

ROLES OF N-LINKED GLYCANS IN THE ENDOPLASMIC RETICULUM

Ari Helenius¹ and Markus Aebi²

*Institute of Biochemistry¹ and Institute of Microbiology,² Swiss Federal Institute of
Technology Zurich, Zurich 8093, Switzerland; e-mail: ari.helenius@bc.biol.ethz.ch,
markus.aebi@micro.biol.ethz.ch*

Key Words glycoprotein folding, quality control, ER-associated degradation,
calnexin/calreticulin, glycan biosynthesis

■ **Abstract** From a process involved in cell wall synthesis in archaea and some bacteria, N-linked glycosylation has evolved into the most common covalent protein modification in eukaryotic cells. The sugars are added to nascent proteins as a core oligosaccharide unit, which is then extensively modified by removal and addition of sugar residues in the endoplasmic reticulum (ER) and the Golgi complex. It has become evident that the modifications that take place in the ER reflect a spectrum of functions related to glycoprotein folding, quality control, sorting, degradation, and secretion. The glycans not only promote folding directly by stabilizing polypeptide structures but also indirectly by serving as recognition “tags” that allow glycoproteins to interact with a variety of lectins, glycosidases, and glycosyltransferases. Some of these (such as glucosidases I and II, calnexin, and calreticulin) have a central role in folding and retention, while others (such as α -mannosidases and EDEM) target unsalvageable glycoproteins for ER-associated degradation. Each residue in the core oligosaccharide and each step in the modification program have significance for the fate of newly synthesized glycoproteins.

CONTENTS

INTRODUCTION	1020
N-LINKED GLYCANS AS SEMI-INDEPENDENT APPENDICES	1021
BI-COMPARTMENTAL BIOSYNTHESIS OF CORE OLIGOSACCHARIDE	1022
Oligosaccharyltransferase	1025
GLYCOPROTEIN FOLDING	1027
Direct Effects on Folding	1027
The Calnexin/Calreticulin Cycle	1028
Calnexin and Calreticulin	1029
ERp57	1034
Glucosidase I	1034
Glucosidase II	1035
UDP-Glucose:Glycoprotein Glucosyltransferase	1036
UDPase	1037

ER-DEPENDENT DEGRADATION OF GLYCOPROTEINS	1038
OTHER LECTINS IN THE SECRETORY PATHWAY	1039
Lectins Related to Mannose 6-Phosphate Receptors	1039
Leguminous Lectin Homologues	1040
N-GLYCANS: THE EVOLUTION OF A SIGNAL	1041
THE N-LINKED GLYCAN, A COMPOSITE, MULTIFUNCTIONAL SIGNAL .	1043
PERSPECTIVES	1044

INTRODUCTION

The majority of proteins synthesized in the endoplasmic reticulum (ER) are glycoproteins. The N-linked oligosaccharide moieties of these proteins serve highly diverse functions. They are ligands in a multitude of recognition processes: They stabilize the proteins against denaturation and proteolysis, enhance solubility, modulate immune responses, facilitate orientation of proteins relative to a membrane, confer structural rigidity to proteins, regulate protein turnover, fine-tune the charge and isoelectric point of proteins, and mediate interactions with pathogens. No other covalent protein modification is as common and as complex chemically, and no other modification is employed for so many different purposes.

The biosynthesis of glycoproteins is a task shared by the ER and the Golgi apparatus. The division of labor is such that together with the cytosol the ER is responsible for the synthesis of the polypeptide and the core oligosaccharides, for the covalent coupling of glycan and polypeptide, and for initial modification of the glycans. Once the glycoproteins have folded and oligomerized properly, they move to the Golgi complex where the N-linked glycans are subjected to further trimming and modification. New saccharides are often added to generate the complex glycans found in the mature glycoproteins.

For a long time, it was unclear why cells have evolved such a complicated and apparently wasteful biosynthetic strategy. Why synthesize a large oligosaccharide in the ER, and then—after transferring it to a polypeptide—subject it to trimming in order to build it up again with different sugars? Why would this process be so important as to be conserved and virtually unchanged in all eukaryotes? The apparent lack of acceptable logic has not escaped the thousands of biochemistry and cell biology students who have had to memorize each step in the pathway.

The logic turns out not to be so complicated: The different configurations of the N-linked glycan are not merely intermediates in a biosynthesis pathway but have specific functions of their own. In fact, they play a role in at least three different stages during the existence of a glycoprotein. For each of them the N-linked glycans have to look different.

The first phase occurs in the ER where partially trimmed versions of the core oligosaccharide are needed for proper protein folding and quality control. Here, the glycans help to secure the fidelity of protein production. This is the phase

described in this review. The second phase involves a role in intracellular transport and targeting exemplified by the role of mannose 6-phosphate in targeting of lysosomal hydrolases. This phase occurs in the ER, in the Golgi complex, and in the *trans*-Golgi network. The third phase takes place after extensive modification in the Golgi. It occurs when the mature protein has reached the extracellular space, the lysosome, the plasma membrane, or wherever the protein is targeted. The functions of the glycans in the mature proteins are as varied as the structures themselves.

In this review, we focus on the role of N-linked glycans in the ER. We describe the synthetic events and the trimming enzymes. We also describe the role of glycans in folding and in the regulation of glycoprotein degradation by ER-associated degradation (ERAD). These events involve a network of lectins, folding sensors, glycosyltransferases, and glycosidases. Finally we discuss the evolutionary origin of this machinery. For a general background on the role of N-linked glycans and their functions, several reviews can be recommended (1, 2). We particularly recommend a recent review by Trombetta (3), which covers many of the same topics that we do with comprehensive references to the original literature.

N-LINKED GLYCANS AS SEMI-INDEPENDENT APPENDICES

After analyzing the SWISS-PROT database, Apweiler et al. (4) recently predicted that more than half of all eukaryotic protein species are glycoproteins. About 90% of these are likely to carry N-linked glycans, and there is an average of 1.9 N-linked glycans per polypeptide chain. With a molecular weight up to 3 kDa, the oligosaccharide groups in mammalian glycoproteins frequently make up a sizable portion of the mass of a glycoprotein and cover a large fraction of its surface.

The glycans are exposed on the surface. They form flexible, hydrated branches that can extend 3 nm or further into the solvent. X-ray crystal structures of glycoproteins and NMR studies indicate that, aside from interactions with the first two N-acetylglucosamines (GlcNAcs), the majority of glycans have few contacts with the surface of the protein (5).

Recently, Petrescu and colleagues (6) surveyed 506 glycoproteins listed in the Protein Data Bank crystallographic database and found the glycans tend to be located in positions where the secondary structure changes. They also observed that some glycans (10%) occur in invaginations of the protein surface, and others are bound to asparagines at the edge of indentations in the protein surface partially filled by the glycans (20%). Even in cases where there are interactions between the oligosaccharide and the protein surface, the terminal sugars are generally free. The N-linked glycans and the protein moiety are thus as a rule relatively independent of each other. Glycans can often be enzymatically removed from the

surface of folded glycoproteins without appreciable effects on protein structure and function. X-ray crystallographers take advantage of this to make glycoproteins more amenable to crystallization.

That glycans behave like semi-independent appendices has several consequences. First, they can be modified without appreciable effects on the protein. Every N-linked glycan is, in fact, subject to extensive modification. This allows cells to fine-tune the biophysical and biological properties of glycoproteins and to generate the microheterogeneity so characteristic of glycoproteins (7). Importantly, the semi-independent nature of glycans also allows cell types and cells in different stages of differentiation and transformation to imprint on their glycoprotein pool their own specific biochemical characteristics, and thus give their exposed surface a “corporate identity.” This makes cells recognizable to other cells in a multicellular environment. It allows self-recognition and provides a central theme in development, differentiation, physiology, and disease (8).

Both extra- and intracellularly, N-linked glycans are used as specific tags or signals recognized by a spectrum of carbohydrate-binding proteins (lectins) (9, 10). Glycans are ideally suited for signal functions. They are prominently exposed on the surface of glycoproteins, they are polar, and in each protein they can be present in multiple copies. During evolution, single point mutations can generate or eliminate N-linked glycosylation sites thus providing genetic versatility. Because the chemistry of sugars allows multiple types of linkages, cells can package and expose a large amount of information in a small space. The loss or addition of a single residue, or the alteration in a single bond, can, for example, dramatically alter binding to lectins (9).

As illustrated below, the notion of glycans as modifiable tags for lectin binding in diverse recognition processes is a central theme in glycobiology. That the affinity between carbohydrates and proteins is generally rather low (dissociation constants are typically in the micromolar range or higher) does not seem to be a major disadvantage. In the confined and crowded space of the ER, lectins are abundant and interactions are meant to be transient. In addition, most lectins are multivalent, which dramatically enhances their avidity for ligands.

BI-COMPARTMENTAL BIOSYNTHESIS OF CORE OLIGOSACCHARIDE

N-linked glycans are added to proteins en bloc in the lumen of the ER as presynthesized oligosaccharides. The core oligosaccharides have a clearly defined structure. In virtually all eukaryotes they are composed of a branched oligosaccharide unit made of three glucoses, nine mannoses, and two N-acetylglucosamines ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$) (Figure 1).

The core glycan is the product of a biosynthesis pathway in which monosaccharides are added to a lipid carrier (dolichol-pyrophosphate) by monosaccharyltransferases in the ER membrane (1, 11). Synthesis occurs on both sides of the

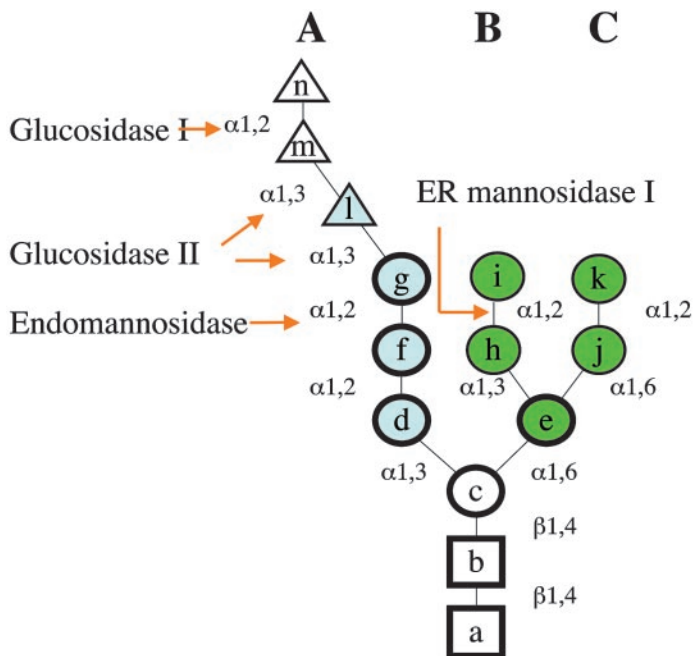
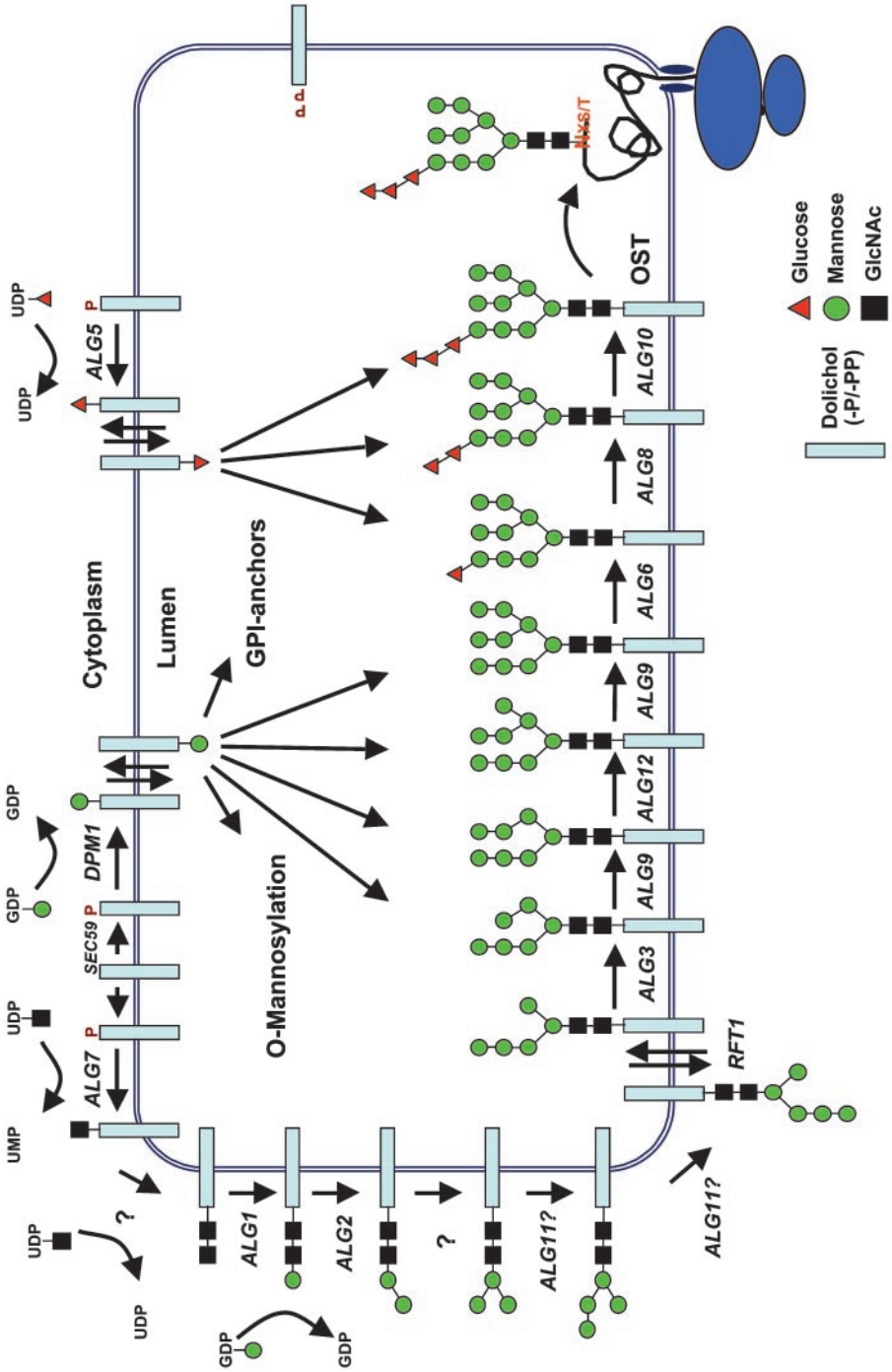


