

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,000

Open access books available

125,000

International authors and editors

140M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Beyond the Sequon: Sites of N-Glycosylation

Benjamin Luke Schulz

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/50260>

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 $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ 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 \neq 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 (EGFR) 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-glycosylation 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 (Izquierdo *et al.*, 2012, Izquierdo *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 glycosylated 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.

4.3.1. Role of accessory OTase proteins Ost3p and Ost6p

The *S. cerevisiae* Ost3p and Ost6p OTase subunits are homologous proteins with the same topology of a thioredoxin-like N-terminal ER luminal domain, followed by four transmembrane helices (Fetrow *et al.*, 2001, Schwarz *et al.*, 2005). These proteins are homologous to the mammalian proteins TUSC3 and MagT1 (Kelleher & Gilmore, 2006). Only one of these homologues is incorporated into a given OTase complex, meaning that there exist two isoforms of OTase in yeast, defined by the presence of either Ost3p or Ost6p (Schwarz *et al.*, 2005, Spirig *et al.*, 2005). These different isoforms have been shown to have different protein substrate specific glycosylation efficiencies at the level of individual glycosylation sites (Karaoglu *et al.*, 1995, Knauer & Lehle, 1999, Schulz & Aebi, 2009). Some Asn residues require the Ost3p-OTase for efficient glycosylation, while other Asn residues require the Ost6p-OTase (Schulz & Aebi, 2009). The mechanistic basis for this difference is likely transient binding of stretches of nascent polypeptide by peptide-binding grooves in the ER luminal domains of Ost3p and Ost6p (Schulz *et al.*, 2009). *In vitro* assays have shown that Ost3p and Ost6p can transiently non-covalently bind peptides complementary to the characteristics of their peptide-binding grooves (Jamaluddin *et al.*, 2011, Schulz *et al.*, 2009). As the amino acids forming these grooves are different in Ost3p and Ost6p, it is proposed that they would tend to bind different stretches of nascent polypeptide, thereby increasing the efficiency of glycosylation at distinct sets of glycosylation sites. The peptide-binding specificity of yeast Ost6p is for short stretches of aliphatic amino acids, with additional affinity provided by the presence of neighbouring acidic residues (Jamaluddin *et al.*, 2011). Such sequences are complementary to the peptide binding groove of Ost6p, as revealed by its 3D crystal structure, which shows a groove with a hydrophobic base, and lined by neutral and basic amino acids (Schulz *et al.*, 2009). The dimensions of the groove are appropriate for binding a ~4-5 amino acid stretch of extended polypeptide, or an amphipathic alpha helix. Ost3p also binds hydrophobic stretches of polypeptide, but with a distinct amino acid characteristic specificity to Ost6p (Jamaluddin *et al.*, 2011). It has also been proposed that the oxidoreductase activity of the thioredoxin-like ER luminal domain of Ost3p and Ost6p could form mixed disulfides with cysteines in nascent polypeptides (Schulz *et al.*, 2009). This would serve to tether nascent polypeptide close to the active site of OTase, and also efficiently inhibit oxidative protein folding. This transient binding of cysteines and hydrophobic stretches of nascent polypeptide is an optimal strategy for inhibiting local protein folding, as such stretches would be normally internal to folded protein domains. This fits with the requirement of the catalytic site of OTase for unfolded or flexible protein acceptor substrate.

4.3.2. Role of accessory OTase protein ribophorin I / Ost1p

Mammalian Ribophorin I (Ost1p in yeast) is required for efficient glycosylation of selected membrane proteins. Ribophorin I physically associates with selected membrane proteins after insertion into the ER membrane (Wilson *et al.*, 2005). This interaction with these selected substrate proteins was also shown to be required for their efficient glycosylation by OTase (Wilson & High, 2007). The interaction between selected membrane proteins and ribophorin I is direct, but the precise mechanisms of the interaction are not clear (Wilson *et al.*, 2008). It is possible that ribophorin I / Ost1p function in a conceptually similar way to

Ost3/6p, in transiently tethering substrate protein close to the catalytic site of OTase to allow efficient glycosylation of a defined subset of glycosylation sites or glycoproteins.

4.3.3. Additional known accessory OTase proteins

An integral membrane protein with homology to the integral membrane domain of Ost3p and Ost6p has been identified in mammalian cells. This protein, DC2 or OSTC, is required for glycosylation of specific substrate glycoproteins (Wilson & High, 2007). A further protein, Keratinocyte-associated protein 2 (KCP2), has been shown biochemically to be a subunit of the mammalian OTase (Sanyal & Menon, 2010, Roboti & High, 2012), and to be required for glycosylation of some proteins (Wilson & High, 2007).

4.3.4. Putative accessory OTase protein presenilin 1

A direct link between site-specific glycosylation and Alzheimer's disease has been made, through the Presenilin-1 protein (Lee *et al.*, 2010). N-glycosylation of the vacuolar ATPase subunit V0a1 is mediated by selective binding of the Alzheimer's disease related protein presenilin-1 (PS1) to unglycosylated V0a1 and OTase. V0a1 glycosylation is required for ER-lysosome trafficking, and so lack of PS1 causes deficiencies in lysosomal acidification and proteolysis during autophagy. It is not clear if PS1 is a truly protein-specific enhancer of glycosylation, or if it interacts with additional substrate glycoproteins to enhance their glycosylation.

