Glycobiology vol. 12 no. 2 pp. 17R–27R, 2002

MINI REVIEW

Complex glycosylation of Skp1 in *Dictyostelium:* implications for the modification of other eukaryotic cytoplasmic and nuclear proteins

Christopher M. West[1,2], Hanke van der Wel[2], and Eric A. Gaucher[3]

[2]Department of Anatomy and Cell Biology, 1600 SW Archer Road, University of Florida College of Medicine, Gainesville, FL 32610-0235, USA, and [3]Department of Chemistry, University of Florida, Gainesville, FL 32611-7200, USA

Accepted on November 2, 2001

Recently, complex *O*-glycosylation of the cytoplasmic/nuclear protein Skp1 has been characterized in the eukaryotic microorganism *Dictyostelium*. Skp1's glycosylation is mediated by the sequential action of a prolyl hydroxylase and five conventional sugar nucleotide–dependent glycosyltransferase activities that reside in the cytoplasm rather than the secretory compartment. The Skp1-HyPro GlcNAc-Transferase, which adds the first sugar, appears to be related to a lineage of enzymes that originated in the prokaryotic cytoplasm and initiates mucin-type *O*-linked glycosylation in the lumen of the eukaryotic Golgi apparatus. GlcNAc is extended by a bifunctional glycosyltransferase that mediates the ordered addition of β1,3-linked Gal and α1,2-linked Fuc. The architecture of this enzyme resembles that of certain two-domain prokaryotic glycosyltransferases. The catalytic domains are related to those of a large family of prokaryotic and eukaryotic, cytoplasmic, membrane-bound, inverting glycosyltransferases that modify glycolipids and polysaccharides prior to their translocation across membranes toward the secretory pathway or the cell exterior. The existence of these enzymes in the eukaryotic cytoplasm away from membranes and their ability to modify protein acceptors expose a new set of cytoplasmic and nuclear proteins to potential prolyl hydroxylation and complex *O*-linked glycosylation.

Key words: glycosyltransferase evolution/cytoplasmic glycosylation/*Dictyostelium*/prolyl hydroxylase/Skp1

Introduction

The secretory pathway has long been considered the provenance of protein glycosylation in eukaryotes, so it has always been interesting when evidence is presented for glycosylation of cytoplasmic or nuclear proteins. This review will focus on complex glycosylation, defined by the addition of a chain of at least two sugars to a protein amino acid, because "simple" modification of Ser/Thr residues by *O*-β-GlcNAc has been recently reviewed (Thornton *et al.*, 1999; Comer and Hart, 2000; Wells *et al.*, 2001). As concluded by a comprehensive summary in 1989 (Hart *et al.*, 1989) and updated in 1999 (in Varki *et al.*, 1999), considerable data support the existence of cytoplasmic/nuclear glycoproteins with complex glycans. At the cellular level, immunocytochemical studies based on lectins, anticarbohydrate antibodies, and glycosaminoglycan-binding proteins localize glycoconjugates to the cytoplasm and nucleus. At the molecular level, cytoplasmic/nucleoplasmic proteins can be labeled with lectins in periodate-sensitive fashion, metabolically labeled with radioactive sugar precursors, or shown to contain chemical quantities of sugars after purification. However, as also summarized in those reviews, direct structural studies are required to support these interpretations. For example, recent work failed to confirm glycosylation of large T antigen and a high mobility group protein except for modification by *O*-β-GlcNAc (Medina and Haltiwanger, 1998a,b). In parallel, the discovery of extracellular functions for some cytoplasmic/nuclear lectins (Hughes, 2001; Lowe, 2001) has diminished the likelihood that intracellular glycan ligands exist. Nevertheless, complex cytoplasmic glycosyation does occur, as reviewed here, and so questions that relate to its prevalence and significance to the cell are relevant.

Evidence for cytoplasmic/nuclear proteins with complex glycans

A compilation of eukaryotic cytoplasmic or nuclear proteins that are known or suspected to be glycosylated is given in Table I. A compelling case has been presented for complex glycosylation of *Paramecium bursaria* Chlorella virus-1 (PBCV-1) capsid proteins in the cytoplasm of *Chlorella* algae. This virus is assembled in the cytoplasm away from membranes, consistent with the lytic mode of viral dispersion (Van Etten *et al.*, 1991). Its VP54 capsid protein is metabolically labeled with [^{3}H]Gal and contains Fuc, Gal, Man, Xyl, Ara or Rha, and Glc (Wang *et al.*, 1993), and its M_r is shifted from 54,000 to 49,000 by chemical deglycosylation. Six spontaneous virus mutants, isolated in a screen for resistance to an anti-carbohydrate virus antibody, express M_r variants of VP54 that correlate with altered sugar composition, providing strong evidence for a covalent association of sugars with the protein (Que *et al.*, 1994). These results led to a proposed glycan structure and a corresponding biosynthetic pathway of six virally encoded glycosyltransferases (GTases) that reside in

[1]To whom correspondence should be addressed

Table I. Candidate cytoplasmic/nuclear glycoproteins

Protein	Organism	Information on sugar chain	Attachment site info	References[a]
Established				
Skp1	*Dictyostelium*	Galα1,6Galα1, Fucα1,2Galβ1,3GlcNAcα1→	HyPro143	Teng-umnuay *et al.*, 1998
Glycogenin	eukaryotes	$(Glc\alpha1,4)_{n\ up\ to\ 1000}$→	Tyr196	Alonso *et al.*, 1995
Many	eukaryotes	GlcNAcβ1→	Ser/Thr	Comer and Hart, 2000
Rho	mammal by Clostridial toxins	Glc→, GlcNAcβ1→	Ser/Thr	Busch *et al.*, 1998
Candidate[b]				
VP54	algal virus	(Fuc,Gal,Rha/Ara,Xyl,Man,Glc)→	?	Wang *et al.*, 1993; Graves *et al.*, 2001
Parafusin	eukaryotes	$Glc\alpha1\text{-}PO_4\text{-}Man\text{-}Gly_{n\leq3}$[c]→	Ser/Thr	Marchase *et al.*, 1993
Nuclear pore protein	Tobacco	$GlcNAc\text{-}Gly_{n\leq4}$→	Ser/Thr	Heese-Peck *et al.*, 1996
Cytokeratin	mammals	$GalNAc\alpha1,3Gly_{n=?}$→	?	Goletz *et al.*, 1997
α-Synuclein	mammalian brain	sialyl-Galβ1,3GalNAcα1→	Ser/Thr	Shimura *et al.*, 2001
Proteoglycans	mammalian brain	chondroitin SO_4→, heparan SO_4→	Ser/Thr (or free)	Margolis and Margolis, 1993

Established examples of both complex and "simple" monosaccharide protein-linked glycans are listed in the upper section. The lower section lists additional, speculative examples as described in the text.
[a]See text for additional references.
[b]The entries in this section remain to be confirmed as described in the text.
[c]Gly denotes unknown sugar.

the cytoplasm. BLAST analysis of the PBVC-1 genome yields six candidate GTase genes. Most of these lack known targeting motifs for the secretory pathway, and one exhibits remote similarity to "Fringe"-type GTases and contributes to VP54 glycosylation based on genetic analyses (Graves *et al.*, 2001).

α-Synuclein is a neural protein that is subject to aggregation in Parkinson's disease. A minor M_r 22,000 isoform of α-synuclein was found in normal brain extracts in a complex with the E3 ubiquitin ligase Parkin (Shimura *et al.*, 2001), which appears to process α-synuclein for polyubiquitination and degradation. Glycosidase digestions converted Parkin-associated α-synuclein to the normal M_r 16,000 valve of the bulk pool, suggesting that this minor form is modified by a sialylated Galβ1,3GlcNAc α1 substituent(s). If glycosylation can be confirmed, it will be interesting to examine whether glycosylation is a trigger for α-synuclein ubiquitination, analogous to the role of phosphorylation for targets of the SCF E3 ubiquitin ligase (Deshaies, 1999).

A potential example of dynamic cytoplasmic glycosylation is that of the phosphoglucomutase-family protein parafusin (Levin *et al.*, 1999). Phosphorylation and membrane association of parafusin oscillates with secretory activity of the cell. Phosphate is proposed to be incorporated in crude extracts into a $Glc\alpha1\text{-}PO_4$-Man moiety on an unknown *O*-linked glycan, based on indirect product characterization from radioactive precursors (Satir *et al.*, 1990; Marchase *et al.*, 1990, 1993; Veyna *et al.*, 1994). Partial corroborating evidence was obtained for parafusin metabolically labeled with [2-^{3}H]Man, but interpretation of the phosphosugar linkages is confounded by recent reports of extensive Ser-phosphorylation (Kussman *et al.*, 1999) that was not originally reported. If this model can be supported by direct structural studies, parafusin has the potential to greatly expand our perspective on dynamic cytoplasmic glycosylation.

Plant nuclear pores contain proteins that are recognized by wheat germ agglutinin and can be radiolabeled by UDP-[^{3}H]Gal and soluble β1,4-Gal-Tase, indicating terminal GlcNAc residues (Heese-Peck *et al.*, 1996). Unlike animal nuclear core complex proteins modified by clustered, single *O*-β-GlcNAc residues, alkaline degradation released products approximately five sugars in length, with no evidence for a peptide based on enzymatic and chemical tests. These proteins might be modified by a glycan(s) with a nonreducing terminal GlcNAc, *O*-linked to Ser or Thr, but proof awaits direct demonstration of the structure.

Another report suggested that cytokeratins are modified by a GalNAcα1,3-terminated oligosaccharide(s), based on lectin binding and compositional analysis of the purified protein (Goletz *et al.*, 1997). Binding was diminished by pretreatment with α-N-acetyl-galactosaminidase and did not occur with recombinant cytokeratin. The implication that cytokeratins might be ligands for endogenous galectin(s) is intriguing but awaits direct structural evidence to rule out the possibility of a contaminating glycoprotein.