Figure 1 The N-linked core oligosaccharide. The core glycan has 14 saccharides: 3 glucoses (*triangles*), 9 mannoses (*circles*), and 2 N-acetylglucosamines. Each saccharide has a letter assigned to it, and these are used in the text to identify the saccharide (3). There are three branches named A, B, and C. The cleavage sites of some glycosidases involved in trimming are indicated. Residues a-g (shown by bold symbols) are added to the glycan on the cytosolic surface of the ER-membrane during biosynthesis; the rest are added lumenally. The blue residues (d, f, g, and l) are involved in the interaction of monoglucosylated glycans with calnexin and calreticulin. The green residues, with the exception of i, are likely to interact with EDEM.

ER membrane (Figure 2). Seven sugars are added on the cytosolic surface after which the sugar moiety is translocated, “flipped,” to the luminal side. Flipping is catalyzed by an ATP-independent, bi-directional flippase (12). In yeast, there is evidence that the flippase is the RFT1 protein, a polytopic membrane protein with about 10 *trans*-membrane spans (13). Unlike flippases in the plasma membrane that translocate phospholipids (14), RFT1 does not have ATP-binding cassettes and hence does not belong to the ABC family of translocators. Genes for homologous proteins occur in the genomes of other eukaryotes (15).

Whether facing the lumen or the cytosol, each individual glycosyltransferase displays strong preference toward a single oligosaccharide substrate (16). This leads to a linear, stepwise biosynthetic pathway of the branched oligosaccharide. The final step is the addition of a terminal α -1,2 linked glucose residue (residue



n) (Figure 1). This residue is needed for efficient recognition by the oligosaccharyltransferase, the enzyme that transfers the finished oligosaccharide from the lipid-bound precursor to the polypeptide (17, 18).

Oligosaccharyltransferase

The oligosaccharyltransferase (OST) is associated with the translocon complex. Together with the signal peptidase, BiP, calnexin, and possibly other factors, it is a member of the welcoming committee that every nascent polypeptide meets as it exits the translocon complex and enters the ER lumen. OST scans the emerging polypeptide for glycosylation sequons (Asn-X-Ser/Thr) and adds N-linked oligosaccharides to the side chain nitrogen of the Asn residue by an N-glycosidic bond. Because the oligosaccharides are added when the sequon is only 12 to 14 residues into the ER lumen, the active site of the enzyme can be no further than 5 nm away from the mouth of the translocon (19). As it passes the OST with an average rate of about 5 residues per second, the polypeptide is still unfolded (20).

The transfer of a glycan to the chemically rather inert side chain of the Asn requires formation of a loop in the polypeptide so that the hydroxyl groups of Ser or Thr can contact the Asn amide and render it more nucleophilic (21, 22). This explains why the middle residue X in the sequon cannot be a proline; proline prevents the formation of such a loop. It also explains why folded polypeptides are poor substrates.

OST is composed of multiple *trans*-membrane subunits. In *Saccharomyces cerevisiae*, there are eight, Ost1p, Ost2p, Wbp1, Swp1, Stt3p, Ost3p/Ost6p, Ost4p, and Ost5p, of which the five first are essential (23). Ost3p and Ost6p are homologues that define distinct oligosaccharyltransferase isoforms (23–25).

←

Figure 2 Synthesis of the N-linked core oligosaccharide and its transfer to a polypeptide chain. Biosynthesis occurs on both sides of the ER membrane. The yeast loci required for the individual biosynthetic steps are indicated. Synthesis starts on the cytoplasmic side where GlcNAc-1-phosphate is transferred from UDP-GlcNAc to dolichylpyrophosphate, followed by an additional GlcNAc and five mannose residues. The Man₅ GlcNAc₂ oligosaccharide, thus generated, is translocated (flipped) into the lumen of the ER (13). On the luminal side, the lipid-linked Man₅ GlcNAc₂ is extended by the addition of four mannose and three glucose residues. The enzymes involved differ from most other glycosyltransferases in spanning the membrane several times and in being quite hydrophobic (16, 17, 34, 179–181). Unlike the cytosol-oriented glycosyltransferases that initiate core oligosaccharide assembly and Golgi-localized glycosyltransferases, these use lipids (dolichol-P-Man and dolichol-P-Glc) as saccharide donors. Little is known about the catalytic mechanism of these interesting enzymes. However, based on sequence similarity, a common origin has been suggested (182).

Except for the small Ost5p and Ost4p, all subunits present in yeast OST have homologues in the mammalian enzyme (26, 27). The homologues of the five essential subunits in yeast (ribophorin I, DAD1, OST48, ribophorin II, and STT3-A/STT3-B) are thought to form a central core complex to which N33 and IAP are associated peripherally (28). There are two isoforms of OST in mammalian cells that differ with respect to the STT3-A/STT3-B subunit. Expression of these alternative subunits is tissue- and cell-type specific (28).

As outlined above, the terminal α -1,2 linked glucose (n) (Figure 1) in the lipid-linked oligosaccharide is a central element in substrate recognition by OST. In vivo analysis of glycosylation efficiency and oligosaccharide specificity in yeast show that when the complete $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ oligosaccharide is limiting, priority is given to the transfer of the correct oligosaccharide structure even with the risk of underglycosylation. However, if the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ oligosaccharide is not available at all, incompletely assembled oligosaccharides are transferred. Gilmore and coworkers (29) developed an allosteric model for yeast OST to explain this interesting ambiguity in substrate specificity.

In higher eukaryotes, in contrast to yeast, the choice between faithful addition of the complete glycan structure and occupancy of glycosylation sites is subject to regulation (28). Mammalian cells contain two OST complexes that differ in selectivity toward the lipid-linked donor substrate. Complexes containing the STT3-A subunit preferentially use the complete oligosaccharide, although those with STT3-B have a higher v_{max} but can also use incomplete core glycans. The variable expression of the two complexes observed in different cell types may reflect differences in the balance between maximal glycosylation versus faithful use of the complete oligosaccharide.

The complexity of OST has made it difficult to assign specific functions to subunits. Cross-linking experiments suggest that STT3 (30, 31), ribophorinI/Ost1 (32), and the Ost48 subunit make direct contact with the substrate, and biochemical data supports an essential role for Wbp1p (Ost48p) (33). However, the distinct influence of different STT3 subunits on the catalytic properties of the mammalian OST (28) and the recent finding that a bacterial homologue of the STT3 subunit, the PglB protein in *Campylobacter jejuni*, is sufficient for OST activity imply that the STT3 subunit is the catalytic subunit (34, 35).

The efficiency with which OST transfers a core glycan to individual sequons varies. Analysis of well-characterized glycoproteins in the SWISS-PROT database and of glycoproteins in the PDB crystallographic database indicates that sequon occupancy by oligosaccharides is about 2/3 (4, 6). Some sequons are ignored, some are glycosylated with partial efficiency, and some are glycosylated with full efficiency. Numerous factors affect efficiency, including the subunit composition of the OST complex, the amino acids in the sequon itself and immediately adjacent to it, the location of the sequon in the polypeptide chain, the availability of dolichol precursor in the ER, and the rate of protein folding.

GLYCOPROTEIN FOLDING

Protein folding in the ER begins cotranslocationally during the entry of a polypeptide chain through the translocon complex either into the lumen of the ER or, in the case of polytopic membrane proteins, into the ER membrane (36–38). It continues after the polypeptide chain has dissociated from the ribosome and the translocon complex. In contrast to proteins in the cytosol, most proteins that fold in the ER acquire disulfide bonds through an oxidation process catalyzed by thiol-disulfide oxidoreductases (39). Finally, because many proteins are oligomeric in their native form, assembly into homo- or hetero-oligomers often completes the folding process.

That many glycoproteins need their N-linked glycans for efficient secretion was realized when tunicamycin, an inhibitor of N-linked glycosylation, became available in the 1970s [see (1, 40)]. It is now clear that when glycosylation is blocked many polypeptides undergo improper or incomplete folding. Failing to reach the native conformation, they do not pass ER quality control (41). They are retained in the ER and eventually are degraded (42).

It is important to point out that all glycoproteins are not equally dependent on their glycans for folding and secretion. If devoid of glycans, many suffer only partial loss of folding and secretion efficiency, others become temperature sensitive, and many remain unaffected [reviewed in (40, 43)]. The importance of glycosylation is thus highly variable. As a rule, the folding of those glycoproteins that have a large number of glycans is more glycan dependent.

When individual sequons in glycoproteins are mutated, it was observed that only some are essential. In the hemagglutinin (HA) of influenza virus (Aichi strain), for example, only one of six glycans (N81) is absolutely essential when glycans are removed one by one (44). Without this glycan, HA fails to acquire any of the six native intrachain disulfide bonds, and it does not exit the ER. The most likely explanation is a local perturbation that prevents oxidation of a nearby disulfide bond (C67-C76). This bond is needed for the onset of further oxidative folding of the molecule. Removing the other five glycans in HA individually does not affect folding, but removing several of them together can lead to misfolding (38, 44). This example illustrates a phenomenon often observed: Although a single glycan may not be essential, the same glycan may prove important when more than one is eliminated. Again, great variability is seen among individual glycoproteins. It is evident that glycans have local effects where their precise location is important, and global effects where their presence is important—but their precise location is not.

Direct Effects on Folding

Part of the fold-promoting effect of oligosaccharides is biophysical. The addition of large, polar carbohydrates affects the properties of a polypeptide chain directly. Although systematic biophysical studies are lacking, it is known that the presence of a glycan can profoundly influence the conformational profile of short

glycopeptides (45). Often, it rigidifies their conformation by limiting the conformational space accessible to the polypeptide chain. An interaction between the N-acetyl group of the first GlcNAcs and the polypeptide, moreover, promotes the formation of β -turns (46). It is likely that some glycans have a role in promoting and stabilizing local structure (6). Other effects on folding are likely to be more global, such as increased solubility of folding intermediates.