4.3.5. How many OTase subunits are there?

Have all OTase subunits been identified? Most known OTase subunit proteins have been identified in the yeast *S. cerevisiae* through genetic screens, and so it is likely that this set is complete. These proteins are identifiable in other eukaryotes including animals and plants. However, the presence of additional subunits cannot be easily predicted. As described above, several such additional subunits have been identified biochemically in recent years in the mammalian OTase enzyme.

It is possible that other, less tightly bound or lowly expressed proteins are yet to be identified. It is also possible that sequential addition of accessory proteins to the OTase complex has proceeded divergently in different eukaryotic lineages. This would mean that biochemical analyses, rather than genomic comparisons, would be necessary to identify any additional OTase complexes in for example the plant or protozoan OTase. Any such additional subunits would likely have diverse additional roles in regulation of OTase core activity.

5. Analytical approaches to determine site-specific glycosylation occupancy

5.1. Glycosylation site identification and occupancy

A goal of understanding the function of OTase in diverse biological systems is to enable accurate prediction of whether a particular Asn will be efficiently glycosylated. However,

such prediction depends on a complete understanding of how OTase interacts with substrate polypeptides in each biological system, and as such is probably a very difficult problem. In addition to the diversity of OTase subunit proteins, OTase activity may also be subject to regulation. In the absence of accurate prediction tools, analytical identification and quantification of glycosylation occupancy is therefore necessary for accurate characterisation of the glycosylation status of a protein. In addition, it is not sufficient to identify that a site is glycosylated, as an Asn can be identified as 'glycosylated' in enrichment experiments, but may actually only be modified at a very low occupancy. The physiological relevance of glycosylation at such sites is therefore questionable. The converse of this is also true, as it appears that with sensitive analytical detection some or even most glycosylation sites are not completely occupied (there exists a small but significant proportion of proteins that are not glycosylated at that particular site) (Hülsmeier *et al.*, 2007). Analytical methods should therefore consider the proportion of a particular Asn that is glycosylated, for instance using LC-MS approaches that can compare the abundance of glycosylated and non-glycosylated versions of the same peptide (Schulz & Aebi, 2009). Although these methods are not in general absolutely quantitative, they can provide relative quantification and a first step towards characterization of the site-specific extent of glycosylation.

5.2. Western blotting for measuring glycosylation occupancy

Numerous studies have made use of Western blotting with antibodies recognizing a specific protein of interest to gauge glycosylation occupancy. However, Western blotting is inherently limited to analysis of proteins for which specific antisera are available, and is constrained to low-throughput assays. Western blotting can also only identify protein-wide glycosylation occupancy, and cannot distinguish between partial glycosylation at different Asn residues on the same protein. Mass spectrometry can overcome both of these key difficulties, as it is a general analysis tool that can be used for site-specific analysis of protein glycosylation.

5.3. Glycoconjugate enrichment strategies

Detection of glycosylation at a specific site is the first step in its quantitative analysis (Schulz *et al.*, 2012). Enrichment of glycoproteins or glycopeptides is key to the success of high sensitivity detection of glycosylation sites. Various enrichment strategies can be employed depending on the biological system of interest, and the analytes of interest within that system. The physical properties of carbohydrates that distinguish them from protein can be used to enrich glycopeptides and glycoproteins. Typical enrichment strategies based on the physical properties of glycans include hydrophilic interaction chromatography (Mysling *et al.*, 2010, Gilar *et al.*, 2011, Christiansen *et al.*, 2010), phenyl boronic acid (Li *et al.*, 2000, Li *et al.*, 2001) and hydrazide (Zhang & Aebersold, 2006, Zhang *et al.*, 2003) attachment. A key mechanism mediating the functional roles of glycans in many biological systems is recognition of specific glycan structures by proteins, or lectins. The specificity of such lectins for defined glycan structures can be used to enrich particular subsets of glycopeptides or glycoproteins bearing those structures (Drake *et al.*, 2006, Zielinska *et al.*, 2010).

5.4. Mass spectrometry for measuring glycosylation occupancy

To obtain quantitative or semi-quantitative measurement of the extent of glycosylation at that site subsequent comparison must be made with the unglycosylated form of the detected peptide. This can be done using comparison of ion intensities of the glycosylated and unglycosylated peptides. The unglycosylated form of the peptide will only be present in one form. However, as glycosylation generally results in a complex mixture of glycan structures at each glycosylation site, measurement of the abundance of the glycosylated form of a given site is not trivial. Some approaches have used detection of entire glycopeptides, although this approach generally requires more specialized and targeted LC-MS technologies (Sumer-Bayraktar *et al.*, 2011). Other approaches have focused on improving quantification of occupancy, and have discarded information on site-specific glycan structure by endoglycosidase treatment (Schulz & Aebi, 2009). For instance, PNGaseF cleaves N-glycans and converts previously glycosylated Asn to Asp, while EndoH leaves a single *N*-acetylglucosamine on previously glycosylated Asn residues. In both of these cases, there is a clear mass and retention shift readily detectable with modern LC-MS technologies, that can be used to differentiate and independently measure the glycosylated and unglycosylated forms of a given peptide.