Although it is a special case, the primer for glycogen synthesis, glycogenin, was the first-known cytoplasmic glycoprotein (Alonso *et al.*, 1995). A Tyr-residue on glycogenin serves as the attachment site for a sequentially added α1,4-linked Glc oligomer that is applied by yet another glycogenin protein doubling as a GlcTase (Lin *et al.*, 1999).

Examples of *N*-linked glycans in the cytoplasm probably originate in the secretory pathway. There is good evidence for the *N*-glycosylation of a mitochondrial protein, but it appears to be imported from the rough endoplasmic reticulum (RER) (Chandra *et al.*, 1998). An N-glycosylated nuclear lectin has been also found in the RER and Golgi (Rousseau *et al.*, 2000). It is now also appreciated that unfolded RER glycoproteins are targeted to the cytoplasm for degradation (Cacan and Verbert, 2000), but their sugar chains are presumably transient and

nonfunctional in this compartment. If evidence can be obtained that *N*-linked glycans associated with tau protein from paired helical filaments of Alzheimer's brains (Sato *et al.*, 2001) is not due to contaminant proteins, it would suggest a breakdown in compartmentalization in a pathological situation. Overall, the observations raise the possiblity of N-glycan function in the cytoplasm or nucleus.

It is worth noting that the bulk of eukaryotic cytoplasmic glycosylation probably occurs on exported polysaccharides, including hyaluronan (DeAngelis, 1999), chitin, cellulose (Delmer, 1999), other glucans, and lipids that are subsequently flipped across membranes as precursors for protein *N*-glycosylation (Schenk *et al.*, 2001; Burda and Aebi, 1999), glycophosphatidylinositol anchors (Tiede *et al.*, 1999), and glycosphingolipids initiated with Glc (Marks *et al.*, 1999). Some glycolipids, such as sterol glucosides or mannosides, appear to remain oriented toward the cytoplasm (Warnecke *et al.*, 1999). A polysaccharide that is normally exported, hyaluronan, also appears to accumulate in the cytoplasm and nucleus based on immunocytochemical studies and evidence for natural cytoplasmic and nuclear receptors (Lee and Spicer, 2000; Huang *et al.*, 2000). Intracellular hyaluronan might result from failure to translocate to the cell exterior. Older biochemical and microscopic evidence also exists for the accumulation of proteoglycans or derived glycosaminoglycans (including heparan, chondroitin, and dermatan sulfates) in the cytoplasm and nucleus of brain and cultured mammalian cells, which was suggested to involve endocytosis and processing of conventionally secreted proteoglycans (Fedarko *et al.*, 1989; Busch *et al.*, 1992; Margolis and Margolis, 1993; Hiscock *et al.*, 1994).

Complex cytoplasmic glycosylation of Skp1 in *Dictyostelium*

The best studied example of a cytoplasmic/nuclear glycoprotein and a corresponding cytoplasmic glycosylation pathway has emerged from the analysis of Skp1 in the "lower" eukaryote *Dictyostelium*, known best for cell biological studies of development. *Dictyostelium* has also been an attractive model organism for glycobiology, because it possesses an *N*-glycosylation pathway lacking typical complex processing, with some novel variations, and several Golgi-associated *O*-linked pathways whose products, including phosphoglycosylation, are immunogenic (Freeze, 1997; in Varki *et al.*, 1999). The International *Dictyostelium* DNA sequencing efforts have nearly saturated the genome, and laboratory strains of this haploid organism are well suited for biochemical analysis, in addition to knockout and knockin genetic manipulations and restriction enzyme–mediated insertional mutagenesis screens to investigate function (Sucgang *et al.*, 2000).

Skp1 is a small (M_r = 21,000), rather remarkable protein found in several multiprotein complexes of the cytoplasm and nucleus of yeast, plants, invertebrates, and vertebrates. Best characterized is the SCF (*S*kp1, *c*ullin-1, *F*-box protein, Roc1/Rbx1) subunit of the E3 ubiquitin ligase complex that targets specific phosphoproteins, including cell cycle regulatory proteins and transcriptional factors, for polyubiquitination and subsequent degradation in proteasomes (Deshaies, 1999). Phosphoprotein target specificity is mediated by the F-box protein subunit, represented by 35 different gene products in mammals. The F-box domain contacts Skp1, which in turn interacts with cullin-1 and an E2 ubiquitin-conjugating enzyme. Recent evidence suggests that SCF can alternatively bind Snf1-related protein kinases and proteasomes (Farras *et al.*, 2001). Skp1 also appears to occur in non-SCF complexes, including the centrosomal CBF3 complex, a kinetochore complex, the RAVE-complex that might regulate assembly of the vacuolar proton transporter, and another complex(es) possibly involved in membrane trafficking (Seol *et al.*, 2001), altogether suggesting a special role for Skp1 at the junction of multiple intracellular regulatory pathways.

Glycosylation of Skp1 was originally detected in *Dictyostelium* by metabolic incorporation of [^{3}H]Fuc (Gonzalez-Yanes *et al.*, 1992), and subsequently confirmed by compositional analysis of the purified protein showing the presence of GlcNAc, Fuc, Gal, and possibly Xyl (Kozarov *et al.*, 1995; Teng-umnuay *et al.*, 1998). Purification of radioactivity from [^{3}H]Fuc-labeled Skp1 peptides showed only a single site of incorporation of Fuc into the protein (Teng-umnuay *et al.*, 1998). Tandem mass spectrometry (MS) and Edman sequencing showed that the glycopeptide contains a pentasaccharide attached to a HyPro residue at position 143 (Figure 1). Based on exoglycosidase digestions, the core trisaccharide has the structure of the type 1 blood group H antigen and is modified by two α-linked Gals. Coordinate loss of the outer three sugars during mild acid hydrolysis suggested linkage of the outer Gals to Fuc, consistent with the absence of significant α-galactosylation in a GDP-Fuc-deficient mutant, but this has not been determined directly.

Confirmation of the attachment of the glycan to Pro143 has come from three lines of evidence: (1) treating cells with inhibitors of prolyl 4-hydroxylases results in a small, downward shift in apparent M_r (Sassi *et al.*, 2001); (2) substitution of P143 of expressed Skp1 with Ala, a change that occurs naturally in some putative Skp1 genes of *Caenorhabditis elegans*, results in a similar downward shift in apparent M_r; and (3) a synthetic peptide including the equivalent of 4-HyPro143 is a good substrate for the first GTase of the pathway (Teng-umnuay

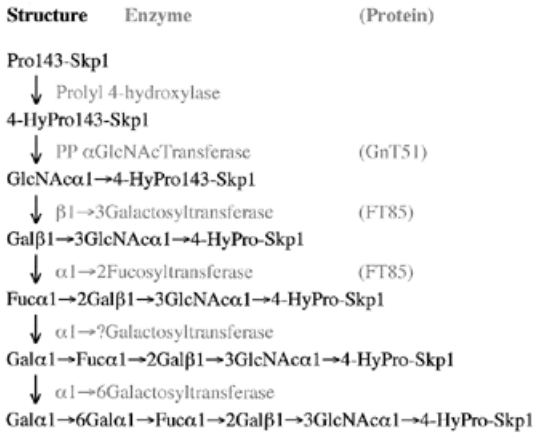

Fig. 1. The *Dictyostelium* Skp1 modification pathway. The sequential pathway deduced from the structure of the pentasaccharide is shown from top to bottom. The modifying enzymes and corresponding proteins or potential genes are shown at each step. Each activity has been detected in cytosolic extracts, and genes have been cloned for the first three GTase activities. The positions of the outer α-linked Gal residues are tentative (see text).

et al., 1999). The latter result suggests that Skp1 Pro143 is normally hydroxylated at the 4-position. The GlcNAc-HyPro143 linkage is alkali-resistant (Spiro, 1972) and differs from the sugar-HyPro linkages of secreted plant and algal proteins, which contain Ara or Gal (Kieliszewski *et al.*, 1995).

The great majority of Skp1 in both growing and developing *Dictyostelium* cells appears to be fully glycosylated based on western blot M_r analysis of whole cell Skp1 (Sassi *et al.*, 2001). Matrix-assisted laser desorption and ionization time-of-flight MS analysis of peptides from purified Skp1 supports this conclusion but does also detect unglycosylated, unhydroxylated peptide (Teng-umnuay *et al.*, 1998). Evidence therefore suggests that Skp1 glycosylation is nearly stoichiometric.

The possibility that Skp1 is similarly glycosylated in other organisms remains largely unexplored. Preliminary sugar composition studies show that highly overexpressed Skp1 in *Saccharomyces cerevisiae* is not significantly glycosylated, and treatment of cultured mammalian cells with prolyl 4-hydroxylase inhibitors fails to shift their apparent M_r values (West *et al.*, unpublished data). The equivalent of Pro143 occurs in microbial, plant, and invertebrate Skp1s, but not in vertebrate homologs, though a related Pro residue is found nearby (West *et al.*, 1997). It is difficult to draw general conclusions from these observations, however, because protein *O*-glycosylation is notoriously polymorphic both phylogenetically and developmentally. Direct structural studies on other Skp1s awaits purification of native material from other organisms. A search for other fucoproteins in the cytosolic fraction of *Dictyostelium* has failed to yield other candidates based on either metabolic labeling or *in vitro* fucosylation of extracts from fucosylation deficient cells (West *et al.*, 1996). However, these results do not rule out membrane-associated, developmentally regulated or low-abundance fucoproteins.