Comparison of native glycoproteins to nonglycosylated versions of the same shows that the presence of glycans increases stability, solubility, and resistance to proteases (45, 47–49). The stabilization effect is mainly entropic. For example, the two N-linked glycans stabilize the first domain of ovomucoid against thermal denaturation by about 2.5 kcal/mol (50). It is hypothesized that glycans stabilize the folded conformation by decreasing the freedom of mobility of unfolded conformations.

The Calnexin/Calreticulin Cycle

Glycans have also an indirect role in protein folding. It is based on binding of the newly synthesized glycoproteins to lectins in the ER. In this process, the glycans serve as sorting signals that the cell modifies to reflect the folding status of the protein. That such a tagging paradigm applies to glycoprotein folding and quality control emerged from four key observations in the early 1990s: (a) N-linked glycans in misfolded glycoproteins retain a glucose residue (residue 1 in Figure 1) in the A branch, and they undergo a cycle of re- and deglycosylation (51); (b) the glucosyltransferase in the ER responsible for the reglycosylation is selective for misfolded glycoproteins (52); (c) newly synthesized glycoproteins bind transiently and selectively to calnexin, a resident ER protein (53); and (d) the binding to calnexin is blocked by inhibitors of ER glucosidases (54).

The existence of a cycle (Figure 3) was first postulated in 1993 for calnexin (54, 55), and later the cycle was expanded to include the soluble, luminal ER protein calreticulin (56, 57). Although controversial in the beginning, the calnexin/calreticulin cycle is now generally accepted. Each step has been extensively analyzed and confirmed in many systems. The literature has been frequently reviewed (3, 58–62).

Before discussing the individual proteins involved in the calnexin/calreticulin cycle in detail, it may be useful to go through the cycle step-by-step (Figure 3). The process begins when a core glycan has been added to the growing, nascent polypeptide chain. The first glucose is rapidly removed by glucosidase I and is followed by removal of the second glucose by glucosidase II. The monoglucosylated core ligand, thus generated, binds the glycoprotein to calnexin or calreticulin. These sequester the nascent or newly synthesized glycopolypeptide chains and serve as molecular chaperones, preventing aggregation and export of the incompletely folded chains from the ER. In many cases, they also protect the folding intermediates against premature degradation and, at the same time, expose them to ERp57, a thiol-disulfide oxidoreductase. It is a cofactor that

provides assistance for proper disulfide bond formation during the ongoing folding process.

To release the bound chains from calnexin and calreticulin, glucosidase II removes the remaining glucose residue. The glycoprotein no longer binds to the lectins and is now free to leave the ER unless recognized by a soluble enzyme, UDP-Glc:glycoprotein glucosyltransferase (GT). GT only reglucosylates incompletely folded glycoproteins, and it serves as a folding sensor in the cycle. If reglucosylated by GT, a glycoprotein rebinds to the lectins. A glycoprotein stays in the cycle until it is either properly folded and oligomerized, or degraded. Virtually all newly synthesized glycoproteins seem to undergo a phase in which they associate transiently with calnexin, calreticulin, or both. Calnexin association is observed for both membrane-bound and soluble glycoproteins. Calreticulin can also bind to both types of proteins, but it is more frequently associated with soluble substrates.

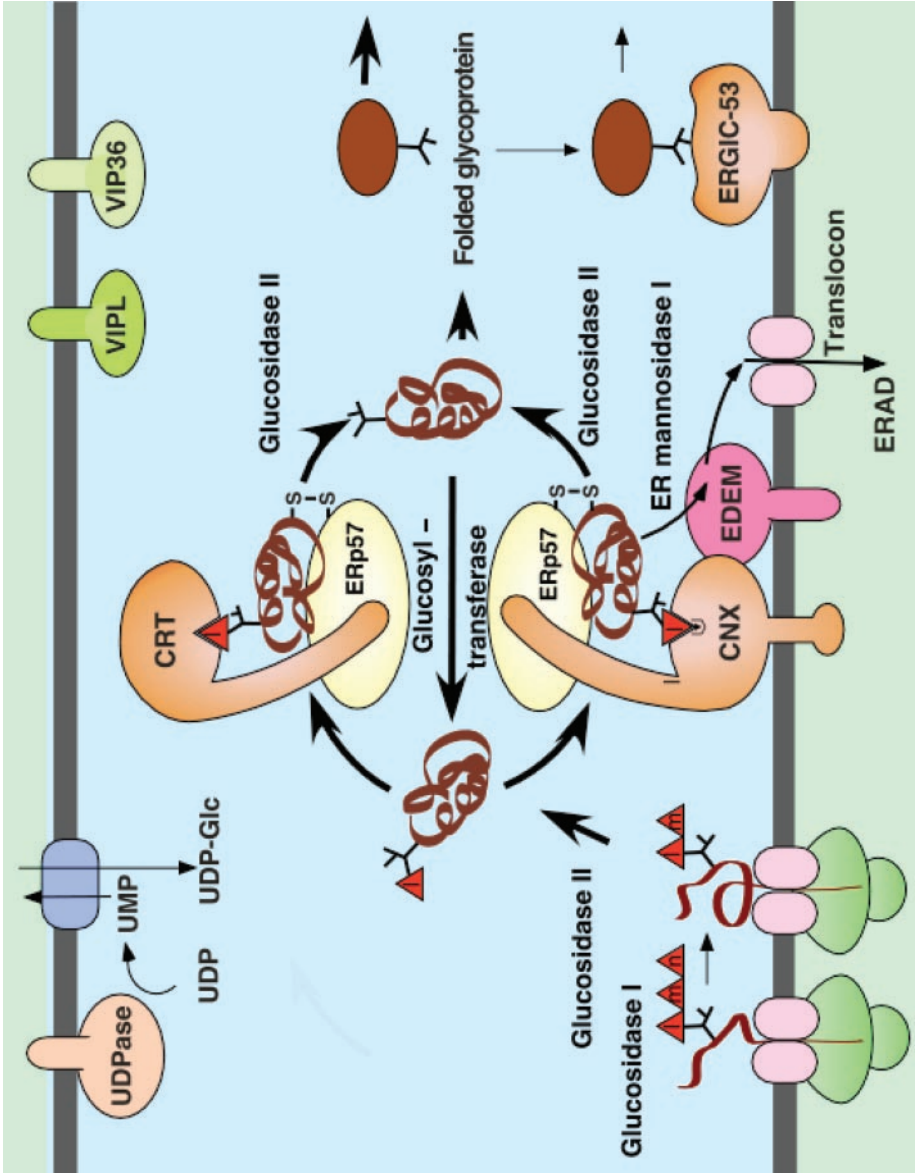
If association with the cycle is inhibited, for example, by blocking the action of the glucosidases, the folding rate of glycoproteins is often increased and folding efficiency decreased. In some cases, quality control breaks down with the consequence that misfolded glycoproteins exit the ER. In other cases, the glycoproteins associate with BiP, an important ER chaperone that cooperates and competes with the calnexin cycle for substrates and that serves as a back-up.

Calnexin and Calreticulin

Calnexin (a transmembrane protein of Type I) and calreticulin (a soluble luminal ER protein) are related members of the legume lectin family. Both are monomeric, calcium-binding proteins, and have ER localization signals. Transgenic mice devoid of calreticulin or calnexin have embryonic lethal phenotypes or are born with strong debilitating phenotypes (63, 64).

The NMR structure of the calreticulin P domain and the X-ray structure of calnexin's ectodomain show unusual domain architectures (60, 65). There are two separate domains, a globular β -sandwich domain with homology to legume lectins and a long extended hairpin fold corresponding to a proline-rich domain (the P domain) not seen in legume lectins (Figure 4). Calnexin has, in addition, a single transmembrane sequence and a cytoplasmic domain of 91 residues that can undergo phosphorylation and allows interactions with ribosomes (66).

The globular domain contains a concave and a convex β -sheet with the sugar binding site located on the concave surface partially shielded by the P domain (60) (Figure 4). The structure is stabilized by a calcium atom, which is not part of the lectin site. That calnexin and calreticulin (in contrast to ERGIC-53 and most other legume lectins) are monomers (67) is most likely explained by the presence of a short helix behind the convex β -sheet that may prevent intermolecular interactions between these surfaces. With its β -sandwich structure, the globular domain is not only similar to legume lectins, to ERGIC-53, and to galectins, but it is also similar to the neurexin family of neuronal cell surface receptors with hundreds of members (68–70).



Extending from a loop that connects two β -strands, the P domain curves away from the globular domain as a narrow hairpin fold stabilized by short antiparallel β -sheets and small hydrophobic clusters (60, 71). The length of the arm is 14 nm in calnexin and 11 nm in calreticulin. The structure suggests a model in which the glycosylated substrate protein, with a glycan moiety bound to the lectin site, occupies the space between the curved P domain arm and the globular domain. In this protected location, the substrate may be shielded from contacts with external factors and other incompletely folded proteins and thus prevented from aggregation. NMR studies suggest that the arm is somewhat flexible, which may allow the chaperone to adapt the space between the arm and the globular domain to the dimensions of a substrate (65).

Calnexin and calreticulin have virtually identical carbohydrate specificities (72–74). In addition to the single α -1,3 linked glucose (l), which is essential, at least three downstream mannoses in the A branch (d, f, and g) contribute to binding (Figure 4). IgG with a single monoglucosylated core glycan binds to calreticulin with a K_d of $\sim 2 \mu\text{M}$ (75, 76).

Modeling and mutational analysis show that the presence of the α -1,3 linked glucose is essential for calreticulin binding mainly because the equatorially oriented 2-hydroxyl forms hydrogen bonds with Asp317 and Tyr109 (75, 77). Similar interactions tie down the glucose in the calnexin lectin domain (60). The three mannoses in the A arm enhance the binding affinity by 25- to 50-fold by

←

Figure 3 The calnexin/calreticulin cycle. Immediately after addition of the core glycan to a growing polypeptide chain by OST, the outermost of the three glucose residues (n) is removed by glucosidase I. Soon thereafter, glucosidase II removes the middle glucose (m). Via the monoglucosylated core glycans thus generated, the glycoprotein binds to calnexin (CNX) and calreticulin (CRT). These sequester the nascent or newly synthesized chains and expose them to ERp57, a thiol-disulfide oxidoreductase that provides assistance during disulfide bond formation. When glucosidase II removes the remaining glucose (l), the glycoprotein dissociates from calnexin and calreticulin. The protein now encounters one of three possible fates. If properly folded, it is free to leave the ER. Exit may be assisted by mannose lectins, such as ERGIC-53, VIP36, and VIPL. If it is incompletely folded, UDP-Glc:glycoprotein glucosyltransferase uses UDP-glucose transported by a UDP-glucose/UMP exchanger from the cytosol to reglucosylate the high-mannose glycans located in improperly folded regions. Through these glycans, the glycoprotein rebinds to calnexin and calreticulin. The third fate is ER-associated degradation (ERAD) after retrotranslocation of the misfolded glycoprotein to the ER most likely through the translocon complex. ERAD of glycoproteins occurs when they have stayed in the ER lumen for some time and when they are recognized by a putative lectin (EDEM) because they have lost a mannose (i) through the action of ER mannosidase I. Red triangles are glucose residues. Abbreviations used are EDEM, ER degradation-enhancing α -mannosidase-like protein; VIP36, vesicular integral protein 36; VIPL, VIP36-like protein; ERAD, ER-associated protein degradation; ERGIC, ER-Golgi intermediate compartment; and ERp57, ER protein 57.

