucose, ensuring efficient transfer of only the mature $Glc_3Man_9GlcNAc_2$ glycan structure (Karaoglu *et al.*, 2001).

2.1. Oligosaccharyltransferase

The OTase enzyme is a multiprotein complex in most eukaryotes, and in yeast consists of 8 protein subunits (Ost1p, Ost2p, Ost3/6p, Ost4p, Ost5p, Swp1p, Wbp1p and Stt3p) (Kelleher & Gilmore, 2006). It is now clear that the Stt3p protein houses the catalytic site of OTase,

while the accessory protein subunits of multiprotein complex OTases are required for complex stability, enzymatic regulation of OTase activity, substrate recognition and OTase enzyme localization (Mohorko *et al.*, 2011). OTase physically associates with the translocon (Shibatani *et al.*, 2005, Yan & Lennarz, 2005) and the ribosome (Harada *et al.*, 2009), and so has direct access to nascent polypeptides immediately as they enter the ER lumen (Dempski & Imperiali, 2002). Glycosylation of many asparagines is co-translocational, and occurs essentially as soon as they enter the ER lumen and can reach the OTase active site (Whitley *et al.*, 1996). Other sites are also glycosylated post-translocationally, with extended residence of protein in the ER lumen (Ruiz-Canada *et al.*, 2009). However, in all cases the protein substrate of OTase must be unfolded for glycosylation to occur.

2.2. Roles of N-glycans in protein folding

The key role of N-glycans on proteins in the ER is to assist in productive protein folding (Helenius & Aebi, 2004). By virtue of their hydrophilic bulk, N-glycans alter the overall biophysical properties of nascent polypeptides, increasing their solubility and constraining local polypeptide conformation (Wormald & Dwek, 1999). N-glycans can also function as signals for incomplete folding of particular domains of proteins, and so direct these to the ER resident thiol oxidoreductase ERp57 via the lectins calnexin and calreticulin (Oliver *et al.*, 1999). Timed trimming of N-glycans on glycoproteins in the ER lumen is also key for regulating retro-translocation of incorrectly folded glycoproteins to the cytoplasm for degradation (Aebi *et al.*, 2010).

3. The 'glycosylation sequon'

The key recognition factor for selection of asparagines for glycosylation by OTase is the 'glycosylation sequon'. This has been historically defined as Asn-Xaa-Ser/Thr (Xaa ≠ Pro). However, it has also long been clear that this is not an adequate predictor of glycosylation, as ~1/3rd of Asn in sequons in secreted proteins are not glycosylated. In addition to this, several examples of glycosylation of Asn residues not in sequons have been reported in recent years.

3.1. Definition of the sequon

The term 'sequon' was likely first used by Derek Marshall (Marshall, 1974) to describe the apparent three amino acid local sequence requirement for N-glycosylation. However, it was long recognized that the presence of a sequon was not sufficient for N-glycosylation to occur at a given Asn in portions of polypeptides entering the ER lumen. Nonetheless, the efficiency of glycosylation at a given asparagine is primarily determined by the flanking amino acids, with the primary factor increasing glycosylation being the presence of a threonine or serine at the +2 position. This has such a strong influence of the efficiency of glycosylation that it has been termed the 'glycosylation sequon' in recognition of its importance. However, the presence of a glycosylation sequon is neither necessary nor sufficient for an asparagine to be glycosylated.

3.2. The '+2' position: Thr, Ser, Cys, Etc

Whilst both Ser and Thr are accepted as amino acids at the +2 position in glycosylation sequons, they are not equal, as glycosylation of Asn-Xaa-Thr sequons is approximately 40 times efficient than of Asn-Xaa-Ser sequons (Kasturi *et al.*, 1995, Kasturi *et al.*, 1997). Far and away the majority of glycosylated asparagines are in traditional Asn-Xaa-Ser/Thr (Xaa≠Pro) sequons. However, several very well validated examples have been reported of asparagines *not* in sequons that are nonetheless efficiently glycosylated.

Several reports have been made of glycosylation at asparagines in the sequence Asn-Xaa-Cys. Human CD69 has such an Asn-Xaa-Cys glycosylation site (Vance et al, 1997). Human beta protein C is glycosylated at an Asn with cysteine at the +2 position (Miletich & Broze, 1990). Interestingly, the Cys in beta protein C is involved in a disulfide bond in the mature protein, and the formation of this disulfide competes directly with glycosylation at the preceding Asn. CHO-cell expressed recombinant human epidermal growth factor receptor (EGRF) also has such a glycosylation site (Sato *et al.*, 2000). Heterologous expression of an insect cathepsin B-like counter-defense protein in *Pichia pastoris* resulted in glycosylation at an asparagine in the sequence Asn-Xaa-Cys (Chi *et al.*, 2010). It is unclear if this site is also natively glycosylated. This shows that both mammalian and fungal OTase are capable of glycosylating selected Asn-Xaa-Cys sequences.