Potential functions of Skp1 glycosylation

As cited above, Skp1 participates in a newly emerging network of intracellular regulatory pathways. Consistent with multiple functions, Skp1 localizes to several subcellular sites (Freed *et al.*, 1999; Gstaiger *et al.*, 1999; Sassi *et al.*, 2001). Biochemically, soluble Skp1 purifies into major and minor pools (West *et al.*, 1997) that appear to differ in the content of a single Gal residue, and a fraction of Skp1 is associated in a salt-sensitive fashion with the microsomal fraction in *Dictyostelium* (Kozarov *et al.*, 1995). Multiple genetic isoforms of Skp1 exist in most eukaryotes in addition to potential glycoforms. In *Dictyostelium*, there is no evidence for a correlation of either of the two genetic isoforms with subcellular location (Sassi *et al.*, 2001) or biochemical fractionation (West *et al.*, 1997), and the great majority (>90%) of these Skp1s are normally glycosylated. However, these results do not exclude the possibility that specific Skp1 functions are mediated by structural variants.

The best available clue for why Skp1 is glycosylated is that when it is inhibited mutationally or pharmacologically, Skp1 is no longer concentrated in the nucleus (Sassi *et al.*, 2001). Mutant analysis shows that the core Galβ1,3GlcNAc disaccharide is sufficient for this function. Proteasomes are enriched in the nucleus, suggesting that nuclear Skp1 might belong to the SCF complex. If the SCF complex is assembled in the cytoplasm, Skp1 might enable its nuclear targeting. Consistent with such a role, the position of the Skp1 pentasaccharide near its C-terminus appears to be oriented away from known contacts with other proteins in the SCF complex (Schulman *et al.*, 2000). However, the effect of glycosylation may be indirect, because the ability of Skp1 to target a Skp1/GFP hybrid to the *Dictyostelium* nucleus does not depend on glycosylation (Erogul, 2001). Nonglycosylated Skp1 can accumulate to normal levels and is stable when cells dedifferentiate (Sassi *et al.*, 2001), as if glycosylation is not required for stability.

The addition of specific sugars is inhibited in Skp1s expressed with point mutations elsewhere in the protein (Sassi *et al.*, 2001), raising the possibility that glycosylation depends on normal Skp1 structure. Though this might simply reflect the inherent specificity of the modification enzymes, it is intriguing to consider these results from a quality control perspective. Roles in quality control of folding have been advanced for collagen prolyl 4-hydroxylation (Lamande and Bateman, 1999) and *N*-linked glycan processing of proteins in the RER (Helenius and Aebi, 2001). If glycosylation is required to certify that Skp1 has folded and assembled into multiprotein complexes properly, this could explain why unglycosylated Skp1 is unable to accumulate in the nucleus.

Another potential role for Skp1 modification is as a metabolic sensor that depends on adequate supply of cofactors and substrates for the modification enzymes. For example, physiologic O_2 availability appears to be rate-limiting for the prolyl 4-hydroxylation of a hypoxia-sensitive transcriptional factor in mammals (hypoxia-inducing factor-α), thereby regulating its ubiquitinylation by the SCF-like von Hippel-Lindau complex (Ivan *et al.*, 2001; Jaakkola *et al.*, 2001). The level of *O*-β-GlcNAc modification of cytoplasmic/nuclear proteins has been suggested to depend on the availability of Glc and GlcNAc over physiological ranges (Akimoto *et al.*, 2001). Fucosylation of Skp1 *in vitro* is highly sensitive to inhibition by nucleotides (West *et al.*, 1996). Reduced O_2 availability might explain why Skp1 expressed during multicellular development is not glycosylated (Sassi *et al.*, 2001).

Mechanism of Skp1 glycosylation

Skp1 glycosylation is expected to require a prolyl hydroxylase and five GTases (Figure 1). The enzyme activities have each been assayed using various mutant and recombinant Skp1s that are incompletely hydroxylated or glycosylated *in vivo* as acceptor substrates. Based on these assays, Skp1 appears to be modified sequentially by a pathway of soluble enzymes in the cytoplasmic rather than the secretory compartment of the cell. Glycosylation can occur posttranslationally *in vivo*, as Skp1 produced in prespore cells becomes glycosylated when cells are dedifferentiated into vegetative cells (Sassi *et al.*, 2001). Overexpressed Skp1 is efficiently glycosylated, suggesting a high capacity for the pathway consistent with the limited glycosylation microheterogeneity of normal Skp1.

Prolyl hydroxylase

The mechanism of prolyl hydroxylation appears to be related to that of collagen, based on the ability of low concentrations of known inhibitors of prolyl 4-hydroxylases to cause a downward shift in M_r similar to that caused by mutation of the Pro143 attachment site to an Ala residue (Sassi *et al.*, 2001). A search

of *Dictyostelium* cDNA and genomic DNA sequence databases reveals a putative gene (gene f01884; http://dicty.sdsc.edu/annotationdicty.html) similar to the prolyl 4-hydroxylase encoded by the above-mentioned *Chlorella* viral genome (Eriksson *et al.*, 1999). The PBCV-1 prolyl 4-hydroxylase contains a canonical catalytic domain but lacks the N-terminal domain thought to interact with protein disulfide isomerase in the RER. *Dictyostelium* gene f01884 contains all five of the key conserved residues considered to be diagnostic for these enzymes but, in contrast to known prolyl 4-hydroxylases, lacks sequences for targeting its product to the RER, where prolyl 4-hydroxylases usually reside (Kivirikko and Pihlajaniemi, 1998). Hydroxylation of Skp1 Pro143 is thus suggested to occur by a conventional posttranslational mechanism using a prolyl 4-hydroxylase that is novel both in its compartmentalization and in its lack of a requirement for repeat motifs typically found in prolyl 4-hydroxylase substrates.

GlcNAcTase

The Skp1 GlcNAcTase was assayed by measuring transfer [^{3}H] of from UDP-[^{3}H]GlcNAc to a mutant Skp1 deficient in GlcNAc (Teng-umnuay *et al.*, 1999). A synthetic peptide containing 4-HyPro143 was a moderately competitive substrate, whereas nonhydroxylated forms of Skp1 or peptide were not. Little or no transferase activity was detected in microsomal fractions, suggesting the enzyme is cytosolic. The GlcNAcTase activity is dependent on a divalent cation and reducing conditions, and exhibits submicromolar K_ms for its substrates, properties expected to support efficient function in the cytoplasm. Purification of the Skp1 GlcNAcTase activity was facilitated by the construction of a novel affinity resin linking UDP-GlcNAc at its 5-uridyl position that might be applicable to other enzymes of this class. Activity was associated with a single M_r 51,000 protein (GnT51), after 130,000-fold purification to near homogeneity, which contains the binding site for UDP-GlcNAc based on photoaffinity labeling.

Peptide sequences obtained from tryptic digests of GnT51 were sequenced *de novo* on a hybrid quadrupole time-of-flight (Q-TOF) tandem MS instrument (West *et al.*, 2001), which was well-suited for the limited amount of protein available from the purification. These sequences made possible a polymerase chain reaction–based approach to identify partial GnT51 genomic sequence which was extended with sequences from the International *Dictyostelium* genome project. This yielded a predicted 2-exon ORF that encodes a protein with an M_r of 52,000 similar to that obtained by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and contains many of the GnT51 peptides characterized by Q-TOF MS. Expression of the predicted cDNA in *Escherichia coli* resulted in Skp1 GlcNAcTase activity in the soluble cell extracts. The N-terminus of the predicted sequence lacks motifs for targeting to the RER, consistent with the biochemical evidence that this is a cytoplasmic enzyme.

The N-terminal 260 amino acids of GnT51 display sequence similarity to the catalytic domain of mucin-type UDP-GalNAc:Ser/Thr polypeptide (PP) αGalNAcTases of family 27 (Campbell *et al.*, 1997), though this was too weak to be detected by BLAST algorithms (Altschul *et al.*, 1990, 1997). Greatest similarity lies in two regions: the N-terminal 125 amino acids known as the NRD2-subdomain (Kapitonov and Yu, 1999), which includes a DxH sequence related to the commonly occurring DxD motif, and downstream in the previously recognized Gal/GalNAc box (Hagen *et al.*, 1999), a sequence of about 60 amino acids in what we call the cat27 domain. The alignment in Figure 2A highlights the regions of greatest similarity, and is shaded to emphasize the chemical relatedness of amino acid positions in light of the divergence of specific amino acid identities. Sequences from more distantly related families are compared to illustrate the significance of the family 27 relatedness. A phylogenetic analysis of the sequences supports the close relationship of GnT51 to family 27 (Figure 3, green clade). Family 7 β4GalTases and certain putative family 2 GTases with unknown function are also found near the base of this clade, but GnT51 appears to be more related to family 27 sequences based on the greater motif similarity seen in the Figure 2A alignment and the presence of a short C-terminal domain that recognizes acceptor substrates in some family 27 enzymes (Hassan *et al.*, 2000). Animal PP GalNAcTases catalyze formation of GalNAcα1-Ser/Thr linkages, which are related to the GlcNAc-HyPro linkage formed by the Skp1 GlcNAcTase. By analogy, the GlcNAc of Skp1 is predicted to be α-linked, in contrast to the β-linkage of GlcNAc catalyzed by the unrelated cytoplasmic/nuclear Ser/Thr *O*-β-GlcNAcTase (Comer and Hart, 2000).