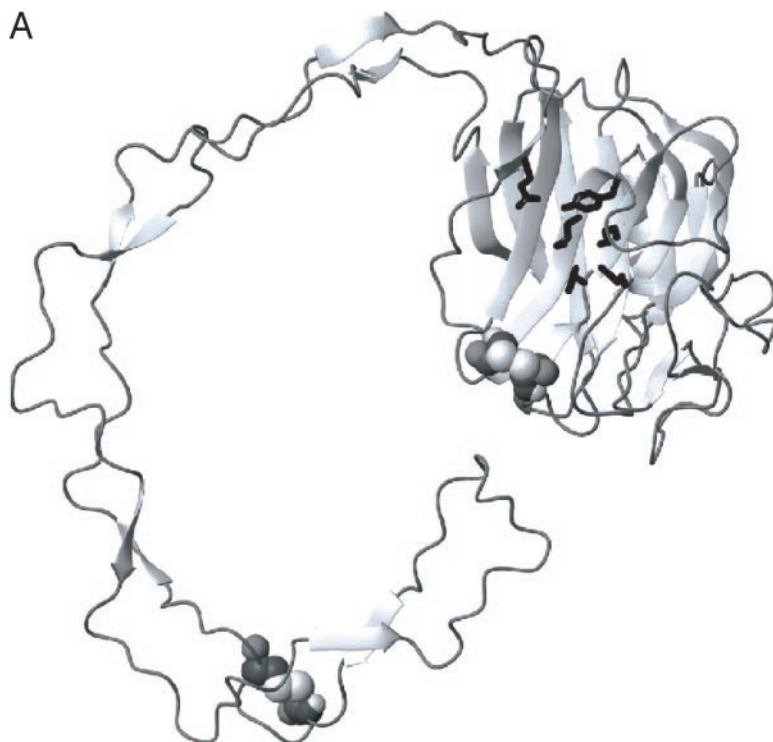


Figure 4 The structure of the calnexin ectodomain and a model of the calreticulin oligosaccharide-binding site. In calnexin (A) the oligosaccharide-binding site is situated in the globular domain composed mainly of a β -sandwich (60). The side chains of the residues interacting with the terminal glucose (I) are shown. The P domain is seen as a long antiparallel loop with four repeat units that have the same fold. The two intrachain disulfides are also shown. (B) A model of the oligosaccharide-binding site of calreticulin (77). The tetrasaccharide bound is $\text{Glc}\alpha_{1-3} \text{Man}\alpha_{1-2} \text{Man}\alpha_{1-2} \text{Man}$. Mutational analysis shows that several of the residues that interact with the sugar in the model are essential for binding (76).

occupying additional subsites for hexapyranosyl residues and thus increasing the number of hydrogen bonds and van der Waals interactions. By involving the entire A branch of the monoglucosylated oligosaccharide, the binding site is larger than generally observed for carbohydrate-protein interactions. Hence, the thermodynamics of the interaction shows unusual properties, such as enthalpy-entropy compensation (76).

Although the lectin activity is no longer questioned as the main principle of substrate binding to calnexin and calreticulin, the significance of protein-protein interactions remains an open issue. One main argument in favor of functionally

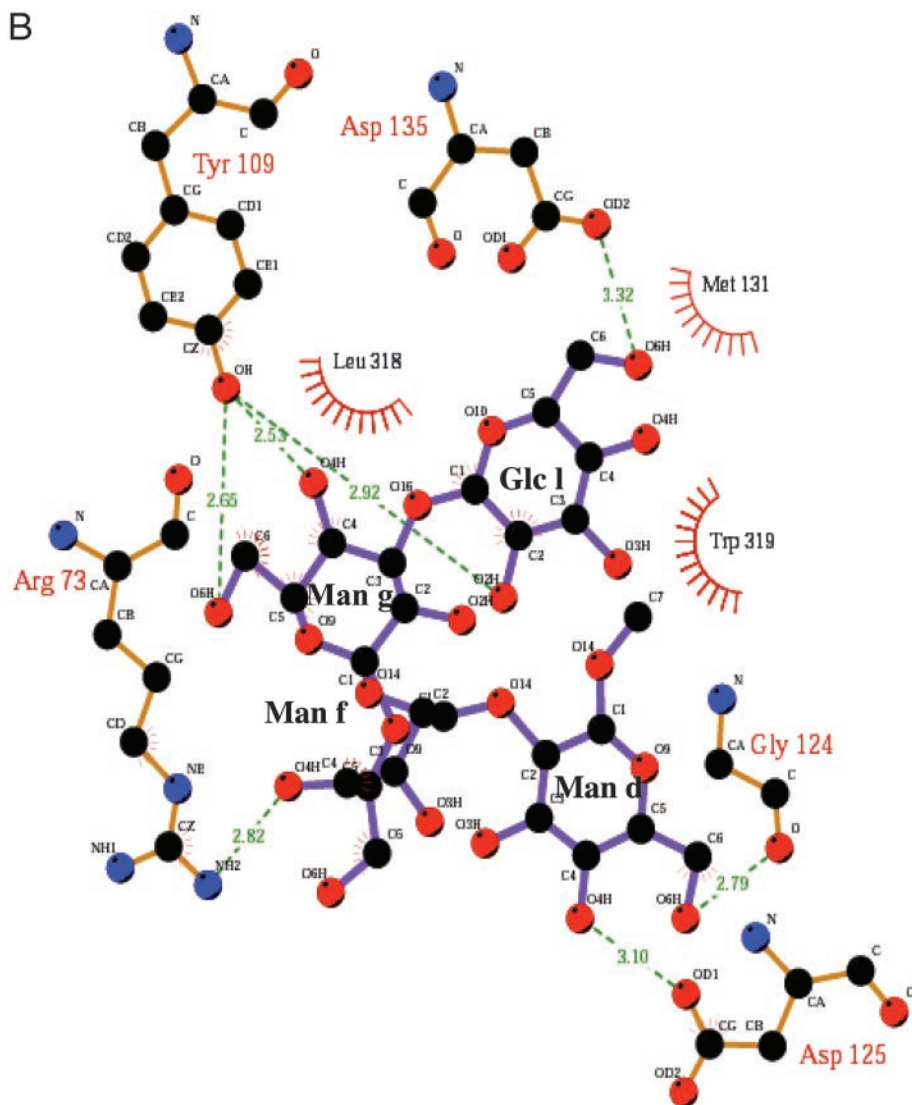


Figure 4 (Continued)

significant interactions between the protein moieties is that calnexin can be coimmunoprecipitated from cell lysates with some incompletely folded cargo proteins lacking glycans. Moreover, calreticulin and the soluble ectodomain of calnexin associate selectively with certain denatured nonglycosylated proteins *in vitro* and preserve them in a refolding competent state at elevated temperatures (78, 79). For further arguments, see Danilczyk & Williams (80).

The facts that weigh against a central role for protein-protein interactions are by no means more compelling. The vast majority of calnexin and calreticulin substrates fail to bind if glucose trimming is inhibited. Provided they have the monoglucosylated glycans, binding of model glycoproteins, such as RNase B and IgG, is independent of polypeptide conformation, suggesting that calnexin does not distinguish between conformers (75, 81, 82). In the case of IgG, the protein moiety does not contribute to the binding energy (76). Finally, the X-ray structure of the calnexin ectodomain does not reveal obvious binding sites for hydrophobic peptides or patches.

Other open issues include the significance of ATP binding and hydrolysis. The addition of ATP causes a change in the fluorescence emission and other properties of calnexin and calreticulin (78, 79, 83, 84). Also, a weak ATPase activity has been reported for both calnexin and calreticulin (78, 79). The reported presence of calreticulin in the cytosol and nucleus of cells has also generated discussion, as well as the presence of calnexin and calreticulin on the surface of cells (85).

ERp57

The majority of proteins synthesized in the ER acquire disulfide bridges. Oxidation is catalyzed by protein disulfide isomerase (PDI) and other thiol-disulfide oxidoreductases (39). Formation of correct disulfides is generally essential for proper folding. With four thioredoxin-like domains, ERp57 is a close homologue of PDI (86). The extreme N- and C-terminal domains have characteristic CXXC active site sequences. *In vivo*, ERp57 has been shown to form mixed disulfides with incompletely folded glycoproteins (87), and it acts as a reductase in the case of the partially folded major histocompatibility complex Class I heavy chain (88).

ERp57 differs from PDI in that it forms a complex with calnexin and calreticulin, and it specifically interacts with glycoproteins (89, 90). NMR studies and deletion mutants show that it binds to the distal end of the P domain (91, 92). Although the interaction is weak ($K_d \sim 9 \mu\text{M}$) the complex is likely to be stabilized by the formation of mixed disulfides with the substrate glycoprotein. The presence of ERp57 at the tip of the P domain is likely to further confine the space in which a glycoprotein substrate is trapped.

Glucosidase I

Glucosidase I removes the outermost glucose residue attached via an α -1,2 linkage to the middle glucose (residue n) (Figure 1). It is a membrane glycoprotein (type II) of about 82 kDa with a short N-terminal cytosolic peptide, a single transmembrane sequence, and a large, glucosidase-active ectodomain. Together with glucosidase II, glucosidase I prevents binding of the protein-bound glycan to OST and makes possible the entry of a glycoprotein into the calnexin/

calreticulin cycle. Castanospermine and other polyhydroxylated indolizidine alkaloids inhibit both glucosidase I and II (93). They are frequently used to inhibit the entry of newly synthesized glycoproteins into the calnexin/calreticulin cycle.

A glucosidase I defect was recently described in a neonate with severe hypotonia and dysmorphic features (94). The syndrome has now been named congenital disorder of glycosylation type IIb. Disruption of the glucosidase I gene (*gsc-1*) in *Arabidopsis* results in defects in the accumulation of seed storage proteins and in the formation of protein bodies, as well as in deficiencies in cell differentiation during embryonal development (95). Tissue culture cell lines lacking glucosidase I activity are viable.

Glucosidase II

Glucosidase II plays a double role in the calnexin/calreticulin cycle. It prepares the substrate for entry into the cycle and allows substrate exit from the cycle. By using the same enzyme for feeding substrate into the cycle and removing the product, cells may make sure that the cycle cannot be oversaturated.

Glucosidase II is a soluble luminal enzyme composed of two tightly associated glycopolypeptide chains, α and β , with molecular weights of 107 and 54 kDa (96–98). The sequence in the C-terminal half of the α chain contains a catalytic domain belonging to the hydrolase family 31 (99). Both α and β chains occur in several differentially spliced forms (100).

The β chain is a highly conserved glycoprotein. In addition to an N-terminal signal sequence and a C-terminal Lys-Asp-Glu-Leu (KDEL) sequence, it contains a domain homologous to the lectin domain of mannose 6-P receptors (81, 101). As the α chain lacks known ER retention sequences, it is likely that the β chain might serve as a localization subunit. It is also possible that the β chain is needed to allow the α chain to fold properly because coexpression studies have shown that both subunits are essential for enzymatic activity, solubility and/or stability, as well as ER retention of the enzyme (102–104).

Inhibition or genetic disruption of glucosidase II in tissue culture cells leads to accelerated but less efficient glycoprotein folding and secretion, partial breakdown of the quality control system with incompletely folded proteins being secreted, induction of the unfolded protein response, and premature degradation of misfolded glycoproteins (102, 105–108). Disruption of the α chain in *Schizosaccharomyces pombe* results in total loss of glucosidase II activity, induction of the unfolded protein response, and accumulation of cargo proteins in the ER (102, 109). Disruption of the β chain has similar effects except for the formation of small amounts of Glc₁Man₈₋₇GlcNAc₂ chains. Germline mutations in the β subunit in humans are associated with autosomal dominant polycystic liver disease (110, 111). This is an inherited condition in which multiple cysts of biliary epithelial origin occur in the liver.

UDP-Glucose:Glycoprotein Glucosyltransferase

GT, the folding sensor, is undoubtedly the most interesting of the enzymes in the calnexin/calreticulin cycle. Because the properties of GT have been recently reviewed in this series (59), only some of the most important features will be discussed here. GT is a large, soluble, luminal protein with a C-terminal ER-retrieval sequence (112–114). In addition to its localization in the ER, it is present in ER exit sites and in the ER-Golgi intermediate compartment (ERGIC) (115). GT has a C-terminal, catalytic segment of 300 amino acids with homology to members of the glucosyltransferase family 8 (113, 114, 116). The N-terminal sequence of about 1200 residues are presumed to participate in substrate glycoprotein recognition. The N- and C-terminal domains of the enzyme are intimately connected functionally and structurally (117, 118).

GT transfers a glucose residue from UDP-glucose to protein-bound high-mannose glycans. Efficiency decreases with decreasing number of mannoses in branches B and C (52). This is probably important for the interplay between the calnexin cycle and ERAD (see below). Interestingly, glucosidase II shows a similar decrease in efficiency with decreasing mannose number (119). Taken together these observations suggest that a glycoprotein passes more slowly through the cycle as mannoses are progressively lost.

How does GT sense the folding status of a substrate glycoprotein? Currently there is only a partial answer to this question. It is clear that selectivity must be based on general features shared by misfolded proteins. The enzyme makes, for example, no distinction between glycoproteins and neoglycoproteins of bacterial or nonsecretory origin (59). There are at least two general ways in which a protein might be able to sense whether another protein is incompletely folded: through exposed hydrophobic peptides or patches or through excessive surface dynamics. Which of these apply to GT is not clear. That the data is still incomplete is in part due to unsatisfactory *in vitro* assays that make use of aggregated substrates with heterogeneous conformations and glycan configurations.

