Daniel J. Kelleher and Reid Gilmore

Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605-2524

Accepted on November 23, 2005

Asparagine-linked glycosylation (ALG) is one of the most common protein modification reactions in eukaryotic cells, as many proteins that are translocated across or integrated into the rough endoplasmic reticulum (RER) carry N-linked oligosaccharides. Although the primary focus of this review will be the structure and function of the eukaryotic oligosaccharyltransferase (OST), key findings provided by the analysis of the archaeabacterial and eubacterial OST homologues will be reviewed, particularly those that provide insight into the recognition of donor and acceptor substrates. Selection of the fully assembled donor substrate will be considered in the context of the family of human diseases known as congenital disorders of glycosylation (CDG). The yeast and vertebrate OST are surprisingly complex heterooligomeric proteins consisting of seven or eight subunits (Ost1p, Ost2p, Ost3p/Ost6p, Ost4p, Ost5p, Stt3p, Wbp1p, and Swp1p in yeast; ribophorin I, DAD1, N33/IAP, OST4, STT3A/STT3B, Ost48, and ribophorin II in mammals). Recent findings from several laboratories have provided overwhelming evidence that the STT3 subunit is critical for catalytic activity. Here, we will consider the evolution and assembly of the eukaryotic OST in light of recent genomic evidence concerning the subunit composition of the enzyme in diverse eukaryotes.

Key words: dolichol/endoplasmic reticulum/N-linked oligosaccharide/oligosaccharyltransferase

Introduction

In eukaryotic cells, a critical step in asparagine-linked glycosylation (ALG) of proteins is catalyzed by the oligosaccharyltransferase (OST), an integral membrane protein that mediates the en bloc transfer of a preassembled high-mannose oligosaccharide onto asparagine residues of nascent polypeptides entering the lumen of the rough endoplasmic reticulum (RER). The dolichol pyrophosphate-linked oligosaccharide (OS-PP-Dol) donor (Glc3Man9GlcNAc2-PP-Dol in higher eukaryotes) for glycosylation is sequentially assembled by glycosyltransferases located on the cytoplasmic and lumenal faces of the RER membrane. The assembly pathway for dolichol oligosaccharides has been reviewed in detail (Burda and Aebi, 1999), so here we will focus on recent discoveries that impact donor-substrate recognition, selection and catalysis by the OST. Although the OST has been the focus of a several reviews during the past decade (Silberstein and Gilmore, 1996; Knauer and Lehle, 1999; Dempski and Imperiali, 2002; Yan and Lennarz, 2005), recent developments in this field make this review timely.

N-Linked glycosylation of nascent glycoproteins is temporally coupled to the protein translocation reaction that occurs as, or immediately after, the polypeptide is synthesized. The coordination of N-linked glycosylation and protein translocation necessitates a spatial localization of the OST adjacent to the protein translocation channel. Even though the OST has unrestricted access to nascent polypeptides entering the ER lumen, a surprising percentage of consensus glycosylation sites are not modified in vivo. Here, we review the effect of acceptor peptide sequence as well as sequence context factors that influence N-linked glycosylation of N-X-T/S sites in vivo and in vitro.

The elucidation of the roles of the noncatalytic OST subunits remains an important issue. In this article, we review the biochemical and genetic characterization of the yeast and mammalian OST subunits, in the context of emerging information about the evolution of the OST that is being deduced from genomics data. Potential roles for several subunits can now be proposed based upon the available data and upon the differences in the enzymatic properties of the OST in diverse organisms.

Ancient origin of N-linked glycosylation

N-Linked glycosylation of proteins was initially thought to be restricted to eukaryotic organisms. The first evidence that this was not the case came from the structural analysis of alkali-insensitive glycopeptides derived from Halobacterium cell surface glycoprotein (CSG) and flagellins (for a review, see Lechner and Wieland, 1989). These Halobacterium salinarum proteins are modified with tetrasaccharides composed of glucose and sulfated glucuronic and iduronic acids that are linked to asparagine residues through glucose (Wieland et al., 1983, 1985). The glycosylated asparagines are in an N-X-T/S motif that matches the sequon for N-linked glycosylation in eukaryotic organisms (Wieland et al., 1985). Furthermore, a dolichol-phosphate-linked tetrasaccharide serves as the oligosaccharide donor for N-linked glycosylation in H. salinarum (Lechner et al.,
A second glycosaminoglycan-like oligosaccharide is attached to the *H. salinarum* CSG through an asparaginyl-N-acetylgalactosamine linkage at an N-A-S site (Paul et al., 1986) using a dolichol-pyrophosphate-linked oligosaccharide as a donor (Lechner and Wieland, 1989). In contrast to what we know about the acceptor-substrate specificity for the eukaryotic OST, addition of the glycosaminoglycan-like oligosaccharide to CSG still occurred when the serine in the N-A-S sequon was replaced with a valine or leucine residue (Zeitler et al., 1998). Thus, in *H. salinarum*, two structurally different N-linked oligosaccharides are attached to N-X-S/T sites by what are likely distinct enzymes based upon differences in donor substrate structure and acceptor substrate specificity. The flagellins of *Methanococcus voltae* are N-glycosylated at multiple N-X-T/S sites with a trisaccharide through an asparaginyl-β-GlcNAc linkage (Voisin et al., 2005). The most abundant oligosaccharide donor in *Haloferax volcanii* is mannosyl-(β1→4)galactosyl phosphodolichol (Kuntz et al., 1997). Taken together, it appears that donor substrates for N-glycosylation in archaeabacterial organisms are structurally diverse, yet are assembled upon dolichol phosphate or dolichol pyrophosphate carriers.

Recent evidence indicates that several eubacteria, including *Campylobacter jejuni*, have acquired a 14-gene pgl (protein glycosylation) operon that is involved in N-linked glycosylation of cell surface proteins. Protein sequence analysis of the pgl operon (Szymanski et al., 1999) revealed that PglB is homologous to the STT3 subunit of the eukaryotic OST. Mutagenesis of the pglB locus and chemical deglycosylation of C. jejuni membrane proteins led to the conclusion that the pgl operon is involved in protein glycosylation (Szymanski et al., 1999). We refer the reader to a recent review of eubacterial glycosylation for more information concerning the roles of the pgl gene products in the assembly pathway of the undecaprenyl-pyrophosphate-linked oligosaccharide donor for N-linked glycosylation in eubacteria (Szymanski and Wren, 2005).

### Oligosaccharide donors for the eukaryotic OST

In eukaryotic organisms, the oligosaccharide donor for N-linked glycosylation of proteins is assembled on the carrier lipid dolichol pyrophosphate by the sequential addition of the monosaccharides GlcNAc, Man, and Glc (Figure 1). Chemically synthesized donor substrate analogues that replace dolichol-PP with phytanyl-PP, dihydrofarnesyl-PP, or citronellyl-PP are not substrates for the OST (Fang et al., 1995), but polyisoprenol-PP-linked oligosaccharides are utilized as donors in the Chinese hamster ovary cell line B211 (Kaiden et al., 1998).