Several large-scale discovery projects for identification of N-glycosylation sites have been performed. The largest of these, from mouse, identified over 5000 putatively glycosylated asparagines (Zielinska *et al.*, 2010). While the vast majority of these were in conventional Asn-Xaa-Ser/Thr sequons, a small but significant number of Asn not in such sequons were identified as being glycosylated. Asn-Xaa-Cys sites represented 65/5052, and Asn-Xaa-Val 20/5052. It was also reported that Asn-Gly sites were modified. However, this result must be treated with extreme caution, given the propensity for non-catalyzed spontaneous deamidation (asparagine-aspartate conversion) is especially high at Asn-Gly sequences (Palmisano *et al.*, 2012, Robinson *et al.*, 2004)

It was proposed that the hydroxyl group of Ser/Thr amino acids at the +2 position was directly involved in catalysis, via the formation of an 'Asparagine turn' (Imperiali & Hendrickson, 1995). This proposal was certainly powerful, and could withstand the observation of rare Asn-Xaa-Cys glycosylation sequons with the relatively weak hydrogen bonding capacity of the cysteine sulfhydryl group. However, apparent glycosylation of Asn-Xaa-Val sequons could not be explained by this mechanism. Resolution of the role of the +2 amino acid in determining glycosylation needed to wait until an atomic resolution structure of OTase was available.

3.3. Further a field: The 'X' position and beyond

The amino acids immediately proximal to the glycosylated Asn also influence the efficiency of its glycosylation. Experimental manipulation of model proteins has shown that the +1 position of an Asn has a strong effect on its extent of glycosylation, with bulky hydrophobic or acidic amino acids strongly reducing glycosylation occupancy, and small, hydrophilic or basic amino

acids giving high levels of modification (Shakin-Eshleman *et al.*, 1996). These results may be misleading, as glycosylation only occurs before protein folding, and so mutations which disrupt or slow local protein folding could make extrapolation of such results difficult. However, roughly this same overall pattern has also been observed in non-experimental comparisons of glycosylated and non-glycosylated Asn (Petrescu *et al.*, 2004). Interplay with the amino acid at the +2 position has also been shown to be important. Studies in a model glycoprotein showed that amino acid substitutions at the +1 position that reduced glycosylation efficiency with Ser at the +2 position were still completely modified if Thr was at the +1 position (Kasturi et al., 1997). The major difficulty in interpreting these results is that the amino acids in the vicinity of a glycosylated Asn residue influence both specific interactions with OTase *and* local protein folding, stability and dynamics. As it is clear that protein folding and glycosylation are intimately linked, separating these effects is difficult.

In addition to local sequence dependency, the position of an asparagine within its protein sequence also contributes to the extent or probability of glycosylation. For instance, probability and extent of glycosylation increases with increasing distance from the C-terminus of a protein. This has been measured both experimentally using manipulation of model proteins and by *in silico* surveys of large sets of experimentally characterized native glycoproteins (Bano-Polo *et al.*, 2011, Rao *et al.*, 2011). This effect is perhaps due to the increased relative protein folding or translocation rates towards the C-terminus.

3.4. The extended bacterial glycosylation sequon

The discovery of N-glycosylation systems in bacteria that are homologous to those in eukaryotes promised rapid progress in understanding the molecular basis for their specificity and activity, given their comparative simplicity and ease of manipulation (Szymanski *et al.*, 1999, Wacker *et al.*, 2002). Initially it was observed that the *C. jejuni* N-glycosylaiton system modifies Asn with very similar local sequence requirements to eukaryotic N-glycosylation sites, that is an Asn-Xaa-Ser/Thr sequon was required but not sufficient for glycosylation (Wacker et al., 2002, Nita-Lazar *et al.*, 2005). Later, it was found that an extended 'sequon' was needed for bacterial glycosylation, with the added requirement of an acidic residue at the -2 position: Asp/Glu-Xaa-Asn-Xaa-Ser/Thr (Xaa≠Pro) (Kowarik *et al.*, 2006b). Close homologues to the *C. jejuni* PglB OTase showed a less strict sequon (Schwarz *et al.*, 2011b). In either case, such an extended sequon was not sufficient for modification. A key defining factor determining glycosylation was that such a sequon was efficiently glycosylated in unfolded polypeptide or in flexible stretches of folded proteins (Kowarik *et al.*, 2006a). Thus, as in the eukaryotic system, flexible acceptor substrate was a key requirement for bacterial OTase.