The enzyme does not use free glycans or small glycopeptides as substrate, hence the name glycoprotein glucosyltransferase. Studies with monomeric, nonaggregated, well-characterized glycoprotein substrates, such as RNase B, show that a random coil is not a substrate; whereas molten globule-like, partially folded proteins are good substrates (120–122). However, a recent study using Man₉GlcNAc₂-containing tryptic peptides showed that GT does reglucosylate glycopeptides with high efficiency provided that they have more than 12 amino acid residues and that the sequence includes hydrophobic residues on either the N-terminal or C-terminal side of the glycan (123). Such short peptides are unlikely to have secondary or tertiary structure.

Because the presence of misfolded, nonglycosylated proteins does not inhibit glucose transfer to denatured thyroglobulin (124), it was hypothesized that GT recognizes the protein and the glycan moieties together. This was supported by

the observation that the presence of the innermost GlcNAc residue of the N-linked glycan was sufficient to make a denatured protein inhibitory to the enzyme. The tendency of GT to bind to hydrophobic peptides in the absence of glycans argues, on the other hand, that GT can recognize and bind to proteins in the absence of glycans (124). Recent studies in our group have shown, furthermore, that misfolded RNase A devoid of glycans interferes with GT activity, suggesting sugar-independent interactions (C. Ritter, K. Quirin, and A. Helenius, unpublished results).

In studies with RNase B dimers, it was found that GT reglucosylates glycans that are located in a misfolded domain while ignoring glycans in a nearby identical but folded domain of the same protein (121). When local misfolding was induced by mutating one of the disulfide bonds, only glycans in the misfolded region were glucosylated, suggesting highly localized glucosylation (C. Ritter, K. Quirin, and A. Helenius, unpublished results).

In summary, current data indicate that GT interacts with glycoproteins and possibly nonglycoproteins but only if they are improperly folded. If the incompletely folded regions contain high-mannose glycans, these are selectively reglucosylated. The enzyme recognizes relatively small local folding defects, and the presence of exposed hydrophobic residues improves recognition. A partially ordered, molten globule-like structure seems to be optimal for efficient recognition.

UDPase

GT uses as a substrate a nucleotide sugar, UDP-Glc, transported into the ER lumen from the cytosol. An import activity for UDP-Glc has been described in the ER of rat liver and in *S. cerevisiae*, and it has been established that import of UDP-Glc is coupled to the exit of UMP (125). To make use of this antiporter, the ER must first convert the UDP generated by GT to UMP. In the secretory pathway of mammalian cells, three nucleoside diphosphatases are known, two in the ER and one in the Golgi complex.

Two phosphatases have been reported in the ER, a soluble and a membrane-bound enzyme (126, 127). The soluble UDPase is a glycoprotein that needs Ca^{2+} , Mg^{2+} , or Mn^{2+} to work, whereas the membrane-bound UDPase only functions with Ca^{2+} . Both enzymes belong to the ecto-nucleoside triphosphate diphosphohydrolase (E-NTPDase) family, and hydrolyze UDP, GDP, and IDP but not nucleoside triphosphates or ADP. Although ubiquitously expressed, the two enzymes have somewhat different tissue distribution. The soluble UDPase is enriched in the ER, but it has no known ER retention sequence. When expressed in Cos-7 cells, it is secreted (128). The membrane-bound enzyme occurs on the ER and the intermediate compartment. A RXRXR sequence at the N terminus may serve as an ER retention sequence. The protein is anchored to the membrane via sequences close to the N terminus and has its active domain in the ER lumen.

ER-DEPENDENT DEGRADATION OF GLYCOPROTEINS

As outlined above, N-linked oligosaccharides play an essential role in the folding of glycoproteins in the ER. If glycoproteins fail to fold or to oligomerize, they are retained in the ER and eventually degraded. The degradation process, termed ER-associated degradation (ERAD), is important because it prevents accumulation of unsalvageable, misfolded proteins in the ER.

ERAD involves three functionally distinct steps: the recognition of a glycoprotein as malformed, its retrotranslocation to the cytoplasm, and the subsequent, ubiquitin-dependent degradation by the proteasome. Several excellent articles have recently reviewed aspects of this process (129–132). Here, we will focus exclusively on the recognition step.

Targeting of substrates to ERAD is a delicate process. Misfolded and unassembled protein subunits should be degraded, but bona fide folding intermediates should not. Current data shows that such a distinction does indeed take place. It is based on the length of time that a glycoprotein has spent in the ER. In mammalian cells, degradation of a misfolded glycoprotein typically starts after a lag period of 30–90 min followed by exponential decay (133–135). This means that cells give newly synthesized proteins a fair chance to fold and assemble before degradation sets in.

The use of a timer mechanism does, however, create a problem: how to spare resident ER glycoproteins destined to stay permanently in the ER? Unless misfolded, they should not be targeted for destruction after the initial lag. It is possible that the system does not rely exclusively on a timer but also on a folding sensor (136).

The delay in degradation of glycoproteins by ERAD is clearly linked to trimming of mannoses. The most important mannosidase is an α -1,2 exomannosidase, a membrane-bound ER enzyme that specifically removes the terminal mannose (residue i) from the B branch of the oligosaccharide to yield the $\text{Man}_8\text{GlcNAc}_2$ B-isomer (Figure 1) (137). If the mannosidase is inhibited by kifunensin (a specific inhibitor) or mutated, glycoprotein degradation is dramatically slower (138). In contrast, if ER mannosidase I is overexpressed, the onset of degradation is accelerated (139). Because the activity of this and other mannosidases in the ER is relatively low (136), the mannoses are removed over a period that varies from 10 min in yeast to more than an hour in some mammalian cells. The slow-acting mannosidase is likely to serve as the timer that protects newly synthesized proteins and nascent chains. That the lag times vary greatly for different proteins may depend on the number of glycans, their locations in the protein, and other effects.

Genetic evidence in *S. cerevisiae* suggests that the $\text{Man}_8\text{GlcNAc}_2$ B-isomer generated by ER mannosidase I is recognized by a membrane-bound ER protein called Htm1/Mnl1 (140, 141). Htm1/Mnl1 is a mannosidase homologue, but because it does not have detectable mannosidase activity, it is thought to serve as a mannose lectin and to be responsible for directing glycoproteins into the

retrotranslocation and degradation pathway. Elimination of this membrane protein retards ERAD of glycoproteins but not of nonglycosylated proteins (140). The generation of $\text{Man}_8\text{GlcNAc}_2$ B-isomer is required for efficient ERAD, yet it is not sufficient because only malformed proteins with $\text{Man}_8\text{GlcNAc}_2$ are degraded (136).

In mammalian cells, ER mannosidase I, the homologue of the yeast mannosidase, plays a similar role. Glycoprotein degradation is dramatically reduced by kifunensin, and when overexpressed, a homologue of Htm1p/Mnl1p, called EDEM [for ER degradation-enhancing α -mannosidase-like protein (142)], accelerates the degradation of malformed glycoproteins (134, 143). The presence of a functional calnexin/calreticulin cycle makes the system more complex than in yeast. In some cases, the calnexin cycle seems to protect glycoproteins against degradation during the initial lag period. If glycoproteins are not allowed to enter the calnexin cycle, ERAD starts without lag (134, 135). It is likely that once mannose trimming has occurred, EDEM begins competing for substrates with the calnexin/calreticulin cycle. Recall that when mannoses are lost from the B and C branches, both GT and glucosidase II become less efficient (52, 119). The substrate is therefore likely to pass more slowly through the cycle and may in this way give EDEM an advantage. It has also been shown that EDEM and calnexin coimmunoprecipitate, suggesting that they associate with each other in a complex (143). It is possible that EDEM acquires substrate glycoproteins directly from calnexin.

The role of additional mannosidases is not yet fully clarified. In addition to the ER α -1,2-mannosidase I, mammalian cells have an ER mannosidase II that removes a mannose (residue k) from the C branch (Figure 1) (137). Further α -1,2 mannosidases of the same enzyme family occur in the Golgi of mammalian cells. Because some of the mannosidases preferentially act on either the A or the C arm (137), they could, if present in the ER, be used to extract glycoproteins from the calnexin cycle or prevent ERAD. Indeed, it has been recently reported that ERAD substrates can be trimmed to $\text{Man}_7\text{GlcNAc}_2$ or even $\text{Man}_5\text{GlcNAc}_2$, suggesting exposure to several mannosidases (144, 145). These data suggest that a degradation pathway independent of the $\text{Man}_8\text{GlcNAc}_2$ isomer B exists in mammalian cells (146, 147).

It is noteworthy that most mammalian cells have in the ERGIC and in the Golgi complex an endomannosidase that cleaves the linkage between the first glucose-substituted mannose (g) and the second mannose (f) in the A branch (Figure 1) (18). This enzyme is thought to remove the glucose from monoglucosylated chains that have somehow escaped the action of ER glucosidases and thus permit normal processing of the glycans in the Golgi complex.

OTHER LECTINS IN THE SECRETORY PATHWAY

Lectins Related to Mannose 6-Phosphate Receptors

That the secretory pathway contains functionally important lectins was first recognized during an effort to clarify the molecular causes of the hereditary

disease mucopolipidosis II (I-cells disease). It was realized that modified N-linked glycans are address tags in the intracellular targeting of lysosomal enzymes and that lectins serve as receptors (148). The lectins in question are the two mannose 6-phosphate binding receptors (MPRs) that transport lysosomal enzymes from the *trans*-Golgi network to endosomes. They have been thoroughly analyzed with respect to function and structure (149).

The crystal structure of the lectin domain in Ca^{2+} -dependent (CD)-MPR, the smaller of the receptors, shows a flattened β -barrel structure with three disulfide bonds (150). The carbohydrate binding site is located in a relatively deep pocket formed by the β -sheets and the loops that connect them. The site contains a bound Mn^{2+} that interacts with the phosphate group of mannose 6-phosphate. Binding of the ligand causes a conformational change in the receptor, and binding is reversed by mildly acidic pH.

Analysis of sequence databases recently allowed Munro (101) to identify additional MPR-like domains in a variety of genes and proteins. He classified these as members of a family and called them mannose 6-phosphate receptor homology domains (MRHs). The conservation of sequences in the region of the glycan binding pocket in CD-MPR suggested that they may also be lectins. Consistent with this, he observed that in addition to genes of unknown function, the proteins with MRH domains include two N-glycan modifying enzymes: the ER glucosidase II β -subunit discussed above and the lysosomal hydrolase N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) γ -subunit (151). This is the *cis*-Golgi enzyme responsible for adding GlcNAc-phosphate to lysosomal hydrolases. Whether these enzyme-associated MRH domains act as lectins remains to be determined. Loss of the γ -subunit leads to a defect in mannose 6-phosphate signal generation, mistargeting of lysosomal hydrolases, and a variant of pseudohurler polydystrophy (mucopolipidosis IIIC) (152).

Leguminous Lectin Homologues

In addition to calnexin and calreticulin, the early secretory pathway contains several other lectins that have been classified as members of the leguminous lectin family. Best characterized is the ubiquitously expressed and abundant ERGIC-53 (p58, MR60), a homo-oligomeric, type I transmembrane protein with mannose lectin activity (153). The X-ray crystal structure of the carbohydrate-binding domain shows the β -sandwich structure characteristic for this lectin family (68). No long appendices like the P domain in calnexin and calreticulin are present.

Equipped with a complex repertoire of cytoplasmic tail signals for interaction with COPI and COPII coats, ERGIC-53 cycles between the ERGIC and the *cis*-Golgi. Although it has been difficult to define physiological ligands, it is apparent that ERGIC-53 is responsible for ER-to-Golgi export of a subset of secretory and lysosomal glycoproteins. Association with cathepsin Z, a lysosomal enzyme, seems to be dependent on glucose trimming, implying that

ERGIC-53 binds to glycoproteins that have successfully emerged from the calnexin/calreticulin cycle (154). Mutations in ERGIC-53 can, in humans, result in an autosomal recessive bleeding disorder caused by a deficiency in coagulation factors V and VIII (F5F8D) (155).