A BLAST search for potential GnT51 homologs identified candidate genes in *Yersenia pestis*, *Dictyostelium*, and *Leishmania major* (labeled *Ye*, *Dd*, and *Lm*, respectively, in figures 2 and 3). The prokaryotic (*Yersenia*) gene is predicted to encode a cytoplasmic protein, and the latter two (eukaryotic microbial) genes encode predicted type II Golgi membrane proteins. All three sequences contain key DxH and Gal/GalNAc motifs of family 27 α-HexNAcTases (Figure 2A), and this relation is supported by the phylogenetic analysis of the sequences shown in Figure 3 (green clade). Functional support comes from the finding that a null mutant for the *Dictyostelium* gene (*cis4C* for cisplatin-resistance) (Li *et al.*, 2000) is deficient in the *O*-glycosylation of the spore-associated proteins SP29 and SP85 (Kaplan and West, unpublished data), which have mucin-like domains containing clustered GlcNAc-residues in α-linkage to Ser/Thr (Zachara *et al.*, 1996; Zhang *et al.*, 1999). The closer relation of GnT51 to these predicted enzymes compared to the animal family 27 sequences argues that this family evolved prior to eukaryotic radiation and that Golgi compartmentalization of family members occurred more than once.

β1,3Galactosyltransferase and α1,2fucosyltransferase

The second and third sugars of the Skp1 glycan are added by a single protein, FT85, with two GTase domains. FT85 was originally purified from the cytosolic fraction as the FucTase, which was assayed as transfer of [^{3}H] from GDP-[^{3}H]Fuc to Skp1(*modC*) from a GDP-Fuc synthesis mutant, resulting in formation of a Fucα1,2Gal linkage (West *et al.*, 1996; Trinchera and Bozzarro, 1996). Little or no activity was detectable in the microsomal fraction, which was not due to an inhibitor based on mixing studies. Purification over 1 million–fold led to the identification of an M_r 85,000 protein, FT85, which copurified with the activity and was photoaffinity-labeled with GDP-hexanolamine-[^{125}I]ASA (West *et al.*, 1996). Like the Skp1-GlcNAcTase, the FucTase also has properties of a cytoplasmic enzyme in its requirement for a reducing agent, its neutral pH optimum, and submicromolar

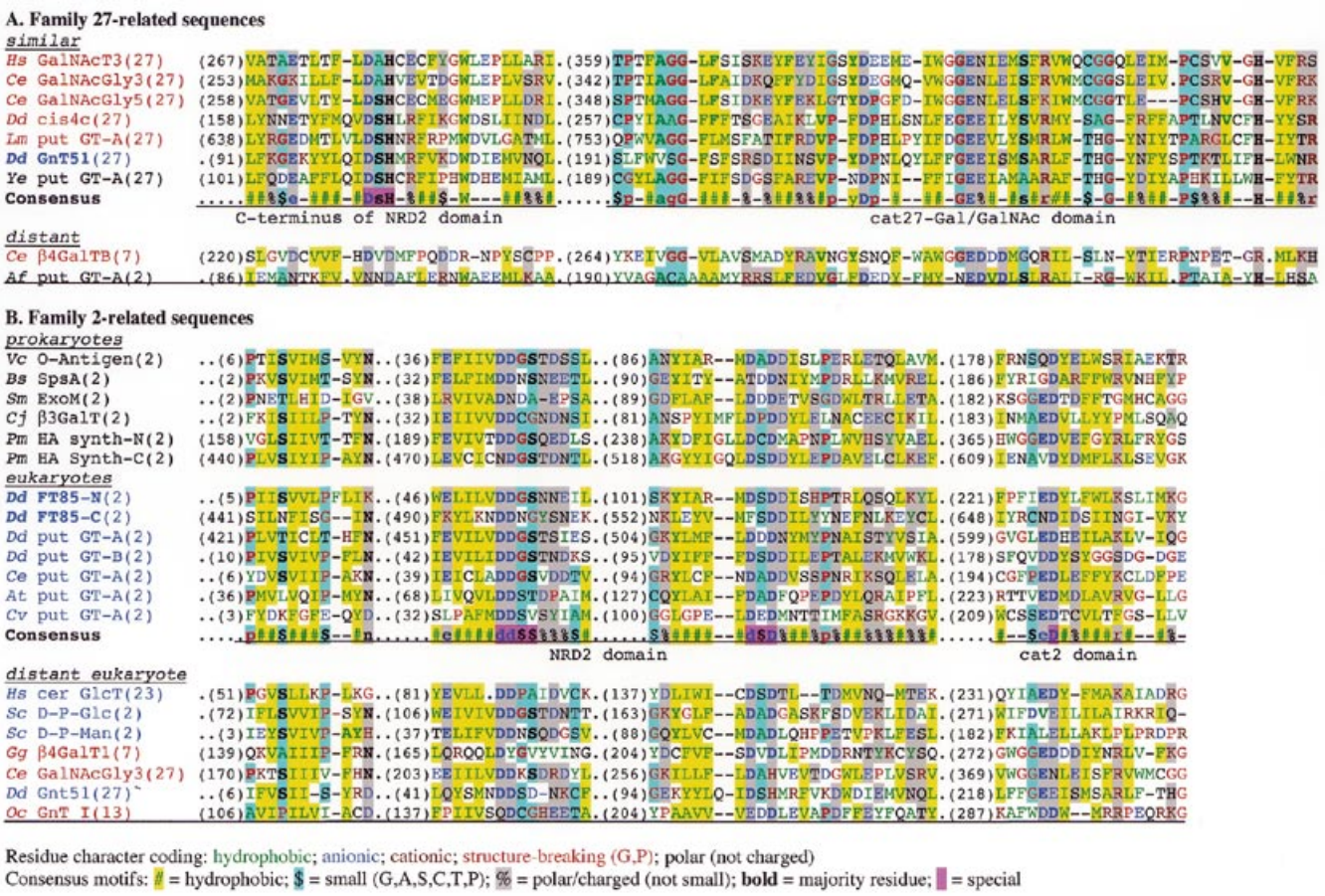

Fig. 2. Sequence alignments of GTase catalytic domains. (**A**) Sequences related to the Skp1 GlcNAcTase (GnT51). Upper group: family 27 GTases; lower group: more distantly related GTases. (**B**) Sequences related to the Skp1 β-GalTase/α-FucTase (FT85). Upper group: prokaryotic family 2 GTases; middle group: eukaryotic soluble GTases, most of which are hypothetical (*put*ative); lower group: more distantly related GTases. Names are color-coded corresponding to whether the (predicted) protein product is compartmentalized in the eukaryotic Golgi (red), eukaryotic cytoplasm (blue), or prokaryotic cytoplasm (black). GenBank gi numbers are given in Figure 3, and family classifications (Campbell *et al.*, 1997) are in parentheses. Sequences initially identified by BLAST were aligned using CLUSTAL and then manually optimized to maximize matching of sequence motifs, hydrophobic or polar residues, and location of gaps adjacent to P or G residues in a larger master alignment of 46 sequences (not shown). Hydrophobic residues are green, acidic residues are blue, basic residues are dark red, P and G are bright red, and polar, noncharged residues are black. Because of the high degree of sequence divergence, the alignments are highlighted to emphasize chemical similarities at each position. Positions where the majority of residues have similar chemical characteristics are highlighted in yellow if hydrophobic; teal if small, hydroxylated, or Pro; or gray if polar/charged; and tabulated in the consensus sequence as hash mark, dollar sign, or percent symbol, respectively. Key motifs cited in the text are highlighted in magenta. Numbers in parentheses identify the position of the next amino acid position to the right; periods denote deleted sequence.

K_ms for its substrates. It is remarkably sensitive to cytochrome *c* and GDP *in vitro*, the significance of which is not known.

FT85 can also fucosylate the acceptor disaccharide Galβ1,3GlcNAc when attached to a hydrophobic aglycon, albeit with a K_m three orders of magnitude higher (West *et al.*, 1996). Galβ1,4GlcNAc (type 2) and Galβ1,6GlcNAc (type 3) moieties are much poorer acceptors. GalNAc can substitute for GlcNAc and may be either α- or β-linked to the aglycone. The substrate preferences of the enzyme most resembles that of the mammalian Se-type α1,2-FucTase, but, in contrast to the Golgi enzyme, the Skp1 FucTase is absolutely dependent on a divalent cation such as Mg^{2+} for activity.

The proteomics strategy used to clone the GnT51 gene was also successfully applied to FT85 (Van der Wel *et al.*, 2001a), yielding a predicted coding region of 768 amino acids. FT85 is required for fucosylation of Skp1 *in vivo* as determined by targeted disruption of its gene locus. In addition, expression of FT85 in *E. coli* rendered soluble extracts capable of fucosylating highly purified Skp1. Consistent with the biochemical data, FT85 lacks known motifs for targeting to the RER or for membrane association. Thus FT85 is necessary and sufficient for Skp1 FucTase function in the cytoplasm (or nucleus) of the cell.