3.5. Structural insights into the requirement for the glycosylation sequon

The high-resolution 3D crystal structure of the *Campylobacter lari* PglB OTase finally provided a structural basis for the requirement of a glycosylation sequon (Lizak *et al.*, 2011b). This structure was solved with co-crystallization of an acceptor peptide. The key pertinent feature of the structure was that the +2 position Thr was too far away from the Asn

to be directly involved in catalysis. Instead, this Thr was hydrogen bonded with two tryptophans and the aspartate in the WWDYG motif conserved in all known OTase homologues. Thr also formed van der Waals interactions with Ile572 of PglB, which Ser at the +2 position could not form, explaining the preference for Thr over Ser in sequons. Proline at the +1 or -1 position would not have allowed this binding conformation, providing a structural basis for the requirement that proline not be present at these positions at glycosylation sites. The requirement of bacterial OTases for an acidic amino acid in the -2 position (Kowarik et al., 2006b) was also explained by formation of a salt bridge from this residue to Arg331 that is conserved in bacterial, but not eukaryotic, PglB/Stt3p OTases.

This structure of the PglB OTase provides clear evidence that the role of the glycosylation sequon is to increase the binding affinity of asparagines to the active site of OTase (Lizak et al., 2011b). Accessory subunits of multiprotein complex OTases in many eukaryotes have been shown to bind substrate polypeptide, perhaps contributing to increasing the binding affinity of specific Asn and leading to the short requirement of specific binding of an Asn-Xaa-Ser/Thr. In contrast, the single protein OTases such as the bacterial PglB may have evolved the requirement for an extended sequon in the absence of such additional binding by accessory OTase subunits.

3.6. The future of the sequon

How to best define the glycosylation 'sequon'? Many factors influence whether a particular asparagine is glycosylated, including: binding affinity of the region immediately proximal to the Asn to the polypeptide acceptor site of OTase; local folding, such as secondary structural elements, disulfide bond formation or hydrophobic collapse; the regulatory state of OTase, including the concentration and structure of lipid-linked oligosaccharide donor; protein expression rate, both global (rate of protein secretion saturates OTase catalytic ability) and local (position of Asn within the protein sequence); and the affect of glycosylation at an Asn on the total possibility of protein folding. (If glycosylation at a given Asn would not allow correct folding of the protein, such that the portion of nascent polypeptides that were glycosylated there would never correctly fold, then that Asn would appear to never be glycosylated. The converse is also true, that if glycosylation is strictly required at a particular Asn for correct protein folding, then that Asn will appear to always be glycosylated, even if most of the nascent polypeptide is not modified and degraded by the quality control systems of the ER.)

It is the combination of these factors that determines if a particular Asn reaches the threshold for modification by OTase. However, even the definition of this threshold is an analytical artefact, as it is increasingly apparent that most glycosylated Asn are only partially modified, with some portion ranging from a fraction of a percent to essentially all copies of a protein, actually glycosylated (Hülsmeier *et al.*, 2007, Sumer-Bayraktar *et al.*, 2011). This pattern seems to contrast with the general requirement of many proteins for N-glycosylation for correct and efficient protein folding (Helenius & Aebi, 2004). Two key factors probably explain this conundrum. Many proteins can fold correctly even without glycosylation at many sites, as long as a certain critical level of glycosylation is present,

perhaps sufficient for ER-lectin chaperone recruitment to crucial protein domains, or overall biophysical solubility. Additionally, Asn residues are inherently likely to be present at the ends of secondary structural elements. This means that glycosylation at such sites is, in general, not likely to strongly disrupt protein folding.

In the end it appears that the descriptive beauty of the 'glycosylation sequon' is actually a dramatic simplification. However, the current state of knowledge is far from being able to quantify the 'glycosylatability' of a particular Asn. In place of this developing skill, the 'sequon' as it is traditionally defined is still a very accurate predictor of the possibility of glycosylation.