The assembly pathway for Glc₃Man₉GlcNAc₂-PP-Dol has been deduced from the analysis of intermediates that accumulate in vitro in membranes from wild-type cells (Chapman et al., 1979; Liu et al., 1979) and by the analysis of the yeast *alg* mutants that lack glycosyltransferase activities (Huffaker and Robbins, 1983; Burda and Aebl, 1999). Most of the glycosyltransferases (ALG proteins) that mediate Glc₃Man₉GlcNAc₂-PP-Dol assembly have now been identified (Figure 1). OS-PP-Dol assembly initiates on the cytoplasmic face of the RER, where sugar nucleotides (UDP-GlcNAc and GDP-Man) serve as the saccharide donors for the addition of the first seven residues (Figure 1, red symbols) to yield Man₃GlcNAc₂-PP-Dol. The initial step in OS-PP-Dol assembly is catalyzed by *N*-acetylglucosaminyl phosphate transferase (ALG7) that forms the phosphodiester linkage between GlcNAc-1-phosphate (derived from UDP-GlcNAc) and dolichol-monophosphate.
The most recently identified glycosyltransferase in the OS-PP-Dol assembly pathway is the ALG13–14 heterodimer that mediates the addition of the second GlcNAc residue to form Dol-PP-GlcNAc2 (Bickel et al., 2005; Chantret et al., 2005; Gao et al., 2005). Dol-PP-GlcNAc2 is the in vivo oligosaccharide donor for N-linked glycosylation in *Giardia lamblia* (Samuelsou et al., 2005). The yeast or vertebrate OST will utilize Dol-PP-GlcNAc2 as a donor substrate in an in vitro assay (Sharma et al., 1981; Imperiali and Shannon, 1991). The importance of the two N-acetyl moieties in Dol-PP-GlcNAc2 has been investigated using several synthetic dolichol-linked oligosaccharides (Tai and Imperiali, 2001). Dol-PP-GlcNAc-Glc, like the minimal in vivo donor Dol-PP-GlcNAc2, is a functional donor for the OST, indicating that the 2'-N-acetyl modification of the second sugar is not critical for OST recognition and catalysis. Although glycopeptide products bearing a single GlcNAc were detected by mass spectroscopy when Dol-PP-GlcNAc was tested as the donor, Dol-PP-GlcNAc is clearly a very poor donor substrate for the OST (Bause et al., 1995; Tai and Imperiali, 2001). Dol-PP-2-deoxyfluoroglucose-GlcNAc and Dol-PP-N-trifluoroacetylglucosamine-GlcNAc are not substrates but are instead OST inhibitors. These results demonstrate that the N-acetyl modification on the proximal sugar is critical for OST catalysis, whereas the presence and structure of a second sugar residue (GlcNAc > Glc > no sugar) increases enzyme activity (Tai and Imperiali, 2001).

Until recently, it was not clear whether Alg2 transfers residue 4 (an α-1,3 linked mannose) or residue 5 (an α-1,6 linked mannose). Analysis of human ALG2 mutant fibroblasts (congenital disorders of glycosylation-Ii [CDG-Ii]) demonstrated that Alg2 is an α-1,3 mannosyltransferase (Thiel et al., 2003), which provides the substrate for addition of residue 5 by a currently unidentified mannosyltransferase (Figure 1). Man₉GlcNAc₂-PP-Dol accumulates as a major lipid-linked oligosaccharide in yeast *alg11Δ* cells providing evidence that Alg11p is an α-1,2 mannosyltransferase responsible for the addition of residue 6 (Helenius et al., 2002). Although structural analysis of the protein-linked oligosaccharides synthesized by *alg11Δ* cells had suggested that Alg11p adds residue 7 (Cipollo et al., 2001), the latter analysis is complicated by elongation of the *alg11Δ* assembly intermediate by luminal mannosyltransferases. Given that both residues 6 and 7 are α-1,2 linked mannose, we speculate that a single enzyme (Alg11p) acts sequentially to mediate both additions. Support for this hypothesis is provided by the absence of a candidate GDP-Man-dependent mannosyltransferase of unknown function encoded by the yeast genome. Man₉GlcNAc₂-PP-Dol is then flipped from the cytoplasmic face of the ER into the lumen in a reaction that is dependent upon Rft1 in vivo (Helenius et al., 2002).

The dolichol monosaccharides Dol-P-Man and Dol-P-Glc serve as the sugar donors for the three luminal mannosyltransferases (ALG3, ALG12, and ALG9) and three glucosyltransferases (ALG6, ALG8, and ALG10) that sequentially transfer residues 8–14 onto the OS-PP-Dol (Figure 1). The ordered addition of the sugar residues is maintained by the remarkable substrate specificity of the glycosyltransferases. For example, the predominant OS-PP-Dol compound synthesized by the *alg3Δ* yeast mutant is Man₉GlcNAc₂-PP-Dol not Glc₃Man₉GlcNAc₂-PP-Dol or Glc₃Man₉GlcNAc₂-PP-Dol (Verostek et al., 1993; Aebi et al., 1996; Kelleher et al., 2001) indicating that Man₉GlcNAc₂-PP-Dol is a poor substrate for the Alg6 and Alg12 enzymes. Additional insight into this stepwise assembly pathway has been provided by the discovery that Alg9 transfers two α-1,2 linked mannose residues (Figure 1, residues 9 and 11). The first addition yields the Man₉GlcNAc₂-PP-Dol intermediate (Burda et al., 1996; Cipollo and Trimble, 2000) that is the substrate for Alg12. Alg9 then adds the ninth and final mannose residue to oligosaccharide donor (Frank and Aebi, 2005).

**Defects in donor assembly are responsible for CDG-I**

OS-PP-Dols lacking the terminal glucose residue are less effective donors for the vertebrate and yeast OST both in vitro and in vivo (Turco et al., 1977; Trimble et al., 1980; Burda and Aebi, 1998; Karaoglu et al., 2001). Enzymatic defects that interfere with the biosynthesis of Glc₃Man₉GlcNAc₂-PP-Dol are responsible for a family of diseases referred to as CDG-I (for a recent review of CDG, see Freeze and Aebi, 2005).

Reductions in phosphomannose isomerase activity (PMI; CDG-Ib), phosphomannomutase activity (PMM, CDG-Ia), and Dol-P-Man synthase activity (CDG-Ie) reduce the biosynthesis of the mannose donors (GDP-mannose and Dol-P-Man), thereby causing an accumulation of diverse assembly intermediates (as reviewed in Aebi and Hennet, 2001). The role of the vertebrate LEC35 gene product in OS-PP-Dol assembly was unclear as Lec35 mutants accumulate Man₉GlcNAc₂-PP-Dol (Lehrman and Zeng, 1989) but are not defective in Alg3 activity and have normal levels of Dol-P-Man. Lec35 cells were recently shown to be defective in all aspects of Dol-P-Man and Dol-P-Glc utilization (Anand et al., 2001), hence the LEC35 gene product is now referred to as mannose-P-dolichol utilization (MPDU). Deficiencies in human MPDU are responsible for CDG-If (Kranz et al., 2001; Schenk et al., 2001).

Several recently identified CDG-I subtypes (Figure 1, subtypes Ig, Ih, and Il) cause the accumulation of an OS-P-PP-Dol assembly intermediate that is a poor substrate for the OST (Chantret et al., 2002; Frank et al., 2004; Schollen et al., 2004). CDG-Ij, Ik, and Ii are caused by mutations that reduce the activity of early acting glycosyltransferases (ALG7, ALG1, and ALG2), thereby reducing the steady-state pool of Glc₃Man₉GlcNAc₂-PP-Dol that can serve as the oligosaccharide donor (Thiel et al., 2003; Wu et al., 2003; Grubemann et al., 2004).

**Acceptor substrates and peptide inhibitors of the OST**

Sequencing of N-glycopeptides derived from eukaryotic glycoproteins established that N-glycosylated asparagine residues are within an N-X-T/S consensus sequon, where X can be any residue except proline (Marshall, 1972). Provided that the N- and C-termini are blocked by reagents that mimic peptide bonds, N-X-T tripeptides are glycosylated by the OST in vitro (Welply et al., 1983). N-Glycosylated N-X-C sites are present in human von Willebrand factor and serum protein C (Titani et al., 1986; Mileitch and Broze,
In vitro assays using synthetic peptide substrates with internal N-X-T/S/C sites showed that the kinetic parameters of the OST (peptide $K_m$ and $V_{max}$) are strongly influenced by the side chain in the +2 position relative to asparagine, with N-X-T serving as a better substrate than N-X-S, which is in turn much better than N-X-C (Bause, 1984; Breuer et al., 2001). Peptides containing asparagine analogues have been tested as potential OST substrates and inhibitors. As expected from the sequence of natural glycopeptides, the sequon analogues Q-X-T and D-X-T are neither substrates nor high-affinity inhibitors (Welply et al., 1983; Bause et al., 1995). Of more than a dozen tested peptides wherein asparagine is replaced by an amino acid analogue (Bause et al., 1998), β-hydroxysperagine was found to yield a low affinity substrate, whereas diaminobutanoic acid (Amb) yielded a promising lead compound for OST inhibitors (Bause et al., 1995), as discussed below.