Vesicular integral protein 36 (VIP36) is another membrane-bound leguminous lectin concentrated primarily in the early secretory pathway (156–158). This lectin is homologous with ERGIC-53 but lacks the coiled-coil structure in the stem. It is specific for high-mannose type N-glycans containing β -1,2 mannoses, and it is thought to mediate forward transport of glycoproteins in the secretory pathway, and possibly their targeting to the apical plasma membrane in polarized epithelial cells.

Examination of genomic and EST databases has led to the identification of additional mammalian proteins with homology to ERGIC-53 and VIP36 (159, 160). The homologues have been classified into three main phylogenetic groups: the VIP36/VIPL, the ERGIC-53/ERGL, and a group of fungal homologues that include Emp46p and Emp47p in *S. cerevisiae* (159). Of these, VIPL is a highly conserved VIP36-like protein with orthologues in many species from humans to *Xenopus* (160). Similar to VIP36, it lacks the coiled stem domain present in ERGIC-53 and in the fungal homologues and is therefore less likely to assemble into an oligomer. It has a cytosolic ER retention motif and is mainly located in the ER. Although the primary sequence suggests that it is a mannose-binding lectin, lectin activity has not been experimentally confirmed for any of these proteins except ERGIC-53 and VIP36. Overexpression of VIPL interferes with ERGIC-53 cycling, suggesting that it may regulate ERGIC-53 function (160). Knockdown experiments have revealed retardation in the export of a subset of proteins (87, 122, 159). Emp46p and Emp47p are homologous membrane proteins in the early secretory pathway of *S. cerevisiae* that cause a partial secretion defect when disrupted (161).

Taken together, it is apparent that proteins with lectin domains occur throughout the secretory pathway. They belong to several families, and many of them are likely to bind mannose (162). Their functions are related to folding, quality control, sorting, and forward transport of glycoprotein cargo. Because they are functionally redundant, and their affinities for their ligands are weak, their biochemical characterization is challenging. However, they are an interesting new class of proteins that have their functions in the lumen of secretory organelles promoting sorting, cargo selection, and efficiency.

N-GLYCANS: THE EVOLUTION OF A SIGNAL

The functions of N-glycans can be further illuminated by considering the evolutionary origin of ER glycosylation. As in most functions of the ER, the synthesis of N-linked glycans stems from homologous processes in the plasma membrane of bacteria or archaea (11, 163). Proteins with N-linked glycans are,

in fact, present in the outermost layer of the archaeal cell wall (164) and in the cell wall of certain gram-negative bacteria, such as *Campylobacter jejuni* (35, 165). The synthesis of prokaryotic glycoproteins shows such striking similarities to the process in eukaryotic cells that there can be no doubt that the processes are homologous.

In archaea, the oligosaccharide moiety is assembled at the cytoplasmic face of the plasma membrane on a lipid carrier, dolichylphosphate or dolichylpyrophosphate. It is then translocated through the plasma membrane and enzymically transferred to a protein on the outside surface of the plasma membrane. The use of a lipid-bound oligosaccharide precursor makes sense in these organisms, because a soluble precursor would diffuse away from the cell. The acceptor sequence in the polypeptide is none other than the familiar Asn-X-Ser/Thr (166). In bacteria, the oligosaccharide precursors are probably bound to bactoprenol, an isoprenoid lipid very similar to dolichol. A homologue of the STT3 subunit of OST in *C. jejuni* is essential and sufficient for oligosaccharide transfer to protein (35; M. Feldman, M. Wacker, and M. Aebi, unpublished results), and Asn-X-Ser/Thr serves as an acceptor sequence in the bacterial system.

Although the membrane topology and oligosaccharide transfer mechanisms are similar, there are differences between eukaryotes and prokaryotes. The limited information available suggests, for example, that there is great diversity among oligosaccharides transferred to proteins in different prokaryote species (164, 167, 168). In contrast, almost all eukaryotes transfer the same structure, $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ (Figure 1). A notable exception is within the clade of the trypanosomatids, primitive eukaryotes, in which the $\text{Man}_{6-9}\text{GlcNAc}_2$ oligosaccharides are transferred to proteins (169, 170). These organisms lack specific Dol-P-Man or Dol-P-Glc-dependent mannosyl- and glycosyltransferase required for the assembly of the complete $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$, but they do have GT.

Thus, we can view the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ oligosaccharide as a bipartite structure in the evolutionary sense: The $\text{Man}_5\text{GlcNAc}_2$ oligosaccharide, assembled in the cytoplasm, probably corresponds to the glycan transferred to protein in the archaeal ancestor of all eukaryotic cells, whereas the three glucose and the four mannose residues, added in the lumen of the ER, are extensions that have occurred during eukaryotic evolution. The situation in trypanosomatids might represent an evolutionary intermediate.

The $\text{Man}_5\text{GlcNAc}_2$ oligosaccharide (residues a-g) (Figure 1) is closer to the protein and functionally distinct. These sugars seem to be primarily responsible for the direct effects that N-linked glycans have in protein folding (3, 6). Five of them escape trimming in the ER and the Golgi complex when complex glycans are synthesized.

We speculate that the addition of further saccharides (mannoses h, i, j, and k, and glucoses m, n, and l) to the archaeal $\text{Man}_5\text{GlcNAc}_2$ oligosaccharide during eukaryotic evolution was driven by the internalization of glycoprotein biosynthesis from the plasma membrane to the ER and by the concomitant need to export newly synthesized proteins to the cell surface. Internalization made it

possible to sequester biosynthesis, folding, and early maturation in a closed compartment in which conditions could be better controlled than in a space connected to the extracellular medium. It also opened new possibilities to increase the volume of production and to expand the product repertoire to include more complex proteins and proteins that required assistance from a variety of chaperones. The disadvantages included the need to control the fidelity of protein maturation and necessity to deal with misfolded proteins, an inevitable side effect of protein folding and oligomeric assembly. The distal mannose and glucose residues (h-n) of the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ oligosaccharide were introduced to serve as recognition tags in the calnexin/calreticulin cycle during quality control and during substrate selection for ERAD. These distal sugars made it possible for the cell to use the N-linked glycans as a composite signal.

THE N-LINKED GLYCAN, A COMPOSITE, MULTIFUNCTIONAL SIGNAL

The different signaling functions of the protein-bound core oligosaccharide are organized in such a way that the A branch provides information about the folding status, and the B and C arms are used for ERAD (Figure 1). The signals are interpreted by specific lectins: calnexin and calreticulin in the case of the A branch and EDEM/HTM1 in the case of the B and C arms. As discussed above, efficient binding of calnexin to the A branch requires the presence of the α -1,3-linked glucose residue (l); however, the three mannose residues g, f, and d significantly enhance the affinity. The monoglucosylated A branch acts as a positive signal for binding. However, this signal can be efficiently masked by the presence of the additional α -1,3-linked glucose (m). This residue functions as a negative signal for calnexin and calreticulin binding.

Similarly the three mannose residues (h, k, and l) of the B and C arm of the oligosaccharide may act as an optimal substrate for EDEM and as a positive signal in the ERAD pathway. Indeed, the structure of the EDEM homologue ER mannosidase I with a bound Man_8 oligosaccharide (171) is compatible with this hypothesis. Again, genetic evidence suggests that the α -1,2-linked mannose residue i of the B branch acts as a negative signal for EDEM/HTM1 binding.

On the basis of these considerations, we can now propose functions for each of the individual trimming steps in the ER. The outermost α -1,2 glucose residue (n) is needed for efficient recognition of the processor oligosaccharide by OST. OST actually sees the complete A branch, but glucose n seems to be a key residue (16). Hydrolysis of this glucose residue is likely to promote release of the glycoprotein substrate from OST and prevent rebinding of the substrate after transfer to a polypeptide chain. Removal of this glucose, furthermore, exposes the α -1,3-linked glucose residue (m), which as discussed above serves as a

negative signal for binding to calnexin and calreticulin. Only after removal of m is the critical third glucose (l) exposed terminally, allowing association of calnexin and calreticulin with the A branch. Similarly, removal of mannose is likely to expose a branched oligomannose structure for binding to EDEM/HTM1. Our guess would be that EDEM binds a tetra mannose unit of h, e, j, and k with h as the key residue. Loss of mannoses from the C branch may weaken the interaction (136).

PERSPECTIVES

The ER is a highly specialized compartment. Although hard numbers are missing, it is estimated that many of the chaperones and folding enzymes are present in nearly millimolar concentrations. With such extreme crowding, it is not surprising that the resident proteins tend to form complexes and extensive networks (173–175). It is a world where affinities tend to be low (in the micro- to millimolar range), and molecular interactions transient. The spectrum of chaperones, lectins, and enzymes is delicately balanced to best serve the secretory process of each specific cell type under physiological conditions (176–178). In most cell types, the conditions are optimized for the synthesis of glycoproteins.

Among the challenges in this field are a better grasp of the complexities of the physiological context and more complete understanding of the rules that prevail in this compartment. It is important to learn to manipulate the conditions, to develop cures for various ER-storage diseases, to control the production of glycoproteins, and to explore in further detail the complex pathways by which proteins mature in the ER.

The analysis of glycoprotein folding and quality control in the ER has opened a window into the intracellular functions of oligosaccharides. One of the most important functions is clearly to secure efficient protein production. In the eukaryotic cell, the fidelity of this biosynthetic process has top priority. What might have started as a simple structural component of the extracellular matrix of an archaebacterium has evolved into a highly coded recognition signal whose idiom is now modulated and interpreted by a spectrum of enzymes and lectins in the ER and other intracellular compartments.

ACKNOWLEDGMENTS

We thank Lars Ellgaard, Matthias Gautschi, Paola Deprez, Sergio Trombetta, and Claude Jakob for advice. Work in our labs was supported by the Swiss National Science Foundation and ETH Zurich.

The *Annual Review of Biochemistry* is online at <http://biochem.annualreviews.org>

LITERATURE CITED

1. Kornfeld R, Kornfeld S. 1985. *Annu. Rev. Biochem.* 54:631–64
2. Varki A. 1993. *Glycobiology* 3:97–130
3. Trombetta ES. 2003. *Glycobiology* 13: R77–91
4. Apweiler R, Hermjakob H, Sharon N. 1999. *Biochim. Biophys. Acta* 1473:4–8
5. Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA. 2002. *Chem. Rev.* 102:371–86
6. Petrescu AJ, Milac A-L, Petrescu SM, Dwek RA, Wormald MR. 2004. *Glycobiology* 14:103–14
7. Rudd PM, Wormald MR, Stanfield RL, Huang M, Mattsson N, et al. 1999. *J. Mol. Biol.* 293:351–66
8. Lowe JB, Marth JD. 2003. *Annu. Rev. Biochem.* 72:643–91
9. Weis WI, Drickamer K. 1996. *Annu. Rev. Biochem.* 65:441–73
10. Lis H, Sharon N. 1998. *Chem. Rev.* 98: 637–74
11. Burda P, Aebi M. 1999. *Biochim. Biophys. Acta* 1426:239–57
12. Hirschberg CB, Snider MD. 1987. *Annu. Rev. Biochem.* 56:63–87
13. Helenius J, Ng DTW, Marolda CL, Walter P, Valvano MA, Aebi M. 2002. *Nature* 415:447–50
14. Balasubramanian K, Schroit AJ. 2003. *Annu. Rev. Physiol.* 65:701–34
15. Helenius J, Aebi M. 2002. *Semin. Cell Dev. Biol.* 13:171–78
16. Burda P, Jakob CA, Beinbauer J, Hege- mann JH, Aebi M. 1999. *Glycobiology* 9:617–25
17. Burda P, Aebi M. 1998. *Glycobiology* 8: 455–62
18. Spiro RG. 2000. *J. Biol. Chem.* 275: 35657–60
19. Nilsson IM, von Heijne G. 1993. *J. Biol. Chem.* 268:5798–801
20. Kowarik M, Kung S, Martoglio B, Hele- nius A. 2002. *Mol. Cell* 10:769–78
21. Bause E. 1983. *Biochem. J.* 209:331–36
22. Imperiali B, Shannon KL, Unno M, Rickert KW. 1992. *J. Am. Chem. Soc.* 114:7944–45
23. Knauer R, Lehle L. 1999. *J. Biol. Chem.* 274:17249–56
24. Karaoglu D, Kelleher DJ, Gilmore R. 1995. *J. Cell Biol.* 130:567–78
25. Karaoglu D, Kelleher DJ, Gilmore R. 1997. *J. Biol. Chem.* 272:32513–20
26. Dempski RE Jr, Imperiali B. 2002. *Curr. Opin. Chem. Biol.* 6:844–50
27. Knauer R, Lehle L. 1999. *Biochim. Biophys. Acta* 1426:259–73
28. Kelleher DJ, Karaoglu D, Mandon EC, Gilmore R. 2003. *Mol. Cell* 12:101–11
29. Karaoglu D, Kelleher D, Gilmore R. 2001. *Glycobiology* 11:54
30. Yan G, Lennarz WJ. 2002. *J. Biol. Chem.* 277:47692–700
31. Nilsson I, Kelleher DJ, Miao YW, Shao YL, Kreibich G, et al. 2003. *J. Cell Biol.* 161:715–25
32. Yan Q, Prestwich GD, Lennarz WJ. 1999. *J. Biol. Chem.* 274:5021–25
33. Pathak R, Hendrickson TL, Imperiali B. 1995. *Biochemistry* 34:4179–85
34. Burda P, te Heesen S, Brachat A, Wach A, Düsterhöft A, Aebi M. 1996. *Proc. Natl. Acad. Sci. USA* 93:7160–65
35. Wacker M, Linton D, Hitchen PG, Nita- Lazar M, Haslam SM, et al. 2002. *Sci- ence* 298:1790–93
36. Bergman LW, Kuehl WM. 1979. *J. Biol. Chem.* 254:8869–76
37. Chen W, Helenius J, Braakman I, Hele- nius A. 1995. *Proc. Natl. Acad. Sci. USA* 92:6229–33
38. Daniels R, Kurowski B, Johnson AE, Hebert DN. 2003. *Mol. Cell* 11:79–90
39. Fassio A, Sitia R. 2002. *Histochem. Cell Biol.* 117:151–57
40. Olden K, Parent JB, White SL. 1982. *Biochim. Biophys. Acta* 650:209–32
41. Hurlley SM, Helenius A. 1989. *Annu. Rev. Cell Biol.* 5:277–307
42. Klausner RD, Sitia R. 1990. *Cell* 62: 611–14