5.5. Selected-reaction-monitoring mass spectrometry

Recent years have seen impressive success with targeted mass spectrometry approaches, using selected-reaction-monitoring (Lange *et al.*, 2008, Gallien *et al.*, 2011). N-glycosylation has been used as a useful tag to specifically enrich otherwise low abundant components of biological fluids (Stahl-Zeng *et al.*, 2007). Often this has been performed not out of direct interest in glycosylation per se, but because of the ubiquity of glycosylation, and its proven utility in biomarker discovery. However, some analyses have used this approach to specifically measure glycosylation occupancy, for instance in patients with congenital disorders of glycosylation (Hülsmeier *et al.*, 2007).

5.6. Future analytical directions

Use of tools such as those outlined above, in combination with experimental manipulation of growth conditions, N-glycan biosynthetic pathways, protein translation and translocation, and OTase function or composition, will allow identification of the regulation and roles of site-specific N-glycosylation occupancy at a systems level.

6. Is the 'glycosylation sequon' an example of convergent evolution?

Insights into glycosylation site evolution

6.1. HMW-ABC glycosylation in non-typeable *Haemophilus influenzae*

A family of cytoplasmic bacterial enzymes have been recently described that catalyse an N-glycosylation reaction remarkably reminiscent of 'traditional' N-glycosylation. These

enzymes are the HMW-C glycosyltransferase of non-typeable *Haemophilus influenzae* (NTHi), and related organisms primarily in the Pasteurellaceae. NTHi can be a commensal resident of the nasopharynx in humans, but in mixed cultures with *Streptococcus pneumoniae* and *Moraxella catarrhalis* is a primary cause of middle ear infections (otitis media) which in developed countries is the most common reason for children to visit doctors and for antibiotic prescriptions (Murphy *et al.*, 2009). Chronic otitis media is common in indigenous communities worldwide, and can lead to hearing loss with subsequent learning difficulties. NTHi is also associated with chronic obstructive pulmonary disease (COPD), a major burden on health care systems worldwide (Murphy, 2006). Understanding of the role of glycosylation of the outer membrane protein components in these organisms is likely of key importance in the development of effective vaccines against NTHi infection.

A key step in NTHi infection is adherence to the host epithelium. Surface exposed adhesin proteins mediate this adherence, with the high molecular weight (HMW) adhesin system being of key importance in many NTHi clinical isolates. HMW-C is a glycosyltransferase associated with this two-partner secretion system adhesin, encoded in the HMW-ABC locus. Two highly homologous loci are present in the ~80% of NTHi clinical isolates that encode this system, HMW1ABC and HMW2ABC respectively (St. Geme *et al.*, 1998, Ecevit *et al.*, 2004). HMW1A encodes an adhesin glycoprotein (Gross *et al.*, 2008, Grass *et al.*, 2003), which is secreted across the inner membrane via the Sec apparatus, and requires the outer membrane protein HMW1B for correct export across the outer membrane (St Geme & Yeo, 2009). HMW1C encodes a family 41 glycosyltransferase that glycosylates HMW1A (Grass *et al.*, 2010, Kawai *et al.*, 2011, Choi *et al.*, 2010). This glycosylation is required for stability, efficient folding and secretion of the HMW1A glycoprotein adhesin (Grass *et al.*, 2003). In turn, the HMW1A adhesin is important for NTHi colonisation and pathogenesis (St Geme *et al.*, 1993, St Geme, 1994).

Similar to several other described bacterial protein glycosyltransferases, HMW1C glycosylates its HMW1A substrate protein in the cytoplasm, before secretion across the inner membrane (Fleckenstein *et al.*, 2006, Charbonneau *et al.*, 2012, Choi *et al.*, 2010, Schwarz *et al.*, 2011a). Most of these other reported bacterial glycosyltransferases are O-glycosyltransferases, transferring nucleotide-activated monosaccharides to the hydroxyl groups of Ser or Thr. In contrast, HMW1C glycosylates Asn residues, with a strong tendency to glycosylate Asn within glycosylation sequons with the sequence Asn-Xaa-Ser/Thr (Gross *et al.*, 2008).

6.2. HMW-C versus OTase: Unrelated enzymes, same sequon?

The HMW-C and OTase systems are not homologous, as traditional N-glycosylation as described above is catalysed by the integral membrane OTase, which transfers an oligosaccharide from a lipid linked carrier to nascent polypeptide in the lumen of the ER (or periplasm). In contrast, the HMW-C cytoplasmic system of some bacteria is catalysed by a soluble glycosyltransferase that transfers a nucleotide-activated monosaccharide to protein in the cytoplasm. However, it is striking that the bacterial cytoplasmic HMW-C enzymes have very similar site recognition to 'traditional' OTase enzymes: they efficiently glycosylate Asn in

'sequons' with Asn-Xaa-Ser/Thr (Xaa≠Pro), but are capable of glycosylating some selected asparagines lacking S/T at the +2 position (Choi et al., 2010, Grass et al., 2010, Schwarz et al., 2011a, Schwarz & Aebi, 2011). HMW-C enzymes also share the substrate requirement of OTase for unfolded protein, or flexible loops in folded protein (Schwarz et al., 2011a).