The yeast and vertebrate OST have a surprisingly low affinity ($K_m = 10–30 \mu M$) for a synthetic tripeptide substrate (e.g., N-Ac-N-$^{125}$I-T-NH$_2$; Karaougli et al., 2001; Kelleher et al., 2003). Hydrophobic N-terminal blocking groups (N-benzoyl and N-octanoyl) enhance peptide-binding affinity relative to an N-acetyl blocking group, whereas bulky N-terminal blocking groups (N-t-butoxycarbonyl or N-4-benzoyl-benzyl) reduce peptide-binding affinity (Welply et al., 1983; Xu T., Werner M., et al., 1998). Together with the effect of X-residue substitutions within an N-X-S sequon (Shakin-Eshleman et al., 1996), these observations indicate that residues other than asparagine and the hydroxy amino acid (S or T) can facilitate acceptor-substrate binding to the OST active site.

Two strategies have been used to obtain high-affinity substrates and more potent OST inhibitors. Conformationally restrained peptides that adopt β-turn or asparagine-turn conformations have been tested as OST substrates, and it was found that peptides that adopt the latter, but not former, conformation are OST substrates (Imperiali et al., 1992). Cyclization of synthetic OST substrates has been achieved by covalently linking an N-terminal blocking group on asparagine to the cysteine side chain in a N-C-T sequon to obtain the cyclo(hex-Asn-Cys)-Thr series of substrates or to an Amb-C-T tripeptide to obtain the cyclo(hex-Amb-Cys)-Thr series of OST inhibitors (Hendrickson et al., 1996; Kellenberger et al., 1997). Peptide cyclization appears to cause a relatively modest increase in the binding affinity relative to noncyclic-control peptides that have a hydrophobic, N-terminal blocking group (Kellenberger et al., 1997; Peluso et al., 2002). In contrast, a dramatic increase in the binding affinity ($K_i < 50 \text{ nM}$) of OST inhibitors has been achieved by extending the C-terminus of the peptides by several residues to obtain cyclo(hex-Amb-Cys)-Thr-Val-Thr-Nph-NH$_2$, where Nph corresponds to para-nitrophenylalanine (Hendrickson et al., 1996). Peptide extensions that included one or more charged amino acids (e.g., K-K-Nph-NH$_2$, E-K-Nph-NH$_2$, or V-E-Nph-NH$_2$) yielded markedly less-effective inhibitors or substrates (Kellenberger et al., 1997). The design of the Val-Thr-Nph-NH$_2$ extension was based upon a careful reexamination (Imperiali and colleagues) of the statistical frequency of amino acids at the +3 and +4 residues adjacent to utilized and nonutilized N-X-T sites (Gavel and Von Heijne, 1990).

**Context factors that influence N-linked glycosylation of sequons**

It has been estimated that ~35% of the potential glycosylation sites in proteins that enter the ER lumen are not modified (Petrescu et al., 2004), and others are modified with low efficiency. Sequence analysis of a large collection of utilized and nonutilized N-X-T/S sites in proteins that enter the secretory pathway showed that a proline residue in the +3 site relative to asparagine reduces the statistical frequency of N-glycosylation (Gavel and Von Heijne, 1990). Hexapeptides with internal N-P-T sites or N-X-T-P sites are neither substrates nor low-affinity inhibitors (Bause, 1983). The statistical analysis of utilized N-glycosylation sites also supports the preference for N-X-T sites relative to N-X-S sites (Gavel and Von Heijne, 1990). The influence of amino acids in the X position of N-X-T and N-X-S sites has been tested in the context of the rabies virus glycoprotein. Although the identity of the X-residue had little effect upon the glycosylation efficiency of N-X-T sites, several residues (Glu, Asp, Trp, and Leu) significantly reduced the glycosylation efficiency of N-X-S sites (Shakin-Eshleman et al., 1996). Analysis of several proteins with overlapping sequons (N-N-T/S-T/S) has shown that a single glycan is attached to such a site (Lockridge et al., 1987; Reddy et al., 1988, 1999). Recent results support the conclusion that an N-linked glycan at an asparagine residue reduces access of a closely spaced sequon to the OST active site (Karamysh et al., 2005).

Potential glycosylation sites that are located <12–14 residues from a transmembrane (TM) span are not utilized by the vertebrate OST, indicating that the OST active site is ∼30–40 Å above the luminal membrane surface of the RER (Nilsson and von Heijne, 1993). Not only does this distance restraint on glycosylation of membrane proximal N-X-T/S sites explain a lack of glycosylation of some integral membrane proteins, it is relevant to the glycosylation scanning mutagenesis procedure that has been used to elucidate the topology of membrane proteins (Hresko et al., 1994). Sequons inserted into short luminal loops are not glycosylated (Popov et al., 1997). Glycosylation sites near the C-terminus of proteins are also less likely to be utilized (Gavel and Von Heijne, 1990), presumably because chain termination allows the protein to move more rapidly past the OST active site. Glycosylation of N-X-T/S sites can also be reduced by a disulfide bond involving a nearby cysteine residue (Allen et al., 1995) or delayed or inhibited if the N-X-T/S site is near the signal peptidase cleavage site (Chen et al., 2001). Mutations that prevent folding of a glycoprotein may be accompanied by N-glycosylation at sequons that are not normally modified (McGinness and Morrison, 1994), indicating that rapid protein folding may be responsible for inefficient modification of some N-X-T/S sites. A detailed analysis of the extent of glycosylation at each of the 14 sites in yeast invertase suggests that additional sequence context factors may influence the efficiency of modification of individual glycosylation sites (Reddy et al., 1988).
Subunit composition of the OST in eukaryotic organisms

The OST has been most extensively investigated in the yeast (Saccharomyces cerevisiae) and vertebrate (canine pancreas) experimental systems. A combination of protein biochemistry and yeast molecular genetics experiments has led to the identification of eight polypeptides (Figure 2A) that are assembled into a heteroooligomeric yeast OST complex composed of one copy each of Ost1p, Ost2p, Ost3p or Ost6p, Ost4p, Ost5p, Wbp1p, Swp1p, and Stt3p (Karaoglu et al., 1997; Spirig et al., 1997). Although several papers describe yeast OST complexes that apparently lack 2–4 of the above subunits (Kelleher and Gilmore, 1994; Knauer and Lehle, 1994; Pathak and Imperiali, 1997), this discrepancy is explained by the difficulty in detecting the diffusely migrating Stt3 protein as well as the three low molecular weight subunits (Ost2p, Ost4p, and Ost5p) on sodium dodecyl sulfate (SDS)-polyacrylamide gels (Karaoglu et al., 1997). Protein sequence database searches and protein purification results indicate that the yeast and vertebrate OST complexes have very similar subunit compositions (Kelleher and Gilmore, 1997; Kelleher et al., 2003). Canine OST complexes composed of ribophorin I (Ost1p), ribophorin II (Swp1p), OST48 (Wbp1p), DAD1 (Ost2p), and an STT3 homologue have been purified from canine pancreas (Kelleher et al., 2003). Below, the subunits of the OST are discussed in their approximate order of discovery.

Wbp1 (OST48)

The Wbp1 protein (OST48 in vertebrates) was the first identified subunit of the yeast OST (te Heesen et al., 1992) and one of the first three identified subunits (ribophorin I, ribophorin II, and OST48) of the purified canine OST (Kelleher et al., 1992; Silberstein et al., 1992). Wbp1p (OST48) proteins have an N-terminal signal sequence that is cleaved in vivo (te Heesen et al., 1992; Silberstein et al., 1992). N-linked glycans (yeast Wbp1 but not canine OST48), a hydrophobic segment near the C-terminus that serves as a TM span, and a short cytoplasmic tail that has a C-terminal dileucine ER retention/retrieval motif (Gaynor et al., 1994; Fu and Kreibich, 2000). These sequence features demonstrated that Wbp1 is a type I (N-lum-Cytoplasmic integral membrane protein with a large luminal domain (Figure 2A)). Wbp1 proteins are encoded by genomes from diverse eukaryotes including vertebrates, fungi, nematodes, arthropods, plants, Dictostelium discoideum, and many, but not all protists (Figure 2A–D), and show little variation in polypeptide length between organisms. Protein sequence comparisons suggest that Wbp1 proteins are members of the GldG/intraflagellar transport (GIFT) domain protein family that includes diverse prokaryotic and eukaryotic proteins (Beatson and Ponting, 2004). As the X-ray structure of a GIFT domain has not been solved, the biological function and structure of this domain in the Wbp1 proteins is not known.