43. Helenius A. 1994. *Mol. Biol. Cell* 5: 253–65
44. Hebert DN, Zhang JX, Chen W, Foellmer B, Helenius A. 1997. *J. Cell Biol.* 139:613–23
45. Imperiali B, O'Connor SE. 1999. *Curr. Opin. Chem. Biol.* 3:643–49
46. O'Connor SE, Imperiali B. 1996. *Chem. Biol.* 3:803–12
47. Imberty A, Perez S. 1995. *Protein Eng.* 8:699–709
48. Wormald MR, Dwek RA. 1999. *Struct. Fold. Des.* 7:R155–60
49. Kundra R, Kornfeld S. 1999. *J. Biol. Chem.* 274:31039–46
50. DeKoster GT, Robertson AD. 1997. *Biochemistry* 36:2323–31
51. Suh K, Bergmann JE, Gabel CA. 1989. *J. Cell Biol.* 108:811–19
52. Sousa MC, Ferrero-Garcia MA, Parodi AJ. 1992. *Biochemistry* 31:97–105
53. Ou WJ, Cameron PH, Thomas DY, Bergeron JJ. 1993. *Nature* 364:771–76
54. Hammond C, Helenius A. 1994. *Science* 266:456–58
55. Hammond C, Helenius A. 1993. *Curr. Biol.* 3:884–85
56. Nauseef WM, McCormick SJ, Clark RA. 1995. *J. Biol. Chem.* 270:4741–47
57. Peterson JR, Ora A, Van PN, Helenius A. 1995. *Mol. Biol. Cell* 6:1173–84
58. Ellgaard L, Molinari M, Helenius A. 1999. *Science* 286:1882–88
59. Parodi A. 2000. *Annu. Rev. Biochem.* 69: 69–93
60. Schrag JD, Bergeron JJ, Li Y, Borisova S, Hahn M, et al. 2001. *Mol. Cell* 8: 633–44
61. Chevet E, Cameron PH, Pelletier MF, Thomas DY, Bergeron JJ. 2001. *Curr. Opin. Struct. Biol.* 11:120–24
62. Ellgaard L, Helenius A. 2003. *Nat. Rev. Mol. Cell Biol.* 4:181–91
63. Denzel A, Molinari M, Trigueros C, Martin JE, Velmurgan S, et al. 2002. *Mol. Cell Biol.* 22:7398–404
64. Michalak M, Lynch J, Groenendyk J, Guo L, Parker JMR, Opas M. 2002. *Biochim. Biophys. Acta* 1600:32–37
65. Ellgaard L, Riek R, Braun D, Herrmann T, Helenius A, Wuthrich K. 2001. *FEBS Lett.* 488:69–73
66. Chevet E, Wong HN, Gerber D, Cochet C, Fazel A, et al. 1999. *EMBO J.* 18: 3655–66
67. Bouvier M, Stafford WF. 2000. *Biochemistry* 39:14950–59
68. Velloso LM, Svensson K, Schneider G, Pettersson RF, Lindqvist Y. 2002. *J. Biol. Chem.* 277:15979–84
69. Rudenko G, Nguyen T, Chelliah Y, Sudhof TC, Deisenhofer J. 1999. *Cell* 99: 93–101
70. Loris R. 2002. *Biochim. Biophys. Acta* 1572:198–208
71. Ellgaard L, Riek R, Herrmann T, Guntert P, Braun D, et al. 2001. *Proc. Natl. Acad. Sci. USA* 98:3133–38
72. Ware FE, Vassilakos A, Peterson PA, Jackson MR, Lehrman MA, Williams DB. 1995. *J. Biol. Chem.* 270:4697–704
73. Vassilakos A, Michalak M, Lehrman MA, Williams DB. 1998. *Biochemistry* 37:3480–90
74. Spiro RG, Zhu Q, Bhojroo V, Soling HD. 1996. *J. Biol. Chem.* 271:11588–94
75. Patil AR, Thomas CJ, Suroliya A. 2000. *J. Biol. Chem.* 275:24348–56
76. Kapoor M, Srinivas H, Kandiah E, Gemma E, Ellgaard L, et al. 2003. *J. Biol. Chem.* 278:6194–200
77. Kapoor M, Ellgaard L, Gopalakrishnapai J, Schirra C, Gemma E, et al. 2004. *Biochemistry* 43:97–106
78. Ihara Y, Cohen-Doyle MF, Saito Y, Williams DB. 1999. *Mol. Cell* 4:331–41
79. Saito Y, Ihara Y, Leach MR, Cohen-Doyle MF, Williams DB. 1999. *EMBO J.* 18:6718–29
80. Danilczyk UG, Williams DB. 2001. *J. Biol. Chem.* 276:25532–40
81. Rodan AR, Simons JF, Trombetta ES, Helenius A. 1996. *EMBO J.* 15:6921–30
82. Zapun A, Petrescu SM, Rudd PM, Dwek

- RA, Thomas DY, Bergeron JJM. 1997. *Cell* 88:29–38
83. Corbett EF, Michalak KM, Oikawa K, Johnson S, Campbell ID, et al. 2000. *J. Biol. Chem.* 275:27177–85
 84. Ou WJ, Bergeron JJ, Li Y, Kang CY, Thomas DY. 1995. *J. Biol. Chem.* 270:18051–59
 85. Papp S, Opas M. 2003. In *Calreticulin*, ed. P Eggleton, M Michalak, pp. 38–45. Georgetown, TX: Landes Biosci.
 86. Hirano N, Shibasaki F, Sakai R, Tanaka T, Nishida J, et al. 1995. *Eur. J. Biochem.* 234:336–42
 87. Molinari M, Helenius A. 1999. *Nature* 402:90–93
 88. Antoniou AN, Ford S, Alphey M, Osborne A, Elliott T, Powis SJ. 2002. *EMBO J.* 21:2655–63
 89. Elliott JG, Oliver JD, High S. 1997. *J. Biol. Chem.* 272:13849–55
 90. Oliver JD, Roderick HL, Llewellyn DH, High S. 1999. *Mol Biol. Cell* 10:2573–82
 91. Frickel EM, Riek R, Jelesarov I, Helenius A, Wuthrich K, Ellgaard L. 2002. *Proc. Natl. Acad. Sci. USA* 99:1954–59
 92. Leach MR, Cohen-Doyle MF, Thomas DY, Williams DB. 2002. *J. Biol. Chem.* 277:29686–97
 93. Elbein AD. 1983. *Methods Enzymol.* 98:135–54
 94. De Praeter CM, Gerwig GJ, Bause E, Nuytink LK, Vliegenthart JF, et al. 2000. *Am. J. Hum. Genet.* 66:1744–56
 95. Boisson M, Gomord V, Audran C, Berger N, Dubreucq B, et al. 2001. *EMBO J.* 20:1010–19
 96. Brada D, Dubach UC. 1984. *Eur. J. Biochem.* 141:149–56
 97. Hettkamp H, Legler G, Bause E. 1984. *Eur. J. Biochem.* 142:85–90
 98. Trombetta ES, Simons JF, Helenius A. 1996. *J. Biol. Chem.* 271:27509–16
 99. Henrissat B, Bairoch A. 1993. *Biochem. J.* 293(Pt. 3):781–88
 100. Arendt CW, Dawicki W, Ostergaard HL. 1999. *Glycobiology* 9:277–83
 101. Munro S. 2001. *Curr. Biol.* 11:R499–501
 102. D'Alessio C, Fernandez F, Trombetta ES, Parodi AJ. 1999. *J. Biol. Chem.* 274:25899–905
 103. Pelletier MF, Marcil A, Sevigny G, Jakob CA, Tessier DC, et al. 2000. *Glycobiology* 10:815–27
 104. Trembl K, Meimaroglou D, Hentges A, Bause E. 2000. *Glycobiology* 10:493–502
 105. Hebert DN, Foellmer B, Helenius A. 1996. *EMBO J.* 15:2961–68
 106. Branza-Nichita N, Petrescu AJ, Dwek RA, Wormald MR, Platt FM, Petrescu SM. 1999. *Biochem. Biophys. Res. Commun.* 261:720–25
 107. Machold RP, Ploegh HL. 1996. *J. Exp. Med.* 184:2251–59
 108. Zhang JX, Braakman I, Matlack KE, Helenius A. 1997. *Mol. Biol. Cell* 8:1943–54
 109. Labriola C, Cazzulo JJ, Parodi AJ. 1995. *J. Cell Biol.* 130:771–79
 110. Drenth JP, te Morsche RH, Smink R, Bonifacino JS, Jansen JB. 2003. *Nat. Genet.* 33:345–47
 111. Li AR, Davila S, Furu L, Qian Q, Tian X, et al. 2003. *Am. J. Hum. Genet.* 72:691–703
 112. Trombetta S, Bosch M, Parodi AJ. 1989. *Biochemistry* 28:8108–16
 113. Parker CG, Fessler LI, Nelson RE, Fessler JH. 1995. *EMBO J.* 14:1294–303
 114. Arnold SM, Fessler LI, Fessler JH, Kaufman RJ. 2000. *Biochemistry* 39:2149–63
 115. Roth J. 2002. *Chem. Rev.* 102:285–303
 116. Tessier DC, Dignard D, Zapun A, Radominska-Pandya A, Parodi AJ, et al. 2000. *Glycobiology* 10:403–12
 117. Arnold SM, Kaufman RJ. 2003. *J. Biol. Chem.* 278:43320–28
 118. Guerin M, Parodi AJ. 2003. *J. Biol. Chem.* 278:20540–46
 119. Grinna LS, Robbins PW. 1980. *J. Biol. Chem.* 255:2255–58