4. Oligosaccharyltransferase defines the sequon

The enzyme oligosaccharyltransferase (OTase) catalyses transfer of oligosaccharide from lipid to nascent polypeptide in the ER. However, while this enzyme shows a high degree of conservation between species with respect to the small scale reaction it catalyses, the immense range of different polypeptide substrates in various biological systems can be efficiently glycosylated because of co-evolution of these substrate proteins and the acceptor specificities of OTase. In turn, this evolutionary history determines whether a particular asparagine residue will be efficiently glycosylated in a given biological system. The OTase defines the 'sequon'.

4.1. OTase protein subunits

OTase consists of the catalytic protein subunit Stt3p/PglB with varying numbers of additional accessory subunits in different organisms (reviewed in (Kelleher & Gilmore, 2006, Mohorko *et al.*, 2011)). Comparison of the evolutionary tree of eukaryotes with the protein subunit composition of OTase implies that accessory protein subunits have been added sequentially during eukaryotic evolution, starting from an ancestral single protein Stt3p OTase enzyme. The functions of most accessory OTase subunits are not clearly defined, although roles in recognition and regulation of glycan and protein substrate have been proposed.

4.2. Single protein OTases

Some divergent eukaryotes such as *Trypanosoma, Leishmania* and *Giardia* have single subunit OTases, consisting of only a catalytic Stt3p protein. However, many species within these groups have multiple different Stt3p homologues. In all of the systems in which the functions of these homologues have been characterized it is apparent that this duplication is functionally important, as the different enzymes vary in their protein acceptor and/or glycan donor substrate specificities.

4.2.1. Single protein OTases in Trypanosoma brucei

Trypanosoma brucei, the causative agent of sleeping sickness, has a genome encoding three full-length Stt3p homologues. Several lines of evidence *in vivo* in *T. brucei*, in *in vitro*

enzyme assays and in a yeast *ex vivo* system support a model in which these three enzymes transfer different glycan structures to selected sets of Asn residues – they have different specificities for both the glycan they transfer, and the asparagines they modify (Izquiedro *et al.*, 2012, Izquiedro *et al.*, 2009). With regard to the asparagine residues on proteins glycosylated by these homologues, heterologous expression of these proteins in *S. cerevisiae* lacking the yeast *STT3* gene and quantitative analysis of glycosylation site occupancy in cell wall glycoproteins showed that the two of these proteins that allowed survival had different protein substrate specificities. The TbStt3B enzyme efficiently glycoslyated Asn surrounded by basic residues, while the TbStt3C enzyme preferentially glycosylated Asn surrounded by acidic residues. These substrate specificities correlated with the presence of complementary residues near the active site of the TbStt3B (acidic) and TbStt3C (basic) enzymes. This suggests that these enzymes have alternate protein substrate specificities determined by ionic interactions between the peptide-binding site and protein substrates. This specificity can be viewed as a type of ill-defined 'extended sequon', similar to the requirement of bacterial OTase for the extended Asp/Glu-Xaa-Asn-Xaa-Ser/Thr sequon, but with less stringency to the precise location of the charged residues.

4.2.2. Single protein OTases in Leishmania major

The single subunit OTase enzymes of the related *Leishmania major* have also been studied *ex vivo (Nasab et al 2008)*. Heterologous expression of the four different *Leishmania major* STT3 protein homologues in *S. cerevisiae* showed that these proteins do not integrate into the yeast OTase complex, but are instead truly single subunit enzymes. Not all of these homologues were capable of allowing survival of *S. cerevisiae* in the absence of the yeast OTase activity, and those that did complement lack of yeast OTase activity showed different protein substrate specific activities – the enzymes showed differences in the glycosylation sites they glycosylated efficiently.

4.2.3. Role of OTase catalytic subunit homologues STT3A and STT3B

Even when present in multiprotein complexes, STT3 homologues have different activities. OTase complexes containing either of the homologous mammalian STT3A and STT3B proteins have different kinetic parameters (Kelleher *et al.*, 2003), and are also responsible for either co-translocational or post-translocational N-glycosylation (Ruiz-Canada et al., 2009), thereby glycosylating different protein substrates (Wilson & High, 2007). However, it is not clear if there is further definition of protein or glycan substrate specificity defined by the presence of Stt3A or Stt3B in an OTase complex.

4.3. Role of accessory OTase proteins

In organisms with multiprotein complex OTases, there are several lines of evidence that some of these additional non-catalytic subunits provide different protein substrate specificities and allow regulation of oligosaccharide substrate recognition and enzymatics.