Point mutations in Wbp1 reduce in vivo and in vitro OST activity (te Heesen et al., 1992) and destabilize the OST complex (Karaoglu et al., 1997). Deletion of the Wbp1 TM span or replacement of this segment with other hydrophobic sequences is not tolerated because of lack of incorporation of these mutant proteins into the OST complex (Li et al., 2003). Two papers that predated the identification of STT3 as the active-site subunit suggested that Wbp1/OST48 provides the oligosaccharide donor and acceptor binding sites that are critical for enzyme activity. Covalent inactivation of the yeast OST with a sulfhydryl reagent can be blocked by the presence of Dol-PP-GlcNAc3, indicating that a cysteine residue, which was mapped to Wbp1, is located in the dolichol-oligosaccharide donor
binding site of the OST (Pathak et al., 1995). The significance of this result cannot be discounted given that enzyme kinetic experiments now suggest that the OST has two independent OS-PP-Dol binding sites (Karaoglou et al., 2001). OST48 and a 68-kDa protein (thought to be ribophorin I at the time, but probably STT3A instead) were labeled with a suicide substrate analogue, wherein the hydroxy amino acid (T or S) was replaced by epoxyethylglycine (Bause et al., 1997).

**Ost1 (ribophorin I)**

The Ost1/ribophorin I subunit is present in the purified yeast and vertebrate OST complexes. The yeast OST1 gene was isolated based upon peptide sequence data (Silberstein S., Collins P.G., Kelleher D.J., Rapijeko P.J., et al., 1995) and identified in a genetic screen for mutations that reduce OST activity (Reiss et al., 1997). The Ost1 proteins have an N-terminal signal sequence that is cleaved in vivo, N-X-T/S sites that carry N-linked glycans in yeast and vertebrates, a single TM span and a C-terminal cytosolic domain, hence Ost1 proteins have a type I (N\textsubscript{lum}-C\textsubscript{cyt}) topology (Crimaudo et al., 1987; Silberstein S., Collins P.G., Kelleher D.J., Rapijeko P.J., et al., 1995). Protein sequence database searches reveal that Ost1/ribophorin I orthologues are encoded by all fully sequenced eukaryotic genomes except *Encephalitozoon cuniculi*, *G. lamblia*, and the kinetoplastids (Figure 2A–D). The cytoplasmic domains of Ost1/ribophorin I proteins are either <10 residues in length (fungi, insects, and many protists) or 140–160 residues in length (vertebrates, plants, nematodes, and *Plasmodium*). Consensus protein secondary structure prediction algorithms (http://npsa-pbil.ibcp.fr/) indicate that the longer cytoplasmic tails have a high probability of being α-helical.

Point mutations in Ost1 reduce N-linked glycosylation in vivo and reduce in vitro OST activity (Silberstein S., Collins P.G., Kelleher D.J., Rapijeko P.J., et al., 1995). Labeling of Ost1 with photoreactive acceptor substrates was detecting using an immunoprecipitation strategy (Yan et al., 1999). However, subsequent studies did not support the initial conclusion that Ost1 was the active site subunit.

Ribophorins I and II are abundant proteins of the mammalian RER (Kreibich et al., 1978) that are present in roughly equal amounts relative to membrane-bound ribosomes (Marcantonio et al., 1984). Antibodies raised against the cytoplasmic domain of ribophorin I interfere with protein translocation across the RER by preventing ribosome targeting to the Sec61 complex (Yu et al., 1990), thereby indicating that the OST is adjacent to the protein translocation channel. The location of OST adjacent to the translocon allows the cotranslational modification of the nascent polypeptide as it enters the lumen of the RER (Chen et al., 1995).

**Swp1 (ribophorin II)**

The yeast *SWP1* gene was isolated as a high-copy suppressor of a *wbp1* mutant (te Heesen et al., 1993), and Swp1 (ribophorin II) is present in the purified yeast and canine OST complexes. The Swp1 proteins have an N-terminal signal sequence that is cleaved in vivo, an N-glycosylation sequon that is modified in ribophorin II, and three hydrophobic segments located near the C-terminus (Rosenfeld et al., 1984; Crimaudo et al., 1987; te Heesen et al., 1993; Kelleher and Gilmore, 1994). Protease digestion experiments have shown that C-terminus of ribophorin II is insensitive to protease digestion in intact microsomes (Crimaudo et al., 1987) consistent with a three TM (N\textsubscript{lum}-C\textsubscript{cyt}) topology depicted in Figure 2A. Swp1/ribophorin II proteins are only encoded by a subset of eukaryotic genomes (Figure 2A and B) suggesting that this subunit was a relatively late addition to the OST complex. Nematode and fungi genomes encode truncated forms of Swp1/ribophorin II that lack an N-terminal domain that is predicted to be α-helical. RNAi of *Caenorhabditis elegans* ribophorin II causes an embryonic lethal phenotype (Fraser et al., 2000) consistent with the previous demonstration that Swp1 is essential in yeast (te Heesen et al., 1993).

**Ost2 (DAD1)**

Ost2 was initially detected as a 16-kD subunit in the purified yeast OST complex (Kelleher and Gilmore, 1994) and was subsequently shown to be present in the mammalian OST complex (Kelleher and Gilmore, 1997) using antibodies specific for the vertebrate homologue DAD1 (Nakashima et al., 1993). Ost2 (DAD1) subunits are low molecular weight (112–130 residues) proteins that consist of a polar N-terminal segment of variable length, followed by three conserved hydrophobic segments. Protease accessibility experiments have shown that the N-terminus and the hydrophilic segment between TM2 and TM3 are exposed on the cytoplasmic face of the membrane (Silberstein S., Collins P.G., Kelleher D.J., and Gilmore R., 1995; Kelleher and Gilmore, 1997), consistent with the 3-TM (N\textsubscript{cyt}-C\textsubscript{lum}) topology model shown in Figure 2A. An alternative topology model for Ost2 that depicts a 3-TM span protein with an inverted topology (Yan et al., 2003) is incompatible with the protease accessibility data cited above and is not in agreement with an Ost2p-reporter gene fusion experiment that suggests that the C-terminus is located within the RER lumen (Yan et al., 2005). An alternative 2-TM (N\textsubscript{cyt}-C\textsubscript{lum}) span model for DAD1 (Makishima et al., 1997) is based upon the assumption that the third hydrophobic segment is too polar to be integrated into the membrane. We would argue that TM3 of Ost2/DAD1 contacts the TM spans of other OST subunits, rather than phospholipid when Ost2/DAD is assembled into the OST complex. Ost2/DAD1 proteins are encoded by all fully sequenced eukaryotic genomes except *E. cuniculi*, *G. lamblia*, and the kinetoplastids (Figure 2A–D).

Point mutations in the TM spans of Ost2 cause in vivo and in vitro reductions in OST activity at the restrictive temperature (Silberstein S., Collins P.G., Kelleher D.J., and Gilmore R., 1995). A temperature sensitive mutation in the hamster DAD1 (defender against death) protein causes an apoptotic phenotype in the tsBN7 cell line because of instability of DAD1 (Nakashima et al., 1993). Degradation of DAD1 at the restrictive temperature results in a severe defect in OST activity (Makishima et al., 1997; Sanjay et al., 1998). The apoptotic phenotype of tsBN7 cells is most likely because of the irreversible induction of the unfolded protein response (UPR) pathway caused by the accumulation of unfolded glycoproteins, as tunicamycin can also induce apoptosis in plant and vertebrate cell lines (Pérez-Salva and Mollinedo, 1995; Walker et al., 1998; Iwata and Koizumi, 2005).
Mutations in the yeast STT3 locus were isolated in a screen for yeast that showed enhanced sensitivity to staurosporine (Yoshida et al., 1992), an inhibitor of protein kinase C. Additional stt3 alleles were obtained in genetic screens for mutations that confer a synthetic lethality phenotype when combined with the wbp1-2 allele (Zufferey et al., 1995) or the alg5Δ3 allele (Reiss et al., 1997). Because the yeast STT3 gene encodes a 78-kDa protein that was not detected in the purified yeast OST complex (Kelleher and Gilmore, 1994; Knauer and Lehle, 1994; Pathak and Imperiali, 1997), it was initially unclear whether Stt3 was an OST subunit (Yoshida et al., 1995; Zufferey et al., 1995). Subsequent immunopurification experiments using yeast strains that express affinity-tagged OST subunits (STT3-HA, STT3-protein A, or OST3-HA) demonstrated that Stt3 is in an equimolar complex with the other OST subunits (Karaoglu et al., 1997; Spirig et al., 1997).