A high-resolution 3D crystal structure of an HMW-C enzyme from *Actinobacillus pleuropneumoniae* has been reported (Kawai et al., 2011), which is highly similar to HMW-C enzymes from NTHi. This structure was solved in the presence of acceptor peptide substrate, but electron density for the peptide was not visible in the structure. Nonetheless, the ability of HMW-C enzymes to glycosylate asparagines not in Asn-Xaa-Ser/Thr glycosylation sequons strongly suggests that this sequence is not directly involved in catalysis. It is likely that, as with OTase, this local sequence requirement is instead involved in increasing the affinity for acceptor polypeptide binding, and in substrate recognition.

This then raises the very curious observation that two non-homologous enzymatic systems for glycosylation of Asn have independently evolved essentially identical substrate recognition motifs. This suggests convergent evolution of enzyme-substrate interactions in these two systems, which would in turn imply that there is some functional benefit for site recognition of Ser/Thr at +2 amino acid residues of an Asparagine. It is tempting to speculate that this sequence may have evolved to balance the need for sufficient binding affinity of the polypeptide acceptor with the advantages of a general glycosylation system.

The selection pressure for OTase and HMW-C to require unfolded polypeptide substrate is not completely clear. However, this requirement is likely due to the benefit of glycoosylation in increasing both protein folding efficiency and the stability of folded proteins. Addition of glycans to already folded proteins can serve to increase their stability, potentially in a regulated manner (Yuzwa *et al.*, 2012). However, to be of assistance in protein *folding*, glycans must be transferred to proteins while they are unfolded or still in the process of folding. This would then lead to the requirement of binding of flexible polypeptide acceptor to OTase/HMW-C, which in turn limits recognition to stretches of amino acid sequence, rather than a more complex folded protein structural motif. Binding of a single Asn residue in a stretch of unfolded polypeptide would be the smallest possible recognition motif, but hydrogen bonding and van der Waals interactions would likely not be sufficient for binding of adequate affinity to allow efficient glycan transfer. A solution to allow increased affinity would be to increase the length of the recognition sequence, and apparently a 3 or 5 amino acid sequence is sufficient, as shown by the Asn-Xaa-Ser/Thr and Asp/Glu-Xaa-Asn-Xaa-Ser/Thr OTase glycosylation sequons of eukaryotes and bacteria. This is evidenced in the structure of bacterial PglB (STT3), in which most of the length of the extended sequon is involved in direct contacts with the OTase peptide binding site (Lizak *et al.*, 2011a).

6.3. Why is the sequon as it is?

Why then should Ser/Thr be part of a preferred glycosylation recognition sequence, and not any other amino acids? Perhaps part of the answer is that these hydroxyl-containing residues are typically surface exposed and are not charged. The hydrophilic nature of Ser and Thr

means that they are generally not present internally in folded proteins, but are almost always surface exposed. As addition of a glycan in the hydrophobic core of a protein would be incompatible with correct protein folding, a hydrophilic recognition motif is necessary. Charged residues (His, Arg, Lys, Asp, Glu) would also be potential candidates for such a role, but here the generality of the neutral hydroxyl groups of Ser and Thr is perhaps important. Neutral hydrophilic residues such as Ser and Thr are compatible with almost any position on the surface of a folded protein. In contrast, charge-based attraction and repulsion is an important contributor to protein folding, stability and function. Point mutation to insert one of these charged amino acids on the surface of a protein is likely to disrupt the protein structure. Ser/Thr as an extended recognition sequence therefore likely provides the affinity and ubiquity necessary for evolution of OTase/HMW-C enzymes as general glycosylation enzymes capable of glycosylating multiple Asn residues in many different proteins.

7. Conclusion

The structural basis for the glycosylation sequon is now apparent. However, it is also clear that recognition and glycosylation of selected asparagine residues is subject to further control and regulation depending on variation within catalytic STT3 enzymes, and on the presence of accessory protein subunits of multiprotein OTase complexes. In order to understand the roles of these accessory proteins, it is however necessary for them to be completely identified. With recent years showing the identification and preliminary characterization of several novel accessory proteins of mammalian OTase, it is probable that additional subunits still remain to be discovered. Biochemical characterization of OTase complexes in other eukaryotes may well also present additional, non-homologous, accessory protein subunits. Further, OTase enzymatic activity is actively regulated, adding to the complexity of potential OTase function. Mass spectrometry-based future analytics for glycosylation analysis will enable phenotypic characterization of the site-specific activity of OTase in these varied biological circumstances. Such analysis will contribute to, and also benefit from, a complete quantitative understanding of the interplay between glycoprotein folding and N-glycosylation. Finally, understanding of the molecular mechanisms of N-glycosylation site selection is beginning to open the possibilities for co-engineering of glycosylation sites and OTases in synthetic biology approaches outside of natural evolutionary constraints, moving N-glycosylation beyond the sequon.

Author details

Benjamin Luke Schulz

The University of Queensland, School of Chemistry and Molecular Biosciences, St Lucia, QLD, Australia

Acknowledgement

The author acknowledges the support of NHMRC Career Development Fellowship APP1031542 and NHMRC Project Grant 631615.