Skp1(FT85$^-$) from the FT85-null strain was fortuitously discovered to be a good substrate for a UDP-Gal-dependent GalTase activity in the crude cytosolic extract (Van der Wel *et al.*, 2001b). MS structural studies carried out as described previously (Teng-umnuay *et al.*, 1998) showed that the Skp1(FT85$^-$) glycopeptide contains only the reducing terminal GlcNAc attached to HyPro. FT85-null extracts are devoid of Skp1 βGalTase activity, and highly purified *Dictyostelium* FT85 and recombinant FT85, exhibit abundant activity. The

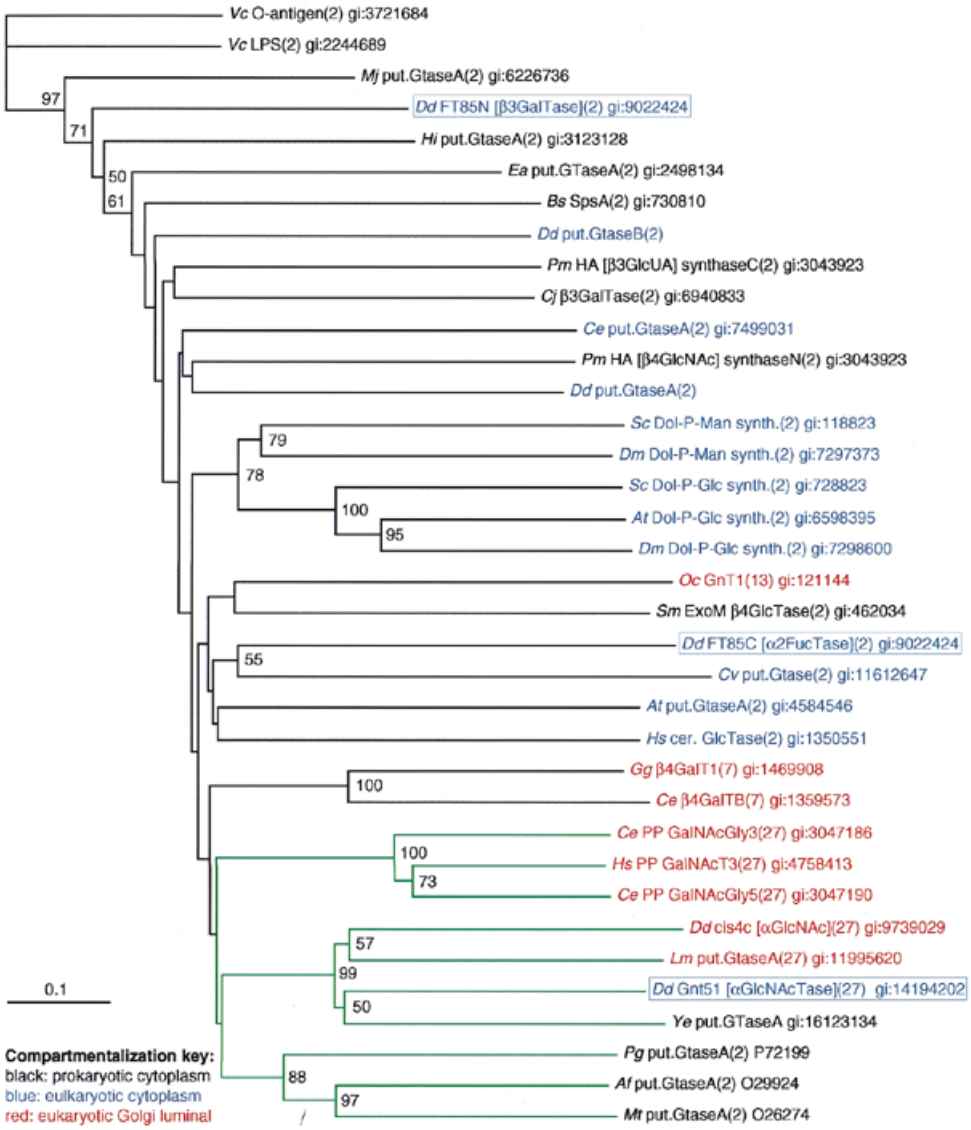

Fig. 3. Phylogenetic tree of the GTase catalytic domains aligned in Figure 2. Prokaryotic sequence names are given in black, eukaryotic cytosolic sequence names in blue, and eukaryotic Golgi sequence names in red. The clade containing family 27–related sequences is in green. Sequences were chosen on the basis of highest BLAST similarity scores against the FT85 N-domain or GnT51 catalytic domain, or relatedness in terms of activity (e.g., β1,3GalTase) or protein architecture (e.g., bifunctional hyaluronan synthase activity). Coalignment of family 2 SpsA, family 13 GnT-I, and family 7 β4GalTases was based on the structure-based alignment shown in Unligil and Rini (2000). Examples of the sequence alignments on which the phylogram is based are in Figure 2. The phylogenetic relationships of the aligned sequences were analyzed using a distance algorithm, because a parsimony method was unable to resolve the Dol-P-Hex synthases or the family 27 GTases from the others. Sequences corresponding to residues 5–128 and 195–254 of FT85 were analyzed. Trees were generated using PAUP* (Swofford, 2000) under the minimum evolution criterion with tree bisection reconnection branch swapping. Heuristic searches were performed with 10,000 replicates. Bootstrap values are given when they exceed 50%. Other family 2 GTases, such as cellulose synthases and class I hyaluronan synthases, are not included because the divergence of their cat2 domains precluded confident alignment of their sequences.

reaction product is susceptible to β1,3-galactosidase digestion. Thus FT85 is a bifunctional glycosyltransferase.

The sequence of FT85 harbors two GTase-like domains, each about 250 amino acids, that are characteristic of bacterial family 2 inverting GTases (Campbell *et al.*, 1997). As shown by the sequence alignment in Figure 2B, the putative domains each contain a variation of the NRD2-subdomain described above for the Skp1 GlcNAcTase, including characteristic Asp residues and a DxD motif (highlighted in magenta) critical for coordinating the sugar nucleotide donor (Charnock and Davies, 1999; Garinot-Schneider *et al.*, 2000). The second half of each putative domain contains sequences associated with the predicted catalytic base of family 2 domains (highlighted in magenta). This region is highly divergent between family 2 and family 27 enzymes correlating with inversion and retention, respectively, of the configuration of the sugar linkages formed. The alignments also include more distantly related sequences from family 2 and other families for comparison.

The family 2 catalytic domain has recently been recognized as belonging to a GTase superfamily that includes the eukaryotic Golgi GTases GlcNAcTase-I (initiates complex N-glycan processing), β1,4GalTase, α1,3GalTase, and a GlcUATase based on structural homologies (Unligil and Rini, 2000; Pederson *et al.*, 2000; Gastinel *et al.*, 2001; Persson *et al.*, 2001).

Recombinant expression of the FT85 N- and C-terminal domains in FT85-mutant cells shows that they mediate the Skp1 βGalTase and FucTase activities, respectively, confirming the two-domain model predicted by the sequence analysis. Amino acid substitutions in these domains specifically diminish the corresponding activities and support the family 2 classification (Van der Wel *et al.*, unpublished data). A phylogenetic analysis of the sequences shows that the β1,3GalTase domain is more related to a bacterial (Figure 3) than known eukaryotic β1,3GalTases, which could not be coaligned. The α1,2FucTase C-terminal domain is found in a clade enriched in eukaryotic GTase or GTase-like sequences and shows no resemblance to previously described α1,2FucTases (Oriol *et al.*, 1999). The FT85 GTase domain sequences are intermingled among evolutionary lineages that have produced multiple modern prokaryotic and eukaryotic GTases, which are themselves intermingled (albeit with low bootstrap support; Figure 3). These observations suggest that the eukaryotic family 2 GTases evolved multiple times, possibly along lineages corresponding to catalytic specificity rather than membrane anchorage or the transition from prokaryotes to eukaryotes.

The two-domain architecture of FT85 resembles that of bifunctional prokaryotic GTases involved in the synthesis of hyaluronan (Jing and DeAngelis, 2000), chondroitin (DeAngelis and Padgett-McCue, 2000), and other bacterial polysaccharides. For example, the class II hyaluronan synthase consists of two family 2 domains that catalyze formation of the alternating β1,3GlcUA- and β1,4GlcNAc-linkages of hyaluronan, based on mutagenesis studies. Bifunctional GTases have also been implicated in the formation of terminal KDO-KDO and NeuAc-NeuAc linkages in bacteria (e.g., Belunis and Raetz, 1992; Gilbert *et al.*, 2000), cellulose (Saxena and Brown, 2000), heparan by EXT1 or EXT2 in the mammalian Golgi (Senay *et al.*, 2000), and animal hyaluronan by class I hyaluronan synthase (Yoshida *et al.*, 2000). FT85 is the first documented example of a two-domain bifunctional GTase in a eukaryote and is the first known to modify a protein substrate. This might ensure processive extension of the Skp1 glycan in the cytoplasm, where spatial control and compartmental protection as in the RER and Golgi are not available.

Other prospective cytoplasmic GTases

The compartmentalization of family 2 GTases in the cytoplasm was previously noted by Kapitonov and Yu (1999). In prokaryotes, the family 2 GTase domain is found in enzymes at the cytoplasmic face of the plasma membrane, where they contribute to the synthesis of polysaccharides and lipid-linked precursors of capsules and exopolysaccharides prior to their export to the cell surface (Whitfield and Roberts, 1999). A topologically similar function is employed in eukaryotes for the synthesis of cellulose, chitin, hyaluronan, and other polysaccharides at the plasma membrane; Dol-P-Man and Dol-P-Glc at the cytoplasmic surface of the RER; and Glc-Cer at the Golgi surface. FT85 conserves the cytoplasmic localization of family 2 GTase domains, except that it is liberated from the membrane surface and extends the range of family 2 GTase acceptor substrates to include proteins.

The evolution of the Skp1 GlcNAcTase represents a similar scenario with the proviso that some of its closest known relatives are the mucin-type polypeptide αGalNAcTases of the Golgi. The phylogenetic analysis suggests that these type II membrane proteins have evolved multiple times by the simple appending of an N-terminal signal anchor and adjacent stalk region. This model may also apply to a recently identified candidate GTase for cytoplasmic glycosylation of PBCV-1 viral capsid VP54 (Graves *et al.*, 2001), as its predicted sequence is most similar to Golgi GTases from families 31 and 34 (Campbell *et al.*, 1997).