The topology of the Stt3 proteins is more difficult to predict and experimentally verify than any of the other OST subunits. Eukaryotic Stt3 proteins consist of an N-terminal hydrophobic domain composed of 12–13 predicted TM spans followed by a C-terminal hydrophilic domain that is located within the RER lumen (Zufferey et al., 1995). The N-termini of canine STT3A and yeast Stt3 is exposed to the cytosol, hence eukaryotic STT3 proteins must have an uneven number of TM spans (e.g., 11 or 13) and a Ncyt-Clum topology (Kelleher et al., 2003; Kim et al., 2005). STT3 topology models with 13 TM spans (as depicted in Figure 2) are based on the assumption that each of the 13 hydrophobic segments are bona fide TM spans, whereas 11 TM span models necessitate that two of the hydrophobic segments are instead exposed on the cytosolic or lumenal faces of the membrane. A comprehensive investigation of the topology of the yeast and mouse STT3 proteins provided support for two conflicting 11 TM span models (Kim et al., 2005), suggesting that structural biology data will be needed to define the topology of the Stt3 proteins.

As initially suggested by the alignment of STT3 protein sequences from *S. cerevisiae*, *C. elegans*, and *M. musculus* (Zufferey et al., 1995), the Stt3 subunit is the most highly conserved polypeptide in the OST complex. With the exception of *E. cuniculi*, which lacks N-linked glycosylation, all fully sequenced eukaryotic genomes encode at least one member of the Stt3 protein family (Figures 2 and 3). Vertebrate, plant, and insect genomes encode two STT3 proteins that we have designated as STT3A and STT3B (Kelleher et al., 2003). Vertebrate STT3A was initially named TM conserved (TMC) (Lissy et al., 1996), whereas STT3B is also known as source of immunodominant major histocompatibility complex-associated peptides (SIMP) (McBride et al., 2002). Phylogenetic tree analysis shows that STT3B proteins are more closely related to the fungal and nematode STT3 proteins, whereas the STT3A protein cluster includes STT3 from *D. discoideum* (Figure 3). Several proist genomes encode multiple STT3 proteins that reflect gene duplication events that are unrelated to the derivation of the STT3A and STT3B subfamilies. For example, four consecutive genes on one chromosome encode the predicted *Leishmania major* STT3 proteins. To our knowledge, it is not known whether multiple STT3 proteins are simultaneously expressed in proist organisms.

**Ost3/Ost6 (N33 and IAP)**

The Ost3 protein was initially detected as a 34-kD subunit in the yeast OST complex (Kelleher and Gilmore, 1994). Ost6p is a homologue of Ost3p that is assembled into a second, less abundant, form of the yeast OST complex that lacks Ost3p as a subunit (Karaoglu et al., 1997; Knauer and Lehle, 1999; Spirig et al., 2005). The Ost3 family of proteins have an N-terminal signal sequence that is cleaved in vivo in yeast, a 160–180 residue N-terminal luminal domain that includes one thioredoxin repeat (Fetrow et al., 2001) followed by four predicted TM spans (Karaoglu et al., 1995). Consequently, the Ost3 proteins are integrated into the RER.
with a 4-TM (N_\text{lum}-C_\text{lum}) topology (Figure 2A). Recently, the 4-TM (N_\text{lum}-C_\text{lum}) topology for the Ost3 family has been disputed based upon results obtained with a reporter–gene fusion strategy (Yan et al., 2003, 2005). An alternative topology model for Ost3 predicts that a very short, weakly hydrophobic segment within the thioredoxin motif of Ost3p functions as a fifth TM span, thereby inverting the topology of Ost3 to yield a 5-TM (N_\text{lum}-C_\text{cyt}) topology (Yan et al., 2003, 2005). Both termini of Ost6 are proposed to face the cytosol in a 4-TM (N_\text{cyt}-C_\text{cyt}) topology (Yan et al., 2005), which would expose the N-terminal signal sequence and the thioredoxin domain of Ost6 on to the cytoplasmic side of the membrane (Chavan et al., 2005). How can we explain the discrepancy between these recent proposals and the previously accepted topology model shown in Figure 2A?

Although reporter–domain fusions have been used extensively to define the topology of bacterial and vertebrate membrane proteins, this strategy can be less satisfactory in yeast (as discussed by Kim H., Melen K., et al., 2003), because integration of membrane proteins can be perturbed by an N-terminal fusion joint that inactivates a signal sequence or by a C-terminal fusion joint that interferes with integration of the final TM span (Green and Walter, 1992).

Protein sequence database searches show that members of the Ost3 family of proteins are encoded by all fully sequenced vertebrate, nematode, fungi (except E. cuniculi), arthropod, and plant genomes. In addition, Ost3 homologues are encoded by D. discoideum, Tetrahymena thermophila, and Cryptosporidium parvum. The gene duplication event that is responsible for the OST6 gene is restricted to the Saccharomyces family of budding yeasts. Phylogenetic tree analysis using currently available fungal Ost3/Ost6 sequences indicates that other fungal genomes encode a protein that is more closely related to Ost3 than to Ost6p (not shown). Vertebrate genomes encode two closely related Ost3 homologues that are annotated as N33 and IAP (MacGrogan et al., 1996; Kelleher et al., 2003). Unlike Ost3 and Ost6 that are present in the purified yeast OST complex (Kelleher and Gilmore, 1994; Karaoglu et al., 1997) and comigrate with the OST on Blue-Native gels (Knauer and Lehle, 1999), N33 and IAP appear to be loosely associated with the canine OST complex and dissociate at various stages during enzyme purification (Kelleher et al., 2003) or during Blue-Native polyacrylamide gel electrophoresis (Shibatani et al., 2005).

Disruption of the yeast OST3 or OST6 genes does not cause a growth defect at any temperature in an otherwise wild-type strain (Karaoglu et al., 1995; Knauer and Lehle, 1999). Loss of Ost3p expression in yeast causes a modest, and substrate dependent, reduction in the in vivo glycosylation of proteins (Karaoglu et al., 1995). The mild phenotypes of the ost3A strain are explained in part by redundancy, as an ost3Δost6Δ strain has a more severe defect in glycosylation than either single mutant (Knauer and Lehle, 1999; Spirig et al., 2005). The C. elegans OST3 gene (ZK686.3) is essential; RNAi of ZK686.3 causes an embryonic lethal phenotype (Kamath et al., 2003) because of a defect in cytokinesis (Gonczy et al., 2000).

Ost4p

The yeast OST4 gene was identified in a screen for yeast mutants that are resistant to sodium vanadate and sensitive to hygromycin B (Chi et al., 1996), two properties that are shared by many yeast mutants that impact different steps in cell-wall biosynthesis (Dean, 1995). Point mutations, or the complete deletion, of the nonessential yeast OST4 gene causes severe in vivo hypoglycosylation of glycoproteins and a temperature sensitive growth defect (Chi et al., 1996). Remarkably, the OST4 gene encodes a 36-residue polypeptide (Chi et al., 1996), which is one of the smallest verified proteins encoded by the yeast genome. The single TM span in Ost4p is preceded by acidic residues and followed by basic residues thus favoring integration in the type I (N_\text{lum}-C_\text{cyt}) topology as initially proposed (Karaoglu et al., 1997; Spirig et al., 1997). Although a type I topology for Ost4p has been disputed (Kim et al., 2000), a more recent study confirmed the original prediction (Kim H., Yan Q., et al., 2003).

Homologues of yeast Ost4 are encoded by fungi, vertebrate, nematode, insect, plant, and several protist genomes (Figure 2A–C). Relatively few of these OST4 sequences are currently annotated in protein sequence databases. For example, the human and mouse expressed sequence tags (EST) databases contain highly expressed mRNAs that encode the 37-residue Ost4 homologue shown in Figure 4.