8. References

- Aebi, M., R. Bemasconi, S. Clerc & M. Molinari, (2010) N-glycan structures: recognition and processing in the ER. *Trends Biochem Sci* 35: 74-82.
- Bano-Polo, M., F. Baldin, S. Tamborero, M. A. Marti-Renom & I. Mingarro, (2011) N-glycosylation efficiency is determined by the distance to the C-terminus and the amino acid preceding an Asn-Ser-Thr sequon. *Protein Sci* 20: 179-186.
- Burda, P. & M. Aebi, (1999) The dolichol pathway of N-linked glycosylation. *Biochim Biophys Acta* 1426: 239-257.
- Charbonneau, M. E., J. P. Cote, M. F. Haurat, B. Reiz, S. Crepin, F. Berthiaume, C. M. Dozois, M. F. Feldman & M. Mourez, (2012) A structural motif is the recognition site for a new family of bacterial protein O-glycosyltransferases. *Mol Microbiol*.
- Chi, Y. H., Y. D. Koo, S. Y. Dai, J. E. Ahn, D. J. Yun, S. Y. Lee & K. Zhu-Salzman, (2010) N-glycosylation at non-canonical Asn-X-Cys sequence of an insect recombinant cathepsin B-like counter-defense protein. *Comp Biochem Physiol B Biochem Mol Biol* 156: 40-47.
- Choi, K. J., S. Grass, S. Paek, J. W. St Geme, 3rd & H. J. Yeo, (2010) The *Actinobacillus pleuropneumoniae* HMW1C-like glycosyltransferase mediates N-linked glycosylation of the *Haemophilus influenzae* HMW1 adhesin. *PLoS One* 5: e15888.
- Christiansen, M. N., D. Kolarich, H. Nevalainen, N. H. Packer & P. H. Jensen, (2010) Challenges of determining O-glycopeptide heterogeneity: a fungal glucanase model system. *Anal Chem* 82: 3500-3509.
- Dempski, R. E. J. & B. Imperiali, (2002) Oligosaccharyl transferase: gatekeeper to the secretory pathway. *Curr Opin Chem Biol* 6: 844-850.
- Drake, R. R., E. E. Schwegler, G. Malik, J. Diaz, T. Block, A. Mehta & O. J. Semmes, (2006) Lectin Capture Strategies Combined with Mass Spectrometry for the Discovery of Serum Glycoprotein Biomarkers. *Mol Cell Proteomics* 5: 1957-1967.
- Ecevit, I. Z., K. W. McCrea, M. M. Pettigrew, A. Sen, C. F. Marrs & J. R. Gilsdorf, (2004) Prevalence of the hifBC, hmw1A, hmw2A, hmwC, and hia Genes in *Haemophilus influenzae* Isolates *J Clin Micro* 42: 3065-3072.
- Fetrow, J. S., N. Siew, J. A. Di Gennaro, M. Martinez-Yamout, H. J. Dyson & J. Skolnick, (2001) Genomic-scale comparison of sequence- and structure-based methods of function prediction: does structure provide additional insight? *Protein Sci* 10: 1005-1014.
- Fleckenstein, J. M., K. Roy, J. F. Fischer & M. Burkitt, (2006) Identification of a Two-Partner Secretion Locus of Enterotoxigenic *Escherichia coli*. *Infect. Immun.* 74: 2245-2258.
- Gallien, S., E. Duriez & B. Domon, (2011) Selected reaction monitoring applied to proteomics. *J Mass Spectrom* 46: 298-312.
- Gilar, M., Y. Q. Yu, J. Ahn, H. Xie, H. Han, W. Ying & X. Qian, (2011) Characterization of glycoprotein digests with hydrophilic interaction chromatography and mass spectrometry. *Anal Biochem* 417: 80-88.
- Grass, S., A. Z. Buscher, W. E. Swords, M. A. Apicella, S. J. Barenkamp, N. Ozchlewski & J. W. St Geme, 3rd, (2003) The *Haemophilus influenzae* HMW1 adhesin is glycosylated in a process that requires HMW1C and phosphoglucomutase, an enzyme involved in lipooligosaccharide biosynthesis. *Molecular Microbiology* 48: 737-751.