Based on the Skp1 GTase studies, proteins containing family 2 and 27 catalytic domain-like sequences, themselves related, might contribute to the putative glycosylation of other cytoplasmic proteins summarized above and listed in Table I. A BLAST search in GenBank and other databases for new eukaryotic family 2– and 27–like sequences that lack RER targeting sequences and are not apparently related to known subclasses that modify lipid or polysaccharide targets (e.g., Dol-P-Hex synthases or containing the processivity motif QXXRW), has yielded candidate open reading frames containing family 2 catalytic domains in the genomes of *C. elegans*, *Arabidopsis thaliana*, PBCV-1, and *D. discoideum* (Figure 2B). The absence of any candidate soluble family 2 GTases from vertebrates is notable but may be due to sequence divergence resulting in the failure of BLAST programs to identify homologous sequences. A phylogenetic analysis of these putative eukaryotic family 2 cytoplasmic enzymes suggests that they evolved multiple times from prokaryotes (Figure 3). For example, they are more related to other prokaryotic GTase genes than to each other, and they appear to have diverged prior to the evolution of the family 2 Dol-P-Hex synthases and the related family 23 ceramide GlcTases based on their distinct clading. These predicted enzymes appear to have remained in the cytoplasm when eukaryotes evolved from prokaryotes.

Two additional GTase families, out of the 55 total recognized by Henrissat (Campbell *et al.*, 1997), contain cytoplasmic members that glycosylate eukaryotic polysaccharides and lipids. For example, glycogenin, the GlcTase that primes glycogen synthesis (Lin *et al.*, 1999), is classified in family 8 with other prokaryotic and eukaryotic presumptive GTases of unknown function, and these are candidates for mediating cytoplasmic glycosylation steps that retain the sugar linkage, based on the precedent of the family 2 member FT85. A similar argument applies to family 4, whose eukaryotic members include PigA, which forms the $GlcNH_2$-inositol linkage in glycophosphatidylinositol-anchor synthesis (Tiede *et al.*, 1999); sucrose synthase; and a plant GalTase that forms digalactosyldiacylglycerol (Froehlich *et al.*, 2001), each of which is oriented toward the cytoplasm.

The frequent addition of *O*-β-GlcNAc to Ser or Thr residues of animal cytoplasmic and nuclear proteins is mediated by a single, novel, soluble, metal-independent family 41 GTase (reviewed in Comer and Hart, 2000). A related enzyme is expressed in plants (Thornton *et al.*, 1999) and probably *Dictyostelium*. An unrelated *O*-β-GlcNAcTase is secreted

from *Clostridium novyi* and enters eukaryotic target cells to modify Rho GTP-binding proteins (Busch *et al.*, 1998). Unlike the α-linked GlcNAc of Skp1, there is as yet no evidence for the extension of *O*-β–linked GlcNAc, but future studies might implicate these enzymes for initiating complex glycosylation of cytoplasmic proteins.

Context of complex cytoplasmic protein glycosylation

Although complex protein-linked glycans are certainly less prominent than *O*-β-GlcNAc, their potential significance is enhanced by the novel functions tentatively associated with their distinctive sugar composition and linkages. Gal- and GalNAc-containing glycans might associate with galectins or discoidins that are prevalent in the cytoplasm (Cooper and Barondes, 1999) and nucleus (Vyakarnam *et al.*, 1998), and have been implicated in RNA splicing (Park *et al.*, 2001) and anti-apoptosis (Akahani *et al.*, 1997). A protein that binds both actin and mannose *in vitro* (Jung *et al.*, 1996) may be a cytoskeletal receptor for a cytoplasmic mannoprotein. Although the size of Skp1's glycan chain may evoke the expectation of a cognate receptor, complex cytoplasmic glycans might alternatively (1) regulate other modifications, such as phosphorylation as for *O*-β-GlcNAc; (2) monitor protein folding as observed for *N*-glycans on proteins in the RER; or (3) reflect the availability of metabolic precursors required for the modification.

The characterization of cytoplasmic hydroxylation and glycosylation is technically challenging, but new MS-based proteomics approaches promise to overcome this barrier (Dell and Morris, 2001). As shown by the application of Q-TOF MS to both characterizing the glycan and sequencing rare GTases that modify Skp1, it is clear that complex cytoplasmic glycosylation has evolved in eukaryotic microorganisms and that this pathway is functionally important. There are strong hints of hydroxylation and complex glycosylation of other cytoplasmic/nuclear proteins in higher plants and animals and of cytoplasmic enzymes that might mediate these modifications. However, as previously observed for Golgi GTases, those of the cytoplasm seem to be equally evolutionarily divergent, necessitating a continuing need for new lead GTase sequences to seed BLAST searches and for refined informatics approaches to identify subtly related sequences already in the databases.

Acknowledgments

The authors are grateful to the many members of the glycobiology and *Dictyostelium* communities who have generously provided invaluable reagents and advice, including Anne Dell, Howard Morris, Monica Palcic, Jacques Baenziger, Adriana Manzi, Eric Holmes, Khushi Matta, Al Elbein, Bernard Henrissat, Gideon Davies, James van Etten, Hud Freeze, Bill Loomis, Dave Knecht, Rick Firtel, Frantisek Puta, and Gerry Weeks. The authors' research in this area has been supported by grants from the NIH (GM-37539) and the Florida Division of the American Cancer Society.

Abbreviations

GTase, glycosyltransferase; MS, mass spectrometry; Q-TOF, hybrid quadrupole time-of-flight; PBCV-1, *Paramecium bursaria* Chlorella virus-1; PP, polypeptide; RER, rough endoplasmic reticulum; SCF; *S*kp1, *c*ullin-1, *F*-box protein.

References

Akahani, S., Nangia-Makker, P., Inohara, H., Kim, H.R., and Raz, A. (1997) Galectin-3: a novel antiapoptotic molecule with a functional BH1 (NWGR) domain of Bcl-2 family. *Cancer Res.*, **57**, 5272–5276.

Akimoto, Y., Kreppel, L.K., Hirano, H., and Hart, G.W. (2001) Hyperglycemia and the *O*-GlcNAc transferase in rat aortic smooth muscle cells: elevated expression and altered patterns of *O*-GlcNAcylation. *Arch. Biochem. Biophys.*, **389**, 166–175.

Alonso, M.D., Lomako, J., Lomako, W.M., and Whelan, W.J. (1995) A new look at the biogenesis of glycogen. *FASEB J.*, **9**, 1126–1137.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Belunis, C.J. and Raetz, C.R.H. (1992) Biosynthesis of endotoxins. *J. Biol. Chem.*, **267**, 9988–9997.

Burda, P. and Aebi, M. (1999) The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta*, **1426**, 239–257.

Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D., and Aktories, K. (1998) A common motif of eukaryotic glycosyltransferases is essential for the enzyme activity of large Clostridial cytotoxins. *J. Biol. Chem.*, **273**, 19566–19572.

Busch, S.J., Martin, G.A., Barnhart, R.L., Mano, M., Cardin, A.D., and Jackson, R.L. (1992) *Trans*-repressor activity of nuclear glycosaminoglycans on Fos and Jun/AP-1 oncoprotein-mediated transcription. *J. Cell Biol.*, **116**, 31–42.

Cacan, R. and Verbert, A. (2000) Transport of free and N-linked oligomannoside species across the rough endoplasmic reticulum membranes. *Glycobiology*, **10**, 645–648.

Campbell, J.A., Davies, G.J., Bulone, V., and Henrissat, B. (1997) A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.*, **326**, 929–939.

Chandra, N.C., Spiro, M.J., and Spiro, R.G. (1998) Identification of a glycoprotein from rat liver mitochondrial inner membrane and demonstration of its origin in the endoplasmic reticulum. *J. Biol. Chem.*, **273**, 19715–19721.

Charnock, S.J. and Davies, G.J. (1999) Structure of the nucleotide-diphospho sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry*, **38**, 6380–6385.

Comer, F.I. and Hart, G.W. (2000) *O*-glycosylation of nuclear and cytosolic proteins. *J. Biol. Chem.*, **275**, 29179–29182.

Cooper, D.N.W. and Barondes, S.H. (1999) God must love galectins; He made so many of them. *Glycobiology*, **9**, 979–984.

DeAngelis, P.L. (1999) Hyaluronan synthases: fascinating glycosyltranferases from vertebrates, bacterial pathogens, and algal viruses. *Cell. Mol. Life Sci.*, **56**, 670–682.

DeAngelis, P.L. and Padgett-McCue, A.J. (2000) Identification and molecular cloning of a chondroitin synthase from *Pasteurella multocida* type F. *J. Biol. Chem.*, **275**, 24124–24129.

Dell, A. and Morris, H.R. (2001) Glycoprotein structure determination by mass spectrometry. *Science*, **291**, 2351–2356.

Delmer, D.P. (1999) Cellulose biosynthesis: exciting times for a difficult field of study. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **50**, 245–276.

Deshaies, R.J. (1999) SCF and cullin/RING H2-based ubiquitin ligases. *Annu. Rev. Cell. Dev. Biol.*, **15**, 435–467.

Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K.I. (1999) Evidence for 4-hydroxyproline in viral proteins. Characterization of a viral prolyl 4-hydroxylase and its peptide substrates. *J. Biol. Chem.*, **274**, 22131–22134.

Erogul, J.Y. (2001) Examination of the role of genetic and post-translational variation in *Dictyostelium* Skp1. MS thesis, University of Florida.

Farras, R., Ferrando, A., Jasik, J., Kleinow, T., Okresz, L., Tiburcio, A., Salchert, K., del Pozo, C., Schell, J., and Koncz, C. (2001) SKP1-SnRK

protein kinase interactions mediate proteasomal binding of a plant SCF ubiquitin ligase. *EMBO J.*, **20**, 2742–2756.

Fedarko, N.S., Ishihara, M, and Conrad, H.E. (1989) Control of cell division in hepatoma cells by exogenous heparan sulfate proteoglycan. *J. Cell Physiol.*, **139**, 287–294.

Freed, E., Lacey, K.R., Huie, P., Lyapina, S.A., Deshaies, R.J., Stearns, T., and Jackson, P.K. (1999) Components of an SCF ubiquitin ligase localize to the centrosome and regulate the centrosome duplication cycle. *Genes Develop.*, **13**, 2242–2257.