![Fig. 4](alignment_of_eukaryotic_ost4_sequences.png)

**Fig. 4.** Alignment of eukaryotic OST4 sequences. OST4 sequences were obtained by protein and DNA sequence database searches using *Saccharomyces cerevisiae* Ost4 as the query sequence. Representative fungal and metazoan OST4 sequences as well as all currently available protist sequences were selected for alignment using ClustalW. Regions of homology are boxed, with amino acid identities designated by dark shading. Replacement of hydrophobic residues (M18 to I24) in the TM span of *Saccharomyces cerevisiae* Ost4 with lysine or aspartic acid caused temperature sensitive growth defects and destabilized the interaction between Stt3 and Ost4 (asterisks) or destabilized the interaction between Ost3 and Ost4 (asterisks and diamonds) (Kim et al., 2000).
Ost4 proteins consist of conserved luminal and TM segments followed by a poorly conserved cytoplasmic domain of variable length (Figure 4). Mass spectroscopy of the purified canine OST complex has shown that a protein with a mass (4193 Da) consistent with the vertebrate Ost4 sequence is present in the complex (Kelleher and Gilmore, unpublished data). Hence we are confident that Ost4 is an OST subunit in diverse eukaryotes (Figure 2A–C).

Ost5

Ost5 was initially detected as a 9-kDa subunit in the yeast OST complex (Kelleher and Gilmore, 1994). The OST5 gene was identified in a screen for yeast genes that encode OST subunits (Reiss et al., 1997). Deletion of the yeast OST5 gene causes a minor defect in the in vivo glycosylation of yeast proteins and a reduction in OST activity that is not caused by an altered affinity for donor or acceptor substrates (Reiss et al., 1997). Ost5p is an 86-residue protein with two-predicted TM spans. The 2-TM N_cyt-C_cyt topology that was predicted for Ost5p (Reiss et al., 1997) has been confirmed (Yan et al., 2005). Although putative Ost5 homologues can be identified in metazoan organisms using protein sequence database searches (Kelleher and Gilmore, unpublished results), biochemical evidence that these polypeptides are OST subunits has not been obtained (Kelleher et al., 2003; Shibatani et al., 2005).

The STT3 subunit contains the catalytic site of the OST

The five OST subunits (Stt3, Ost1, Ost2, Swp1, and Wbp1) that are encoded by essential yeast genes were the logical candidates for the active-site subunit of the OST. Of these five proteins, only the Stt3 protein has archaeobacterial homologues (Spirig et al., 1997). Alignment of eukaryotic and archaeobacterial STT3 proteins resulted in the identification of a short invariant sequence (WWDYG motif) that is present in all members of the STT3 family (Spirig et al., 1997; Burda and Aebi, 1999). During the past several years, overwhelming evidence has been obtained indicating that Stt3 family proteins catalyze N-linked glycosylation in eukaryotes, archaeobacteria and in selected eubacteria.

The most direct evidence that Stt3 proteins are the catalytic subunits of the eukaryotic OST was provided by an elegant in vivo analysis of the C. jejuni pgp operon. N-linked glycosylation of AcrA, a C. jejuni membrane protein was dependent upon a wild-type PgIB locus (Wacker et al., 2002). Moreover, the expression of AcrA and the complete pgp operon in Escherichia coli reconstituted N-linked glycosylation activity in the E. coli host unless point mutations were introduced into the WWDYG motif in PgIB (Wacker et al., 2002). Genetic analysis of PgIB has been facilitated by the nonessential nature of N-linked glycosylation in C. jejuni and in the E. coli host system. Structural analysis of the resulting glycopeptide showed that the modified asparagine (N*) was present within the peptide DEFN*VSK (Wacker et al., 2002), where the underlined sequence matches the N-X-T/S sequon for eukaryotic N-linked glycopeptides. The N-X-T/S sequence is necessary, but not sufficient for N-linked glycosylation in C. jejuni, as several N-X-T/S sites in the AcrA and HisJ glycoproteins are not modified (Nita-Lazar et al., 2005). However, because 35% of N-X-T/S sequons are not modified by the eukaryotic OST, further work will be required to define a minimum consensus site for PgIB. The heptasaccharide (GalNAc-α1,4-GalNAc-α1,4-(Glcβ1,3)-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac) transferred to proteins by the C. jejuni PgIB protein bears little resemblance to the high-mannose oligosaccharide donor assembled by eukaryotic cells (Young et al., 2002), except that the saccharide linked to the undecaprenyl-pyrophosphate carrier is an amino sugar (i.e., Bac = bacillosamine = 2,4,diacetamido-2,4,6-trideoxy-glucose). The relaxed donor-substrate specificity for the eubacterial OST was further demonstrated by the discovery that structurally diverse oligosaccharides attached to undecaprenyl pyrophosphate are utilized when PgIB is expressed in E. coli (Feldman et al., 2005).

Evidence that Stt3 contains the catalytic center of the yeast OST has been provided by a combination of photolabeling experiments and site-directed mutagenesis (Yan and Lennarz, 2002b). Three of the OST subunits (Ost1, Ost3, and Stt3) can be photolabeled with acceptor peptides that incorporate a photoreactive benzophenone side chain near an N-X-T glycosylation site. Mutagenesis of the C-terminal luminal domain of yeast Stt3 yielded a collection of assembly defective and assembly competent stt3 mutants that displayed severe growth defects and dramatically reduced OST activity. Point mutations in the WWDYG motif of yeast Stt3p eliminate or dramatically reduce OST activity (Yan and Lennarz, 2002b).

Highly selective photocross-linking of a nascent polypeptide to the active site of the canine STT3A has been achieved by incorporating a photoreactive derivative of lysine (K*) adjacent to a cryptic glycosylation site (Q-K*-T instead of N-K*-T) in a nascent chain undergoing translocation into the RER lumen (Nilsson et al., 2003). Unlike the photolabeling experiments summarized above (Yan and Lennarz, 2002b), cross-links between other OST subunits and the cryptic glycosylation site were not detected. Photocross-linking of the cryptic glycosylation site to STT3A occurred when the photoprobe was located at least 69 residues away from the peptidyl-transferase site on the ribosome (Nilsson et al., 2003), which correlates well with the minimal length between the peptidyl-transferase site and an N-X-T/S that can be glycosylated in a ribosome-bound nascent polypeptide (Whitely et al., 1996). These results reinforce the conclusion that the N-X-T/S sites pass by the OST active site as the nascent polypeptide emerges from the lumenal face of the protein translocation channel.

Biochemical evidence that STT3 proteins are the active site subunits of the OST complex was provided by the kinetic analysis of canine OST complexes that incorporate different isoforms of the STT3 subunit (Kelleher et al., 2003). OST complexes that have an STT3A subunit display a higher specificity for selecting the fully assembled dolichol-linked oligosaccharide donor (Glc3Man9GlcNAc2-PP-Dol) than OST complexes that contain an STT3B subunit. These differences in donor-substrate selection are explained by STT3 isoform-dependent differences in the kinetic parameters of the OST active site (Kelleher et al., 2003).
Organization of subunits into the OST complex

Biochemical and genetic evidence for specific protein–protein interactions between the OST subunits has steadily accumulated during the past decade. In the absence of a high-resolution OST structure, we consider how these results can be incorporated into a working model for the eukaryotic OST complex.

Nearest-neighbor analysis suggest OST subcomplexes

The first evidence for interactions between OST subunits was provided by the isolation of the SWP1 gene as an allele-specific high-copy suppressor of the wbp1-2 mutant (te Heesen et al., 1993). The overexpression of Ost2, like Swp1, suppresses the temperature sensitive growth defect of a wbp1-2 strain by stabilizing Wbp1 (Silberstein S., Collins P.G., Kelleher D.J., and Gilmore R., 1995). Reciprocal gene product depletion experiments (Knauer and Lehle, 1994), protein cross-linking studies (te Heesen et al., 1993), and yeast two-hybrid experiments (Fu et al., 1997) indicate that Swp1 (ribophorin II) interacts directly with and, by so doing, stabilizes Wbp1 (OST48). Chemical cross-linking experiments have defined direct interactions between DAD1, OST48, and ribophorin II (Kelleher and Gilmore, 1997). Ost48 and ribophorin II are unstable when tsBN7 cells (DAD1-ts cell line) are shifted to the restrictive temperature (Sanjay et al., 1998). Reciprocal loss of stability would be a weak argument for protein interactions between Wbp1, Swp1, and Ost2, if a given mutation (e.g., wbp1-2) reduced the stability of all eight OST subunits. However this does not appear to be the case based upon western blot analysis of OST subunit content in wbp1-2 and ost2-3 mutants (Silberstein S., Collins P.G., Kelleher D.J., and Gilmore R., 1995; Karaoglu et al., 1997). These results led to the proposal that Wbp1, Swp1, and Ost2 are adjacent in the OST and constitute one of three subcomplexes (Karaoglu et al., 1997; Spirig et al., 1997).