- Grass, S., C. F. Lichti, R. R. Townsend, J. Gross & J. W. St Geme, 3rd, (2010) The Haemophilus influenzae HMW1C protein Is a glycosyltransferase that transfers hexose residues to asparagine sites in the HMW1 adhesin *PLoS Pathogens* 6: e1000919.
- Gross, J., S. Grass, A. E. Davis, P. Gilmore-Erdmann, R. R. Townsend & J. W. St Geme, 3rd, (2008) The Haemophilus influenzae HMW1 Adhesin Is a Glycoprotein with an Unusual N-Linked Carbohydrate Modification. *J Biol Chem* 283: 26010-26015.
- Harada, Y., H. Li, H. Li & W. J. Lennarz, (2009) Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site. *Proc Natl Acad Sci U S A* 106: 6945-6949.
- Helenius, A. & M. Aebi, (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem* 73: 1019-1049.
- Hülsmeier, A. J., P. Paesold-Burda & T. Hennet, (2007) N-glycosylation site occupancy in serum glycoproteins using multiple reaction monitoring liquid chromatography mass spectrometry. *Mol Cell Proteomics* 6: 2132-2138.
- Imperiali, B. & T. L. Hendrickson, (1995) Asparagine-linked glycosylation: specificity and function of oligosaccharyl transferase. *Bioorg Med Chem* 3: 1565-1578.
- Izquiedro, L., A. Mehlert & M. A. Ferguson, (2012) The lipid-linked oligosaccharide donor specificities of Trypanosoma brucei oligosaccharyltransferases. *Glycobiology* 22: 696-703.
- Izquiedro, L., B. L. Schulz, J. A. Rodrigues, M. L. Güther, J. B. Procter, G. J. Barton, M. Aebi & M. A. Ferguson, (2009) Distinct donor and acceptor specificities of Trypanosoma brucei oligosaccharyltransferases. *EMBO J* 28: 2650-2661.
- Jamaluddin, M. F. B., U. M. Bailey, N. Y. J. Tan, A. P. Stark & B. L. Schulz, (2011) Polypeptide binding specificities of Saccharomyces cerevisiae oligosaccharyltransferase accessory proteins Ost3p and Ost6p. *Protein Sci* 20: 849-855.
- Karaoglu, D., D. J. Kelleher & R. Gilmore, (1995) Functional characterization of Ost3p. Loss of the 34-kD subunit of the Saccharomyces cerevisiae oligosaccharyltransferase results in biased underglycosylation of acceptor substrates. *J Cell Biol* 130: 567-577.
- Karaoglu, D., D. J. Kelleher & R. Gilmore, (2001) Allosteric regulation provides a molecular mechanism for preferential utilization of the fully assembled dolichol-linked oligosaccharide by the yeast oligosaccharyltransferase. *Biochemistry* 40: 12193-12206.
- Kasturi, L., H. Chen & S. H. Shakin-Eshleman, (1997) Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem J* 323: 415-419.
- Kasturi, L., J. R. Eshleman, W. H. Wunner & S. H. Shakin-Eshleman, (1995) The Hydroxy Amino Acid in an Asn-X-Ser/Thr Sequon Can Influence N-Linked Core Glycosylation Efficiency and the Level of Expression of a Cell Surface Glycoprotein. *J. Biol. Chem.* 270: 14756-14761.
- Kawai, F., S. Grass, Y. Kim, K. J. Choi, J. W. St Geme, 3rd & H. J. Yeo, (2011) Structural insights into the glycosyltransferase activity of the Actinobacillus pleuropneumoniae HMW1C-like protein. *J Biol Chem* 286: 38546-38557.
- Kelleher, D. J. & R. Gilmore, (2006) An evolving view of the eukaryotic oligosaccharyltransferase. *Glycobiology* 16: 47R-62R.

- Kelleher, D. J., D. Karaoglu, E. C. Mandon & R. Gilmore, (2003) Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol Cell* 12: 101-111.
- Knauer, R. & L. Lehle, (1999) The oligosaccharyltransferase complex from *Saccharomyces cerevisiae*. Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex. *J Biol Chem* 274: 17249-17256.
- Kowarik, M., S. Numao, M. F. Feldman, B. L. Schulz, N. Callewaert, E. Kiermaier, I. Catrein & M. Aebi, (2006a) N-linked glycosylation of folded proteins by the bacterial oligosaccharyltransferase. *Science* 314: 1148-1150.
- Kowarik, M., N. M. Young, S. Numao, B. L. Schulz, I. Hug, N. Callewaert, D. C. Mills, D. C. Watson, M. Hernandez, J. F. Kelly, M. Wacker & M. Aebi, (2006b) Definition of the bacterial N-glycosylation site consensus sequence. *The EMBO Journal* 25: 1957-1966.
- Lange, V., P. Picotti, B. Domon & R. Aebersold, (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol Syst Biol* 4: 222.
- Lee, J. H., W. H. Yu, A. Kumar, S. Lee, P. S. Mohan, C. M. Peterhoff, D. M. Wolfe, M. Martinez-Vicente, A. C. Massey, G. Sovak, Y. Uchiyama, D. Westaway, A. M. Cuervo & R. A. Nixon, (2010) Lysosomal Proteolysis and Autophagy Require Presenilin 1 and Are Disrupted by Alzheimer-Related PS1 Mutations. *Cell* 141: 1146-1158.
- Li, Y., E. L. Larsson, H. Jungvid, I. Y. Galaev & B. Mattiasson, (2000) Affinity chromatography of neoglycoproteins. *Bioseparation* 9: 315-323.
- Li, Y., U. Pfüller, E. L. Larsson, H. Jungvid, I. Y. Galaev & B. Mattiasson, (2001) Separation of mistletoe lectins based on the degree of glycosylation using boronate affinity chromatography. *J Chromatogr A* 925: 115-121.
- Lizak, C., Y. Y. Fan, T. C. Weber & M. Aebi, (2011a) N-Linked Glycosylation of Antibody Fragments in *Escherichia coli*. *Bioconjug Chem*.
- Lizak, C., S. Gerber, S. Numao, M. Aebi & K. P. Locher, (2011b) X-ray structure of a bacterial oligosaccharyltransferase. *Nature* 474: 350-355.
- Marshall, R. D., (1974) The nature and metabolism of the carbohydrate-peptide linkages of glycoproteins. *Biochem Soc Symp* 40: 17-26.
- Miletich, J. P. & G. J. J. Broze, (1990) Beta protein C is not glycosylated at asparagine 329. The rate of translation may influence the frequency of usage at asparagine-X-cysteine sites. *J Biol Chem* 265: 11397-11404.
- Mohorko, E., R. Glockshuber & M. Aebi, (2011) Oligosaccharyltransferase: the central enzyme of N-linked protein glycosylation. *J Inherit Metab Dis* 34: 869-878.
- Murphy, T. F., (2006) The role of bacteria in airway inflammation in exacerbations of chronic obstructive pulmonary disease. *Curr Opin Infect Dis* 19: 225-230.
- Murphy, T. F., L. O. Bakaletz & P. R. Smeesters, (2009) Microbial interactions in the respiratory tract. *Pediatr Infect Dis J* 28: S121-126.
- Mysling, S., G. Palmisano, P. Højrup & M. Thaysen-Andersen, (2010) Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics. *Anal Chem* 82: 5598-5609.