Freeze, H.H. (1997) *Dictyostelium discoideum* glycoproteins: using a model system for organismic glycobiology. *New Compr. Biochem.*, **29**, 89–121.

Froehlich, J.E., Benning, C., and Dormann, P. (2001) The digalactosyldiacylglycerol (DGDG) synthase DGD1 is inserted into the outer envelope membrane of chloroplasts in a manner independent of the general import pathway and does not depend on direct interaction with monogalactosyldiacylglycerol synthase for DGDG synthesis. *J. Biol. Chem.*, **276**, 31806–31812.

Garinot-Schneider, C., Lellouch, A.C., and Geremia, R.A. (2000) Identification of essential amino acid residues in the *Sinorhizobium meliloti* glucosyltransferase ExoM. *J. Biol. Chem.*, **275**, 31407–31413.

Gastinel, L.N., Bignon, C., Misra, A.K., Hindsgaul, O., Shaper, J.H., and Joziasse, D.H. (2001) Bovine α1, 3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltranferases. *EMBO J.*, **20**, 638–649.

Gilbert, M., Brisson, J.-R., Karwaski, M.-F., Michniewicz, J., Cunningham, A.-M., Wu, Y., Young, N.M., and Wakarchuk, W.W. (2000) Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384. *J. Biol. Chem.*, **275**, 3896–3906.

Goletz, S., Hanisch, F.-G., and Karsten, U. (1997) Novel αGalNAc containing glycans on cytokeratins are recognized *in vitro* by galectins with type II carbohydrate recognition domains. *J. Cell Sci.*, **110**, 1585–1596.

Gonzalez-Yanes, B., Cicero, J.M., Brown, R.D. Jr., and West, C.M. (1992) Characterization of a cytosolic fucosylation pathway in *Dictyostelium*. *J. Biol. Chem.*, **267**, 9595–9605.

Graves, M.V., Bernadt, C.T., Cerny, R., and van Etten, J.L. (2001) Molecular and genetic evidence for a virus-encoded glyosyltransferase involved in protein glycosylation. *Virology*, **285**, 332–345.

Gstainger, M., Marti, A., and Krek, W. (1999) Association of human SCF (SKP2) subunit p19 (SKP1) with interphase centrosomes and mitotic spindle poles. *Exp. Cell Res.*, **247**, 554–562.

Hagen, F.K., Hazes, B., Raffo, R., deSa, D., and Tabak, L.A. (1999) Structure-function analysis of the UDP-*N*-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase. *J. Biol. Chem.*, **273**, 8268–8277.

Hart, G.W., Haltiwanger, R.S., Holt, G.D., and Kelly, W.G. (1989) Glycosylation in the nucleus and cytoplasm. *Annu. Rev. Biochem.*, **58**, 841–874.

Hassan, H., Reis, C.A., Bennett, E.P., Mirgorodskaya, E., Roepstorff, P., Honllingsworth, M.A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (2000) The lectin domain of UDP-N-acetyl-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase-T4 directs its glycopeptide specificities. *J. Biol. Chem.*, **275**, 38197–38205.

Heese-Peck, A., Colern, A., Borksenious, O.N., Hart, G.W., and Raikhel, N.V. (1996) Plant nuclear pore complex proteins are modified by novel oligosaccharides with terminal N-acetylglucosamine. *Plant Cell*, **7**, 1459–1471.

Helenius, A. and Aebi, M. (2001) Intracellular functions of N-linked glycans. *Science*, **291**, 2364–2369.

Hiscock, D.R.R., Yanagishita, M., and Hascall, V.C. (1994) Nuclear localization of glycosaminoglycans in rat ovarian granulosa cells. *J. Biol. Chem.*, **269**, 4539–4546.

Huang, L., Grammatikakis, N., Yoneda, M., Banerjee, S.D., and Toole, B.P. (2000) Molecular characterization of a novel intracellular hyaluronan-binding protein. *J. Biol. Chem.*, **275**, 29829–29839.

Hughes, R.C. (2001) Galectins as modulators of cell adhesion. *Biochimie*, **83**, 667–676.

Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J.M., Lane, W.S., and Kaelin, W.G. (2001) HIFα targeted for VHL-mediated destruction by proline hydroxylation: implications for O_2 sensing. *Science*, **292**, 464–468.

Jaakkola, P., Mole, D.R., Tian, Y.-M., Wilson, M.I., Gielbert, J., Gaskell, S.J., von Kriegsheim, A., Hebestreit, H.F., Mukherji, M., Schofield, C.J., and others (2001) Targeting of HIFα to the von Hippel-Lindau ubiquitinylation complex by O_2-regulated prolyl hydroxylation. *Science*, **292**, 468–472.

Jing, W. and DeAngelis, P.L. (2000) Dissection of the two transferase activities of the *Pasteurella multocida* hyaluronan synthase: two active sites exist in one polypeptide. *Glycobiology*, **10**, 883–889.

Jung, E., Fucini, P., Stewart, M., Noegel, A.A., and Schleicher, M. (1996) Linking microfilaments to intracellular membranes: the actin-binding and vesicle-associated protein comitin exhibits a mannose-specific lectin activity. *EMBO J.*, **15**, 1238–1246.

Kapitonov, D. and Yu, R.K. (1999) Conserved domains of glycosyltransferases. *Glycobiology*, **9**, 961–978.

Kieliszewski, M.J., O'Neill, M., Leykam, J., and Orlando, D.R. (1995) Tandem mass spectrometry and structural elucidation of glycopeptides from a hydroxyproline-rich plant cell wall glycoprotein indicate that contiguous hydroxyproline residues are the major sites of hydroxyproline-*O*-arabinosylation. *J. Biol. Chem.*, **270**, 2541–2549.

Kivirikko, K.I. and Pihlajaniemi, T. (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl-4-hydroxylases. *Adv. Enzymol. Relat. Areas Mol. Biol.*, **72**, 325–398.

Kozarov, E., van der Wel, H., Field, M., Gritzali, M., Brown Jr., R.D., and West, C.M. (1995) Characterization of FP21, a cytosolic glycoprotein from *Dictyostelium*. *J. Biol. Chem.*, **270**, 3022–3030.

Kussman, M., Hauser, K., Kissmehl, R., Breed, J., Plattner, H., and Roepstorff, P. (1999) Comparison of *in vivo* and *in vitro* phosphorylation of exocytosis-sensitive protein PP63/parafusin by differential MALDI mass spectrometric peptide mapping. *Biochemistry*, **38**, 7780–7790.

Lamande, S.R. and Bateman, J.F. (1999) Procollagen folding and assembly: the role of endoplasmic reticulum enzymes and molecular chaperones. *Sem. Cell Dev. Biol.*, **10**, 455–464.

Lee, J.Y. and Spicer, A.P. (2000) Hyaluronan: a multifunctional, megadalton, stealth molecule. *Curr. Opin. Cell Biol.*, **12**, 581–586.

Levin, S., Almo, S.C., and Satir, B.H. (1999) Functional diversity of the phosphoglucomutase superfamily: structural implications. *Prot. Eng.*, **12**, 737–746.

Li, G., Alexander, H., Schneider, N., and Alexander, S. (2000) Molecular basis for resistance to the anticancer drug cisplatin in *Dictyostelium*. *Microbiology*, **146**, 2219–2227.

Lin, A., Mu, J., Yang, J., and Roach, P.J. (1999) Self-glucosylation of glycogenin, the initiator of glycogen biosynthesis, involves an inter-subunit reaction. *Arch. Biochem. Biophys.*, **363**, 163–170.

Lowe, J.B. (2001) Glycosylation, immunity, and autoimmunity. *Cell*, **104**, 809–812.

Marchase, R.B., Richardson, K.L., Srisomsap, C., Drake, R.R., and Haley, B.E. (1990) Resolution of phosphoglucomutase and the 62-kDA acceptor for the glucosylphosphotransferase. *Arch. Biochem. Biophys.*, **280**, 122–129.

Marchase, R.B., Bounelis, P., Brumley, L.M., Dey, N., Browne, B., Auger, D., Fritz, T.A., Kulesza, P., and Bedwell, D.M. (1993) Phosphoglucomutase in *Saccharomyces cerevisiae* is a cytoplasmic glycoprotein and the acceptor for a Glc-phosphotransferase. *J. Biol. Chem.*, **268**, 8341–8349.

Margolis, R.K. and Margolis, R.U. (1993) Nervous tissue proteoglycans. *Experientia*, **49**, 429–446.

Marks, D.L., Wu, K., Paul, P., Kamisaka, Y., Watanabe, R., and Pagano, R.E. (1999) Oligomerization and topology of the Golgi membrane protein glucosylceramide synthase. *J. Biol. Chem.*, **274**, 451–456.

Medina, L. and Haltiwanger, R.S. (1998a) SV40 large T antigen is modified with *O*-linked N-acetylglucosamine but not with other forms of glycosylation. *Glycobiology*, **8**, 383–391.

Medina, L. and Haltiwanger, R.S. (1998b) Calf thymus high mobility group proteins are nonenzymatically glycated but not significantly glycosylated. *Glycobiology*, **8**, 191–198.

Oriol, R., Mollicone, R., Cailleau, A., Balanzino, L., and Breton, C. (1999) Divergent evolution of fucosyltransferase genes from vertebrates, invertebrates, and bacteria. *Glycobiology*, **9**, 323–334.

Park, J.W., Voss, P.G., Grabski, S., Wang, J.L., and Patterson, R.J. (2001) Association of galectin-1 and galectin-3 with Gemin4 in complexes containing the SMN protein. *Nucleic Acids Res.*, **29**, 3595–3602.