120. Trombetta ES, Helenius A. 2000. *J. Cell Biol.* 148:1123–29
121. Ritter C, Helenius A. 2000. *Nat. Struct. Biol.* 7:278–80
122. Caramelo JJ, Castro OA, Alonso LG, De Prat-Gay G, Parodi AJ. 2003. *Proc. Natl. Acad. Sci. USA* 100:86–91
123. Taylor SC, Thibault P, Tessier DC, Bergeron JJ, Thomas DY. 2003. *EMBO Rep.* 4:405–11
124. Sousa M, Parodi AJ. 1995. *EMBO J.* 14: 4196–203
125. Perez M, Hirschberg CB. 1986. *J. Biol. Chem.* 261:6822–30
126. Trombetta SE, Parodi AJ. 1992. *J. Biol. Chem.* 267:9236–40
127. Failer BU, Braun N, Zimmermann H. 2002. *J. Biol. Chem.* 277:36978–86
128. Mulero JJ, Yeung G, Nelken ST, Ford JE. 1999. *J. Biol. Chem.* 274:20064–67
129. Tsai B, Ye Y, Rapoport TA. 2002. *Nat. Rev. Mol. Cell Biol.* 3:246–55
130. Kostova Z, Wolf DH. 2003. *EMBO J.* 22:2309–17
131. McCracken AA. 2003. *BioEssays* 25: 868–77
132. Yoshida Y. 2003. *J. Biochem.* 134: 183–90
133. Lippincott-Schwartz J, Bonifacino JS, Yuan LC, Klausner RD. 1988. *Cell* 54: 209–20
134. Molinari M, Calanca V, Galli C, Lucca P, Paganetti P. 2003. *Science* 299: 1397–400
135. Mancini R, Aebi M, Helenius A. 2003. *J. Biol. Chem.* 278:46895–905
136. Jakob CA, Burda P, Roth J, Aebi M. 1998. *J. Cell Biol.* 142:1223–33
137. Herscovics A. 2001. *Biochimie* 83: 757–62
138. Liu Y, Choudhury P, Cabral CM, Sifers RN. 1999. *J. Biol. Chem.* 274:5861–67
139. Wu Y, Swulius MT, Moremen KW, Sifers RN. 2003. *Proc. Natl. Acad. Sci. USA* 100:8229–34
140. Jakob CA, Bodmer D, Spirig U, Battig P, Marcil A, et al. 2001. *EMBO Rep.* 2: 423–30
141. Nakatsukasa K, Nishikawa S, Hosokawa N, Nagata K, Endo T. 2001. *J. Biol. Chem.* 276:8635–38
142. Hosokawa N, Wada I, Hasegawa K, Yoriyuzi T, Tremblay LO, et al. 2001. *EMBO Rep.* 2:415–22
143. van Anken E, Romijn EP, Maggioni C, Mezghrani A, Sitia R, et al. 2003. *Immunity* 18:243–53
144. Frenkel Z, Gregory W, Kornfeld S, Lederkremer GZ. 2003. *J. Biol. Chem.* 278:34119–24
145. Kitzmuller C, Caprini A, Moore SE, Frenoy JP, Schwaiger E, et al. 2003. *Biochem. J.* 376:687–96
146. Ermonval M, Duvet S, Zonneveld D, Cacan R, Buttin G, Braakman I. 2000. *Glycobiology* 10:77–87
147. Cabral CM, Choudhury P, Liu Y, Sifers RN. 2000. *J. Biol. Chem.* 275:25015–22
148. Kornfeld S. 1987. *FASEB J.* 1:462–68
149. Ghosh P, Dahms NM, Kornfeld S. 2003. *Nat. Rev. Mol. Cell Biol.* 4:202–12
150. Dahms NM, Hancock MK. 2002. *Biochim. Biophys. Acta* 1572:317–40
151. Bao M, Booth JL, Elmendorf BJ, Canfield WM. 1996. *J. Biol. Chem.* 271: 31437–45
152. Raas-Rothschild A, Cormier-Daire V, Bao M, Genin E, Salomon R, et al. 2000. *J. Clin. Investig.* 105:673–81
153. Hauri HP, Kappeler F, Andersson H, Appenzeller C. 2000. *J. Cell Sci.* 113: 587–96
154. Appenzeller C, Andersson H, Kappeler F, Hauri HP. 1999. *Nat. Cell Biol.* 1: 330–34
155. Zhang B, Cunningham MA, Nichols WC, Bernat JA, Seligsohn U, et al. 2003. *Nat. Genet.* 34:220–25
156. Fiedler K, Simons K. 1994. *Cell* 77: 625–26
157. Fullekrug J, Scheiffele P, Simons K. 1999. *J. Cell Sci.* 112:2813–21
158. Hara-Kuge S, Ohkura T, Seko A, Yamashita K. 1999. *Glycobiology* 9: 833–39

159. Neve EP, Svensson K, Fuxe J, Pettersson RF. 2003. *Exp. Cell Res.* 288:70–83
160. Nufer O, Mitrovic S, Hauri HP. 2003. *J. Biol. Chem.* 278:15886–96
161. Sato K, Nakano A. 2002. *Mol. Biol. Cell* 13:2518–32
162. Schrag JD, Procopio DO, Cygler M, Thomas DY, Bergeron JJ. 2003. *Trends Biochem. Sci.* 28:49–57
163. Bugg TD, Brandish PE. 1994. *FEMS Microbiol. Lett.* 119:255–62
164. Messner P, Schaffer C. 2003. *Fortschr. Chem. Org. Naturstoffe* 85:51–124
165. Young NM, Brisson JR, Kelly J, Watson DC, Tessier L, et al. 2002. *J. Biol. Chem.* 277:42530–39
166. Lechner J, Wieland F. 1989. *Annu. Rev. Biochem.* 58:173–94
167. Moens S, Vanderleyden J. 1997. *Arch. Microbiol.* 168:169–75
168. Schaffer C, Graninger M, Messner P. 2001. *Proteomics* 1:248–61
169. Parodi AJ. 1993. *Glycobiology* 3:193–99
170. McConville MJ, Mullin KA, Ilgoutz SC, Teasdale RD. 2002. *Microbiol. Mol. Biol. Rev.* 66:122–54
171. Vallee F, Lipari F, Yip P, Sleno B, Her-scovics A, Howell PL. 2000. *EMBO J.* 19:581–88
172. Deleted in proof
173. Booth C, Koch LE. 1990. *Cell* 59:729–37
174. Tatu U, Helenius A. 1997. *J. Cell Biol.* 10:555–65
175. Meunier L, Usherwood YK, Chung KT, Hendershot LM. 2002. *Mol. Biol. Cell* 13:4456–69
176. Kaufman RJ. 1999. *Genes Dev.* 13:1211–33
177. Ng DT, Spear ED, Walter P. 2000. *J. Cell Biol.* 150:77–88
178. Mori K. 2003. *Traffic* 4:519–28
179. Aebi M, Gassenhuber J, Domdey H, Heesen ST. 1996. *Glycobiology* 6:439–44
180. Stagljar I, te Heesen S, Aebi M. 1994. *Proc. Natl. Acad. Sci. USA* 91:5977–81
181. Reiss G, te Heesen S, Zimmerman J, Robbins PW, Aebi M. 1996. *Glycobiology* 6:493–98
182. Oriol R, Martinez-Duncker I, Chantret I, Mollicone R, Codogno P. 2002. *Mol. Biol. Evol.* 19:1451–63



CONTENTS

THE EXCITEMENT OF DISCOVERY, <i>Alexander Rich</i>	1
MOLECULAR MECHANISMS OF MAMMALIAN DNA REPAIR AND THE DNA DAMAGE CHECKPOINTS, <i>Aziz Sancar, Laura A. Lindsey-Boltz, Keziban Ünsal-Kaçmaz, Stuart Linn</i>	39
CYTOCHROME C -MEDIATED APOPTOSIS, <i>Xuejun Jiang, Xiaodong Wang</i>	87
NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY OF HIGH-MOLECULAR-WEIGHT PROTEINS, <i>Vitali Tugarinov, Peter M. Hwang, Lewis E. Kay</i>	107
INCORPORATION OF NONNATURAL AMINO ACIDS INTO PROTEINS, <i>Tamara L. Hendrickson, Valérie de Crécy-Lagard, Paul Schimmel</i>	147
REGULATION OF TELOMERASE BY TELOMERIC PROTEINS, <i>Agata Smogorzewska, Titia de Lange</i>	177
CRAWLING TOWARD A UNIFIED MODEL OF CELL MOBILITY: Spatial and Temporal Regulation of Actin Dynamics, <i>Susanne M. Rafelski, Julie A. Theriot</i>	209
ATP-BINDING CASSETTE TRANSPORTERS IN BACTERIA, <i>Amy L. Davidson, Jue Chen</i>	241
STRUCTURAL BASIS OF ION PUMPING BY CA-ATPASE OF THE SARCOPLASMIC RETICULUM, <i>Chikashi Toyoshima, Giuseppe Inesi</i>	269
DNA POLYMERASE , THE MITOCHONDRIAL REPLICASE, <i>Laurie S. Kaguni</i>	293
LYSOPHOSPHOLIPID RECEPTORS: Signaling and Biology, <i>Isao Ishii, Nobuyuki Fukushima, Xiaoqin Ye, Jerold Chun</i>	321
PROTEIN MODIFICATION BY SUMO, <i>Erica S. Johnson</i>	355
PYRIDOXAL PHOSPHATE ENZYMES: Mechanistic, Structural, and Evolutionary Considerations, <i>Andrew C. Eliot, Jack F. Kirsch</i>	383
THE SIR2 FAMILY OF PROTEIN DEACETYLASES, <i>Gil Blander, Leonard Guarente</i>	417
INOSITOL 1,4,5-TRISPHOSPHATE RECEPTORS AS SIGNAL INTEGRATORS, <i>Randen L. Patterson, Darren Boehning, Solomon H. Snyder</i>	437
STRUCTURE AND FUNCTION OF TOLC: The Bacterial Exit Duct for Proteins and Drugs, <i>Vassilis Koronakis, Jeyanthi Eswaran, Colin Hughes</i>	467
ROLE OF GLYCOSYLATION IN DEVELOPMENT, <i>Robert S. Haltiwanger, John B. Lowe</i>	491

STRUCTURAL INSIGHTS INTO THE SIGNAL RECOGNITION PARTICLE, <i>Jennifer A. Doudna, Robert T. Batey</i>	539
PALMITOYLATION OF INTRACELLULAR SIGNALING PROTEINS: Regulation and Function, <i>Jessica E. Smotryst, Maurine E. Linder</i>	559
FLAP ENDONUCLEASE 1: A Central Component of DNA Metabolism, <i>Yuan Liu, Hui-I Kao, Robert A. Bambara</i>	589
EMERGING PRINCIPLES OF CONFORMATION-BASED PRION INHERITANCE, <i>Peter Chien, Jonathan S. Weissman, Angela H. DePace</i>	617
THE MOLECULAR MECHANICS OF EUKARYOTIC TRANSLATION, <i>Lee D. Kapp, Jon R. Lorsch</i>	657
MECHANICAL PROCESSES IN BIOCHEMISTRY, <i>Carlos Bustamante, Yann R. Chemla, Nancy R. Forde, David Izhaky</i>	705
INTERMEDIATE FILAMENTS: Molecular Structure, Assembly Mechanism, and Integration Into Functionally Distinct Intracellular Scaffolds, <i>Harald Herrmann, Ueli Aebi</i>	749
DIRECTED EVOLUTION OF NUCLEIC ACID ENZYMES, <i>Gerald F. Joyce</i>	791
USING PROTEIN FOLDING RATES TO TEST PROTEIN FOLDING THEORIES, <i>Blake Gillespie, Kevin W. Plaxco</i>	837
EUKARYOTIC mRNA DECAPPING, <i>Jeff Collier, Roy Parker</i>	861
NOVEL LIPID MODIFICATIONS OF SECRETED PROTEIN SIGNALS, <i>Randall K. Mann, Philip A. Beachy</i>	891
RETURN OF THE GDI: The GoLoco Motif in Cell Division, <i>Francis S. Willard, Randall J. Kimple, David P. Siderovski</i>	925
OPIOID RECEPTORS, <i>Maria Waldhoer, Selena E. Bartlett, Jennifer L. Whistler</i>	953
STRUCTURAL ASPECTS OF LIGAND BINDING TO AND ELECTRON TRANSFER IN BACTERIAL AND FUNGAL P450S, <i>Olena Pylypenko, Ilme Schlichting</i>	991
ROLES OF N-LINKED GLYCANS IN THE ENDOPLASMIC RETICULUM, <i>Ari Helenius, Markus Aebi</i>	1019
ANALYZING CELLULAR BIOCHEMISTRY IN TERMS OF MOLECULAR NETWORKS, <i>Yu Xia, Haiyuan Yu, Ronald Jansen, Michael Seringhaus, Sarah Baxter, Dov Greenbaum, Hongyu Zhao, Mark Gerstein</i>	1051