- Nasab, F. P., B. L. Schulz, F. Gamarro, A. J. Parodi & M. Aebi, (2008) All in One: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. *Mol Biol Cell* 19: 3758-3768
- Nita-Lazar, A., M. Wacker, B. Schegg, S. Amber & M. Aebi, (2005) The N-X-S/T consensus sequence is required but not sufficient for bacterial N-linked protein glycosylation. *Glycobiology* 15: 361-367.
- Oliver, J. D., H. L. Roderick, D. H. Llewellyn & S. High, (1999) ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol Biol Cell* 10: 2573-2582.
- Palmisano, G., M. N. Melo-Braga, K. Engholm-Keller, B. L. Parker & M. R. Larsen, (2012) Chemical deamidation: a common pitfall in large-scale N-linked glycoproteomic mass spectrometry-based analyses. *J Proteome Res* 11: 1949-1957.
- Petrescu, A. J., A. L. Milac, S. M. Petrescu, R. A. Dwek & M. R. Wormald, (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14: 103-114.
- Rao, R. S., O. T. Buus & B. Wollenweber, (2011) Distribution of N-glycosylation sequons in proteins: how apart are they? *Comput Biol Chem* 35: 57-61.
- Robinson, N. E., Z. W. Robinson, B. R. Robinson, A. L. Robinson, J. A. Robinson, M. L. Robinson & A. B. Robinson, (2004) Structure-dependent nonenzymatic deamidation of glutaminy and asparaginy pentapeptides. *J Pept Res* 63: 426-436.
- Roboti, P. & S. High, (2012) Keratinocyte-associated protein 2 is a bona fide subunit of the mammalian oligosaccharyltransferase. *J Cell Sci* 125: 220-232.
- Ruiz-Canada, C., D. J. Kelleher & R. Gilmore, (2009) Cotranslational and Posttranslational N-Glycosylation of Polypeptides by Distinct Mammalian OST Isoforms. *Cell* 136: 272-283.
- Sanyal, S. & A. K. Menon, (2010) Stereoselective transbilayer translocation of mannosyl phosphoryl dolichol by an endoplasmic reticulum flippase *Proc Natl Acad Sci U S A*.
- Sato, C., J. H. Kim, Y. Abe, K. Saito, S. Yokoyama & D. Kohda, (2000) Characterization of the N-oligosaccharides attached to the atypical Asn-X-Cys sequence of recombinant human epidermal growth factor receptor. *J Biochem* 127: 65-72.
- Schulz, B. L. & M. Aebi, (2009) Analysis of Glycosylation Site Occupancy Reveals a Role for Ost3p and Ost6p in Site-specific N-Glycosylation Efficiency. *Mol Cell Proteomics* 8: 357-364.
- Schulz, B. L., C. U. Stirnimann, J. P. A. Grimshaw, M. S. Brozzo, F. Fritsch, E. Mohorko, G. Capitani, R. Glockshuber, M. G. Grütter & M. Aebi, (2009) Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc Natl Acad Sci U S A* 106: 11061-11066.
- Schulz, B. L., J. C. White & C. Punyadeera, (2012) Saliva proteome research: current status and future outlook. *Crit Rev Biotech*.
- Schwarz, F. & M. Aebi, (2011) Mechanisms and principles of N-linked protein glycosylation. *Curr Opin Struc Biol* 21: 576-582.