Several lines of evidence indicate that Ost3 and Ost4 interact directly with Stt3 in a second subcomplex (Figure 2A). The OST3 and OST4 genes are high-copy allele specific suppressors of several stt3 mutants, and OST3 is a high-copy suppressor of the temperature sensitive growth defect of the ost4Δ mutant (Spirig et al., 1997). Biochemical support for these interactions was provided by the observation that an Stt3-Ost3-Ost4 heterotrimer is recovered when an immunopurified OST complex is incubated in a mixed-detergent micelle (Triton X-100-SDS) wash buffer (Karaoglu et al., 1997). Coimmunoprecipitation experiments have shown that Ost3p dissociates from the OST complex in detergent extracts prepared from an ost4Δ mutant, indicating that Ost4 stabilizes the association between Ost3 and Stt3 (Karaoglu et al., 1997). Point mutations that introduce charged amino acid residues into the C-terminal half of the Ost4 TM span (Kim et al., 2000; Kim H., Yan Q., et al., 2003) cause severe growth defects and destabilize interactions between Stt3, Ost3, and Ost4 (Figure 4). The dissociation of N33 and IAP from canine OST complexes (Kelleher et al., 2003) and the dissociation of Ost3 in an ost4Δ mutant (Karaoglu et al., 1997; Kim et al., 2000) is best explained by a peripheral location for the OST3 proteins in vertebrate and fungal complexes. Notably, disruption of the yeast OST3 gene does not destabilize the remaining subunits in the OST complex (Karaoglu et al., 1997).

Genetic evidence for a third subcomplex (Ost1-Ost5) was provided by the discovery that the OST5 gene is a high-copy suppressor of the ost1-5 allele (Reiss et al., 1997). More recently, chemical cross-linking experiments provide additional support for the Ost1p–Ost5p interaction (Yan et al., 2003). The protein–protein interactions summarized above must be a subset of those that exist within the OST, as the active enzyme does not readily dissociate into three stable subcomplexes when exposed to nonionic detergents. Although recent chemical cross-linking experiments suggest numerous additional subunit–subunit contacts (Yan et al., 2003), it should be noted that the analytical methods used were unable to distinguish between a direct cross-link between two subunits and an indirect link that is mediated by an additional subunit.

Assembly of subcomplexes into the OST; what do we learn from genomics?

We can now consider the OST subcomplex hypothesis in light of the predicted diversity of eukaryotic OST complexes. With the exception of the 1-subunit enzymes predicted for organisms that are at the base of the eukaryotic phylogenetic tree (Figure 2E), the simplest eukaryotic OST is predicted to have four subunits (Figure 2D), including one or more subunits from each of the putative subcomplexes. We can infer from this information that Stt3 provides an interaction surface for the Wbp1-Ost2-Swp1 complex that is not dependent upon Swp1 (Figure 2C and D). A direct interaction between Ost1 and Stt3, but not between Ost1 and the Wbp1-Ost2-Swp1 subcomplex, is suggested by the dissociation of Ost1 from the latter subcomplex in Stt3 depleted yeast cells (Zufferey et al., 1995). These considerations suggest that STT3 provides independent binding sites for Ost1 and the Wbp1-Ost2 heterodimer (Figure 2D).

Three of the OST subunits (Ost2, Ost4, and Ost5) are hydrophobic miniproteins. It is noteworthy that an Ost4 protein is encoded by all fully sequenced genomes that encode an Ost3 protein (Figures 2A–C and 4). This observation supports the hypothesis that Ost4p is an assembly/stabilizing factor for the incorporation of an Ost3 family protein into the complex (Karaoglu et al., 1997; Kim et al., 2000). The topology and extensive conservation of the TM spans of Ost2 suggest the following: (i) the essential function of Ost2 can be mapped to the TM spans in accord with mutagenesis results (Silberstein S., Collins P.G., Kelleher D.J., and Gilmore R., 1995) and (ii) Ost2 may stabilize/mediate an interaction between Stt3 and Wbp1. Likewise, one can speculate that Ost5 stabilizes the interaction between Ost1 and Stt3.

Taken together, the genomic evidence indicates that the three OST subcomplexes are explained, at least in part, by the stepwise addition of subunits to an STT3 catalytic core during evolution of the eukaryotic OST. The first proteins to be added may have been an Ost1 monomer and the Wbp1-Ost2 heterodimer. Subsequent enlargement of the OST involved the addition of four peripheral subunits that were likely added as a heterodimer (Ost3-Ost4) and two
monomers (Swpl and Ost5), with the Wbp1-Ost2 heterodimer contributing the binding site for Swpl. Although a recently published interaction map for the yeast enzyme places Ost1 in the center surrounded by other subunits (Yan et al., 2005), such a location is incompatible with protein cross-linking results obtained using the mammalian experimental system (Kelleher and Gilmore, 1997) and yeast two hybrid experiments that did not detect interactions between ribophorin I and II (Fu et al., 1997). Recently, it has been shown that ribophorin I can be cross-linked to a subset of nascent integral membrane proteins, in a sequon-independent manner, as the nascent chains exit the lateral gate of the Sec61 complex, suggesting an exposed location for Ost1/ribophorin I in the OST complex (Wilson et al., 2005). Clearly, a better understanding of the protein–protein interactions within the OST complex will only be obtained when mid- or high-resolution structures of the OST complex are solved.

**OST isoforms in mammalian cells**

STT3A and STT3B are assembled into alternative forms of the canine OST complex that differ with respect to donor-substrate selection and specific activity (Kelleher et al., 2003). Human multiple-tissue northern blots have shown that both the STT3A and the STT3B mRNAs are widely expressed in human tissues, and this conclusion is supported by the human UniGene database. Protein immunoblot experiments have shown that both STT3 isoforms are expressed in human fibroblasts and tissue culture cells (CH12.LX, HEK293, and HeLa cells; Kelleher et al., 2003; Ruiz-Canada and Gilmore, unpublished data). Nonetheless, tissue-specific differences in the relative expression of the STT3 isoforms are readily apparent and may indicate that the enzymatic properties of the OST are regulated to respond to glycoprotein flux through the secretory pathway.

**Donor-substrate selection by the eukaryotic OST**

The terminal glucose residue on the dolichol-linked oligosaccharide strongly influences the rate of oligosaccharide transfer in vertebrate and fungal organisms (Turco et al., 1977; Trimble et al., 1980). The accumulation of nonoptimal OS-PP-Dol assembly intermediates is responsible for the reduced rate of oligosaccharide transfer to nascent glycoproteins in tissues from CDG-I patients. The utilization of a luminally exposed assembly intermediate (Man₃GlcNAc₂-PP-Dol to Glc₂Man₃GlcNAc₂-PP-Dol) should have a deleterious effect upon higher eukaryotic organisms, because the structure of a protein-bound oligosaccharide has a profound effect upon the ability of a nascent glycoprotein to fold correctly and be packaged into vesicles for intracellular transport or instead to be targeted for ER-associated degradation (ERAD) if folding is not successful (for a review, see Parodi, 2000; Schrag et al., 2003; Helenius and Aebi, 2004).