Pederson, L.C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T.A., and Negishi, M. (2000) Heparan/chondroitin sulfate biosynthesis. Structure and mechanism of human glucuronyltransferase I. *J. Biol. Chem.*, **275**, 34580–34585.

Persson, K., Ly, H.D., Dieckelmann, M., Wakarchuk, W.W., Withers, S.G., and Strynadka, N.C. (2001) Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nat. Struct. Biol.*, **8**, 166–175.

Que, Q., Li, Y., Wang, I.-N., Lane, L.C., Chaney, W.G., and van Etten, J.L. (1994) Protein glycosylation and myristylation in *Chlorella* virus PBCV-1 and its antigenic variants. *Virology*, **203**, 320–327.

Rousseau, C., Muriel, M.P., Musset, M., Botti J., and Seve, A.-P. (2000) Glycosylated nuclear lectin CBP70 also associated with endoplasmic reticulum and the Golgi apparatus: does the "classic pathway" of glycosylation also apply to nuclear glycoproteins? *J. Cell. Biochem.*, **78**, 638–649.

Sassi, S., Sweetinburgh, M., Erogul, J., Zhang, P., Teng-umnuay, P., and West, C.M. (2001) Analysis of Skp1 glycosylation and nuclear enrichment in *Dictyostelium*. *Glycobiology*, **11**, 283–295.

Satir, B.H., Srisomsap, C., Reichman, M., and Marchase, R.B. (1990) Parafusin, an exocytic-sensitive phosphoprotein, is the primary acceptor for the glucosylphosphotransferase in *Paramecium tetraurelia* and rat liver. *J. Cell Biol.*, **111**, 901–907.

Sato, Y., Naito, Y., Grundke-Iqbal, I., Iqbal, K., and Endo, T. (2001) Analysis of N-glycans of pathological tau: possible occurrence of aberrant processing of tau in Alzheimer's disease. *FEBS Lett.*, **496**, 152–160.

Saxena, I.M. and Brown, R.M. (2000) Cellulose synthases and related enzymes. *Curr. Opin. Plant Biol.*, **3**, 523–531.

Schenk, B. Fernandez, F., and Waechter, C.J. (2001) The ins(ide) and outs(ide) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiology*, **11**, 61R–70R.

Schulman, B.A., Carrano, A.C., Jeffrey, P.D., Bowen, Z., Kinnucan, E.R.E., Finnin, M.S., Elledge, S.J., Harper, J.W., Pagano, M., and Pavletich, N.P. (2000) Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature*, **408**, 381–386.

Senay, C., Lind, T., Muguruma, K., Tone, Y., Kitigawa, H., Sugahara, K., Lidholt, K., Lindahl, U., and Kusche-Gullberg, M. (2000) The EXT1/EXT2 tumor suppressors: catalytic activities and role in heparan sulfate biosynthesis. *EMBO Rep.*, **1**, 282–286.

Seol, J.H., Shevchenko, A., Shevchenko, A., and Deshaies, R.J. (2001) Skp1 forms multiple protein complexes, including RAVE, a regulator of V-ATPase assembly. *Nat. Cell Biol.*, **3**, 384–391.

Shimura, H., Schlossmacher, M.G., Hattori, N., Frosch, M.P., Trockenbacher, A., Schneider, R., Mizuno, Y., Kosik, K.S., and Selkoe, D.J. (2001) Ubiquitination of a new form of α-synuclein by parkin from human brain: implications for Parkinson's disease. *Science*, **293**, 263–269.

Spiro, R.G. (1972) Study of the carbohydrates of glycoproteins. *Methods Enzymol.*, **28**, 3–43.

Sucgang, R., Shaulsky, G., and Kuspa, A. (2000) Toward the functional analysis of the *Dictyostelium discoideum* genome. *J. Eukaryot. Microbiol.*, **47**, 334–339.

Swofford, D.L. (2000) PAUP*. Phylogenetic Anlysis Using Parsimony (* and Other Methods), version 4. Sinauer Assocaites, Sunderland, MA.

Teng-umnuay, P., Morris, H.R., Dell, A., Panico, M., Paxton, T., and West, C.M. (1998) The cytoplasmic F-box binding protein Skp1 contains a novel pentasaccharide linked to hydroxyproline in *Dictyostelium*. *J. Biol. Chem.*, **273**, 18242–18249.

Teng-umnuay, P., van der Wel, H., and West, C.M. (1999) Identification of a UDP-GlcNAc:Skp1-hydroxyproline GlcNAc-transferase in the cytoplasm of *Dictyostelium*. *J. Biol. Chem.*, **274**, 36392–36402.

Thornton, T.M., Swain, S.M., and Olszewski, N.E. (1999) Giberellin signal transduction presents…the SPY who *O*-GlcNAc'd me. *Tr. Plant Sci.*, **4**, 424–428.

Tiede, A., Bastisch, I., Schubert, J., Orlean, P., and Schmidt, R.E. (1999) Biosynthesis of glycosylphosphatidylinositols in mammals and unicellular microbes. *J. Biol. Chem.*, **380**, 503–523.

Trinchera, M. and Bozzarro, S. (1996) *Dictyostelium* cytosolic fucosyltransferase synthesizes H type 1 trisaccharide *in vitro*. *FEBS Lett.*, **395**, 68–72.

Unligil, U.M. and Rini, J.M. (2000) Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.*, **10**, 510–517.

Van der Wel, H., Morris, H.R., Panico, M., Paxton, T., Dell, A., Thomson, J.M., and West, C.M. (2001a) A non-Golgi α1, 2-fucosyl-transferase that modifies Skp1 in the cytoplasm of *Dictyostelium*. *J. Biol. Chem.*, **276**, 33952–33963.

Van der Wel, H., Fisher, S.Z., Gaucher, E.A., and West, C.M. (2001b) A bifunctional diglycosyltransferase forms the Fucα1, 2Galβ, 3-disaccharide on Skp1 in the cytoplasm of *Dictyostelium*. *Glycobiology*, **11**, abstract 62.

Van Etten, J.L., Lane, L.C., and Meints, R.H. (1991) Viruses and viruslike particles of eukaryotic algae. *Microbiol. Rev.*, **55**, 586–620.

Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G., and Marth, J., eds. (1999) *Essentials of glycobiology*. Cold Spring Harbor Laboratory Press, Cold SPring Harbor, NY, pp. 171–182, 289–292.

Veyna, N.A., Jay, J.C., Srisomsap, C., Bounelis, P., and Marchase, R.B. (1994) The addition of glucose-1-phosphate to the cytoplasmic glycoprotein phosphoglucomutase is modulated by intracellular calcium in PC12 cells and rat cortical synaptosomes. *J. Neurochem.*, **62**, 456–464.

Vyakarnam, A., Lenneman, A.J., Lakkides, K.M., Patterson, R.J., and Wang, J.L. (1998) A comparative nuclear localization study of galectin-1 with other splicing components. *Exp. Cell Res.*, **242**, 419–428.

Wang, I.-N., Li, Y., Que, Q., Bhattacharya, M., Lane, L.C., Chaney, W.G., and Van Etten, J.L. (1993) Evidence for virus-encoded glycosylation specificity. *Proc. Natl Acad. Sci. USA*, **90**, 3840–3844.

Warnecke, D., Erdmann, R., Fahl, A., Hube, B., Muller, F., Zank, T., Zahringer, U., and Heinz, E. (1999) Cloning and functional expression of UGT genes encoding sterol glucosyltransferases from *Saccharomyces cerevisiae*, *Candida albicans*, *Pichia pastoris*, and *Dictyostelium discoideum*. *J. Biol. Chem.*, **274**, 13048–13059.

Wells, L., Vesseller, K., and Hart, G.W. (2001) Glycosylation of nucleocytoplasmic proteins: signal transduction and *O*-GlcNAc. *Science*, **291**, 2376–2378.

West, C.M., Scott-Ward, T., Teng-umnuay, P., Van der Wel, H., Kozarov, E., and Huynh, A. (1996) Purification and characterization of an α1, 2-L-fucosyltransferase, which modifies the cytosolic protein FP21, from the cytosol of *Dictyostelium*. *J. Biol. Chem.*, **271**, 12024–12035.

West, C.M., Kozarov, E., and Teng-umnuay, P. (1997) The cytosolic glycoprotein FP21 of *Dictyostelium discoideum* is encoded by two genes resulting in a polymorphism at a single amino acid position. *Gene*, **200**, 1–10.

West, C.M., Van der Wel, H., Kaplan, L., Morris, H.R., and Dell, A. (2001) Analysis of a microbial UDP-GlcNAc:hydroxyproline polypeptide GlcNAcTransferase that modifies Skp1 in the cytoplasm of *Dictyostelium*. *Glycobiology*, **11**, abstract 61.

Whitfield, C. and Roberts, I.S. (1999) *Structure*, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol. Microbiol.*, **31**, 1307–1319.

Yoshida, M., Itano, N., Yamada, Y., and Kimata, K. (2000) *In vitro* synthesis of hyaluronan by a single protein derived from mouse HAS1 gene and characterization of amino acid residues essential for the activity. *J. Biol. Chem.*, **275**, 497–506.

Zachara, N.E., Packer, N.H., Temple, M.D., Slade, M.B., Jardine, D.R., Karuso, P., Moss, C.J., Mabbutt, B.C., Curmi, P.M.G., Williams, K.L., and Gooley, A.A. (1996) Recombinant prespore-specific antigen from *Dictyostelium discoideum* is a β-sheet glycoprotein with a spacer peptide modified by *O*-linked N-acetylglucosamine. *Eur. J. Biochem.*, **238**, 511–518.

Zhang, Y., Zhang, P., and West, C.M. (1999) A linking function for the cellulose-binding protein SP85 in the spore coat of *Dictyostelium discoideum*. *J. Cell Sci.*, **112**, 4667–4677.