- Schwarz, F., Y. Y. Fan, M. Schubert & M. Aebi, (2011a) Cytoplasmic N-glycosyltransferase of *Actinobacillus pleuropneumoniae* is an inverting enzyme and recognizes the NX(S/T) consensus sequence. *J Biol Chem* 286: 35267-35274.
- Schwarz, F., C. Lizak, Y. Y. Fan, S. Fleurkens, M. Kowarik & M. Aebi, (2011b) Relaxed acceptor site specificity of bacterial oligosaccharyltransferase in vivo. *Glycobiology* 21: 45-54.
- Schwarz, M., M. Knauer & L. Lehle, (2005) Yeast oligosaccharyltransferase consists of two functionally distinct sub-complexes, specified by either the Ost3p or Ost6p subunit. *FEBS Lett* 579: 6564-6568.
- Shakin-Eshleman, S. H., S. L. Spitalnik & L. Kasturi, (1996) The amino acid at the X position of an Asn-X-Ser sequon is an important determinant of N-linked core-glycosylation efficiency. *J Biol Chem* 271: 6363-6366.
- Shibatani, T., L. L. David, A. L. McCormack, K. Frueh & W. R. Skach, (2005) Proteomic analysis of mammalian oligosaccharyltransferase reveals multiple subcomplexes that contain Sec61, TRAP, and two potential new subunits. *Biochemistry* 44: 5982-5992.
- Spirig, U., D. Bodmer, M. Wacker, P. Burda & M. Aebi, (2005) The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology* 15: 1396-1406.
- St Geme, J. W., 3rd, (1994) The HMW1 adhesin of nontypeable *Haemophilus influenzae* recognizes sialylated glycoprotein receptors on cultured human epithelial cells. *Infect Immun* 62: 3881-3889.
- St Geme, J. W., 3rd, S. Falkow & S. J. Barenkamp, (1993) High-molecular-weight proteins of nontypeable *Haemophilus influenzae* mediate attachment to human epithelial cells. *Proc Natl Acad Sci U S A* 90: 2875-2879.
- St Geme, J. W., 3rd & H.-J. Yeo, (2009) A prototype two-partner secretion pathway: the *Haemophilus influenzae* HMW1 and HMW2 adhesin systems. *Trends in Microbiology* 17: 355-360.
- St. Geme, J. W., III, V. V. Kumar, D. Cutter & S. J. Barenkamp, (1998) Prevalence and Distribution of the hmw and hia Genes and the HMW and Hia Adhesins among Genetically Diverse Strains of Nontypeable *Haemophilus influenzae*. *Infect. Immun.* 66: 364-368.
- Stahl-Zeng, J., V. Lange, R. Ossola, K. Eckhardt, W. Krek, R. Aebersold & B. Domon, (2007) High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* 6: 1809-1817.
- Sumer-Bayraktar, Z., D. Kolarich, M. P. Campbell, S. Ail, N. H. Packer & M. Thaysen-Andersen, (2011) N-glycans modulate the function of human corticosteroid-binding globulin. *Mol Cell Proteomics* 10: M111.009100.
- Szymanski, C. M., R. Yao, C. P. Ewing, T. J. Trust & P. Guerry, (1999) Evidence for a system of general protein glycosylation in *Campylobacter jejuni*. *Molecular Microbiology* 32: 1022-1030.
- Vance, B. A., W. Wu, R. K. Ribaud, D. M. Segal & K. P. Kears, (1997) Multiple dimeric forms of human CD69 result from differential addition of N-glycans to typical (Asn-X-Ser/Thr) and atypical (Asn-X-cys) glycosylation motifs. *J Biol Chem* 272: 23117-23122.

- Wacker, M., D. Linton, P. G. Hitchen, M. Nita-Lazar, S. M. Haslam, S. J. North, M. Panico, H. R. Morris, A. Dell, B. W. Wren & M. Aebi, (2002) N-Linked Glycosylation in *Campylobacter jejuni* and Its Functional Transfer into *E. coli*. *Science* 298: 1790-1793.
- Whitley, P., I. Nilsson & G. von Heijne, (1996) A Nascent Secretory Protein May Traverse the Ribosome/Endoplasmic Reticulum Translocase Complex as an Extended Chain. *J Biol Chem* 271: 6241-6244.
- Wilson, C. M. & S. High, (2007) Ribophorin I acts as a substrate-specific facilitator of N-glycosylation. *J Cell Sci* 120: 648-657.
- Wilson, C. M., C. Kraft, C. Duggan, N. Ismail, S. G. Crawshaw & S. High, (2005) Ribophorin I associates with a subset of membrane proteins after their integration at the sec61 translocon. *J Biol Chem* 280: 4195-4206.
- Wilson, C. M., Q. Roebuck & S. High, (2008) Ribophorin I regulates substrate delivery to the oligosaccharyltransferase core. *Proceedings of the National Academy of Sciences* 105: 9534-9539.
- Wormald, M. R. & R. A. Dwek, (1999) Glycoproteins: glycan presentation and protein-fold stability. *Structure* 7: R155-160.
- Yan, A. & W. J. Lennarz, (2005) Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. *Glycobiology* 15: 1407-1415.
- Yuzwa, S. A., X. Shan, M. S. Macauley, T. Clark, Y. Skorobogatko, K. Vosseller & D. J. Vocadlo, (2012) Increasing O-GlcNAc slows neurodegeneration and stabilizes tau against aggregation. *Nat Chem Biol* 8: 393-399.
- Zhang, H. & R. Aebersold, (2006) Isolation of glycoproteins and identification of their N-linked glycosylation sites. *Methods Mol Biol* 328: 177-185.
- Zhang, H., X. J. Li, D. B. Martin & R. Aebersold, (2003) Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol* 21: 660-666.
- Zielinska, D. F., F. Gnad, J. R. Wisniewski & M. Mann, (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141: 897-907.