The mechanism of oligosaccharide-donor selection has been investigated for more than two decades using enzyme kinetics methods (Trimble et al., 1980; Breuer and Bause, 1995; Gibbs and Coward, 1999). The OST catalyzes a bisubstrate reaction, where both the donor and acceptor substrates must be present at the same time. The dolichol pyrophosphate and glycopeptide products are not OST inhibitors (Gibbs and Coward, 1999; Peluso et al., 2002). Competition experiments using donor substrate mixtures have shown that the fully assembled donor (Glc₃Man₃GlcNAc₂-PP-Dol) is transferred 5–20 times faster than an assembly intermediate (e.g., Man₀GlcNAc₂-PP-Dol) when the OST is assayed using microsomal membranes (Karaoglu et al., 2001) or a detergent-solubilized enzyme (Trimble et al., 1980). Contrary to expectations, several early studies using detergent-solubilized OST showed that the apparent $K_m$ for the donor oligosaccharide was not significantly influenced by the terminal glucose residues (Trimble et al., 1980; Sharma et al., 1981). Although the structure of the oligosaccharide donor does have an effect upon the apparent $K_m$ for the acceptor substrate (Breuer and Bause, 1995; Gibbs and Coward, 1999), the relatively modest enhancement in peptide-binding affinity that is observed in the presence of Glc₃Man₃GlcNAc₂-PP-Dol relative to that observed with an assembly intermediate cannot by itself explain the efficiency of donor-substrate selection.

Enzyme kinetic experiments using purified yeast (Karaoglu et al., 2001) or canine (Kelleher et al., 2003) OST and purified OS-PP-Dols of defined composition have now revealed unexpected properties of the OST complex. In agreement with previous studies (Welply et al., 1983; Breuer and Bause, 1995; Hendrickson and Imperiali, 1995; Gibbs and Coward, 1999), acceptor-substrate saturation experiments yield linear Lineweaver-Burk plots consistent with a single binding site for the acceptor-peptide substrate (Karaoglu et al., 2001; Kelleher et al., 2003). In the case of the yeast OST, the apparent $K_m$ for the acceptor substrate is enhanced by −1.5-fold when Glc₃Man₃GlcNAc₂-PP-Dol is the donor substrate relative to an assembly intermediate (Karaoglu et al., 2001). Donor-substrate saturation curves for Glc₃Man₃GlcNAc₂-PP-Dol were observed to be sigmoidal, not hyperbolic, and indicate that the yeast OST binds the donor substrate in a cooperative manner (Karaoglu et al., 2001). Importantly, donor-substrate saturation curves obtained with an assembly intermediate (Man₀GlcNAc₂-PP-Dol) are also sigmoidal, indicating that a two-part binding site that recognizes both the terminal glucose residue and the dolichol-PP-GlcNAc₂ moiety cannot account for the cooperative-binding kinetics of the donor substrate. Two donor-substrate binding sites and one acceptor substrate binding site are compatible with a substrate activation model for the OST, wherein binding of the donor substrate to a regulatory OS-PP-Dol binding site is a prerequisite for binding of both the acceptor substrate and a second donor substrate to the catalytic site. The kinetic analysis of canine OST isoforms confirmed that higher eukaryotic OST complexes have two independent binding sites for the donor oligosaccharide and further showed that the identity of the catalytic subunit (STT3A versus STT3B) affected the active site kinetic parameters, but not the regulatory site kinetic parameter (Kelleher et al., 2003).

How can a donor-substrate activation model for the OST explain the selective utilization of the fully assembled OS-PP-Dol? According to the substrate-activation model we have proposed (Figure 5), the ability of the OST to discriminate between the glucosylated and nonglucosylated donor substrates is determined by the ratio of two pairs of interaction
factors (γ/α and δ/β). The interaction factors describe how the structure of the OS-PP-Dol molecule bound to the regulatory site influences the binding affinity of the catalytic site for Glc3Man9GlcNAc2-PP-Dol or an assembly intermediate. Effectively, the substrate-activation mechanism allows the OS-PP-Dol-binding affinity of the catalytic site to be fine-tuned to favor the utilization of Glc3Man9GlcNAc2-PP-Dol when a mixed population of oligosaccharide donors is present. Because the affinity of the regulatory site for Glc3Man9GlcNAc2-PP-Dol and Man9GlcNAc2-PP-Dol is similar (94 and 67 nM, respectively), occupation of the regulatory site will reflect the composition of the donor oligosaccharide pool, which under normal physiological conditions will be enriched in Glc3Man9GlcNAc2-PP-Dol. When the regulatory site is occupied by Glc3Man9GlcNAc2-PP-Dol, the relative affinity of the catalytic site for the donor substrate will favor binding of the fully assembled donor by -1.5-fold (γ/α) relative to an assembly intermediate. Consequently, the OST will primarily utilize Glc3Man9GlcNAc2-PP-Dol in vivo both because this compound is the most abundant donor and because of the homotropic interactions favoring Glc3Man9GlcNAc2-PP-Dol binding to the catalytic site. When the regulatory site is occupied by an assembly intermediate (e.g., Man9GlcNAc2-PP-Dol), the enzyme will display a 7- to 10-fold preference (δ/β) for Glc3Man9GlcNAc2-PP-Dol relative to an assembly intermediate, hence any available Glc3Man9GlcNAc2-PP-Dol will be preferentially utilized. What in vivo evidence supports this kinetic model for donor-substrate selection? The original yeast alg3 mutant, which is slightly leaky, preferentially utilizes the residual Glc3Man9GlcNAc2-PP-Dol to glycosylate proteins in vivo (Verostek et al., 1993). The latter situation also occurs in wild-type yeast cells as the dextrose in growth media is depleted during late log phase (Trimble et al., 1980) or in human CDG fibroblast cell lines that have leaky mutations in human ALG genes (Korner et al., 1999). The complete absence of Glc3Man9GlcNAc2-PP-Dol from the donor oligosaccharide pool, as occurs in alg gene null mutants (e.g., alg6Δ; Reiss et al., 1996), causes a reduction in the in vivo glycosylation of glycoproteins in accord with the reduced transfer rate predicted by the kinetic model. For a more detailed presentation of this model for donor-substrate selection by the eukaryotic OST, we refer the interested reader to Karaoglu et al. (2001).

Trypanosomes (de la Canal and Parodi, 1987) and many other protists (Samuelson et al., 2005) assemble OS-PP-Dols that lack glucose residues, hence these organisms provide useful experimental systems to analyze donor-substrate selection by the OST. Unlike the yeast and mammalian OST which select Glc3Man9GlcNAc2-PP-Dol from a complex mixture of dolichol-linked oligosaccharides, the trypanosome OST does not show a preference for either the optimal vertebrate donor (Glc3Man9GlcNAc2-PP-Dol) or the native donor (e.g., Man9GlcNAc2-PP-Dol for Trypanosoma cruzi) relative to other compounds in a donor oligosaccharide mix (Bosch et al., 1988). These results suggest that donor-substrate selection is a property of the multi-subunit OST complexes that are expressed in organisms that assemble the triglucosylated oligosaccharide donors. It will now be important to examine the kinetic properties of the OST from organisms that are...
predicted to have simpler OST complexes (Figure 2C and D) to determine whether cooperative binding of OS-PP-Dol can be ascribed to the presence of a specific subunit.

Conclusions

The conclusion that STT3 proteins are the active site subunits of the OST in eukaryotes, archaeabacteria and certain eubacteria ended a three-decade long search for the protein that catalyzes this critical step in glycoprotein biosynthesis. The discovery that the eubacterial enzyme is a single subunit (PglB), rather than the 300-kDa hetero-oligomeric present in most eukaryotes, greatly improves the prospect that a high-resolution structure of the OST can be obtained in the foreseeable future. Although the crystallization of a membrane protein is not a trivial endeavor, the successful expression of *C. jejuni* PglB in *E. coli* has paved the way for progress in this area. The expression of PglB in *E. coli* will also allow investigators to examine the OST mechanism in a simpler system. At the same time, it remains clear that there are important differences between the eubacterial and eukaryotic OST in terms of donor- and acceptor-substrate specificity and donor-substrate selection. We can anticipate that analysis of the OST from diverse organisms will reveal new and unexpected findings about this important enzyme.

Abbreviations

ALG, asparagine-linked glycosylation; Amb, diaminobutyric acid; CDG, congenital disorders of glycosylation; CSG, cell surface glycoprotein; OS-PP-Dol, dolichol pyrophosphate-linked oligosaccharide; OST, oligosaccharyltransferase; RER, rough endoplasmic reticulum; TM, transmembrane.

References


