Review

N-Glycan-based ER Molecular Chaperone and Protein Quality Control System: The Calnexin Binding Cycle

Lydia Lamriben, Jill B. Graham, Benjamin M. Adams and Daniel N. Hebert

Department of Biochemistry and Molecular Biology, Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, MA 01003, USA

Corresponding author: Daniel N. Hebert, dhebert@biochem.umass.edu

Abstract

Helenius and colleagues proposed over 20-years ago a paradigm-shifting model for how chaperone binding in the endoplasmic reticulum was mediated and controlled for a new type of molecular chaperone— the carbohydrate-binding chaperones, calnexin and calreticulin. While the originally established basics for this lectin chaperone binding cycle holds true today, there has been a number of important advances that have expanded our understanding of its mechanisms of action, role in protein homeostasis, and its connection to disease states that are highlighted in this review.

Keywords endoplasmic reticulum, molecular chaperones, N-glycans, protein folding, quality control

Cells are challenged by the need to efficiently fold an expansive assortment of proteins that span three orders of magnitude in amino acid number under a variety of conditions. Frequently, small single-domain proteins spontaneously acquire their native fold with high efficiency immediately upon translation (1). In contrast, larger complex proteins fold more slowly aided by the intervention of cellular factors such as molecular chaperones that are dedicated to assisting and optimizing the folding process (2). How molecular chaperones help with the maturation of proteins is of important biological and medical interest, as there are a growing number of diseases associated with protein folding defects involving difficult to fold multidomain proteins.

Molecular chaperones are defined as proteins that assist other proteins to reach their native structures through transient interactions (3). There are two important concepts to understand for chaperones: (i) how chaperones specifically recognize non-native structures for a large range of client proteins; and (ii) how chaperones control their binding cycle so substrates are transiently bound then released. Protein folding involves sequestration of hydrophobic residues in the core of the protein (or membrane) and the exposure of hydrophilic side chains to the aqueous environment. Classical chaperones from the Hsp60, 70 and 90 families assist proteins by binding to substrates possessing aqueous-exposed hydrophobic domains that provide a signal indicating that the protein is either a protein folding intermediate, misfolded or part of an unassembled multimer (4). In this way, a chaperone can specifically differentiate the folding stages for a large range of proteins. The substrate binding affinities of Hsp60s, 70s and Hsp90s are controlled by adenine nucleotides. The forward progression of the chaperone binding cycle is mediated by the form of the bound adenine nucleotide that alternates the conformation of the chaperone.
The Calnexin Binding Cycle

from its high affinity bound state to the low affinity released state.

An estimated one-third of the proteome is targeted to the secretory pathway in mammalian cells and these proteins fold in the endoplasmic reticulum (ER) (5). The ER of metazoans contains Hsp70 (BiP/GRP78) and 90 (GRP94) family members, but lack an Hsp60 component. The vast majority of membrane and soluble proteins targeted for the secretory pathway receive multiple N-linked carbohydrates as they are translated and translocated into the ER (6,7). A molecular chaperone network of factors that binds carbohydrates further assists the maturation of these proteins (8–10).

N-linked glycans are originally transferred by oligosaccharyltransferase (OST) as a Glc₃Man₉GlcNAc₂ appendage preassembled on a dolichol phosphate that resides in the ER membrane (11) (Figure 1). Once attached to the protein, the glycan is rapidly restructured by ER resident glycosidases and a transferase to a glycoform that is dependent upon the structure, age and the location of the protein within the ER. The composition of glycans act as protein fitness tags that signal the status and age of the protein to which they are attached (8–10). Central to the implementation of this glyco-code or signal is a specialized carbohydrate-binding chaperone system that supports the binding of the lectin chaperones, calnexin and calreticulin, specifically to monoglucosylated glycans. Therefore, binding to the ER lectin chaperones is directly controlled by the composition of the glycan. At first glance, the ER lectin chaperone system provides a completely different framework for assisting the maturation of proteins that traverse the secretory pathway by binding the exposed hydrophilic modification rather than misplaced exposed hydrophobic side chains on a maturing polypeptide. In this review, we will describe the discovery, mechanism and role of this important ER lectin chaperone system in cellular homeostasis and its association with disease states.

The Discovery of the Calnexin Binding Cycle

Understanding the mechanism of the lectin binding cycle provided a paradigm shift in our thinking on how an alternative chaperone system could operate and assist in the cellular folding process. It also offered an explanation for the

![Figure 1: Structure of N-linked glycan.](image_url)

N-linked glycans are transferred en bloc to asparagine residues in the sequon Asn-X-Ser/Thr, where X denotes any amino acid except for Pro. The precursor structure is composed of three glucose (blue circles), nine mannoses (green circles) and two N-acetylglucosamines (blue squares). Glycosidic bonds are noted.

longstanding questions regarding the changes in composition of glycans as proteins progress through the secretory pathway. Why build up an elaborate glycan structure in the ER just to dismantle it and then build it up again in the Golgi? This innovative carbohydrate-binding chaperone model first proposed by Ari Helenius and his colleagues at Yale University in 1994 was based on integrating a number of seemingly disparate but important observations in the literature with seminal results from his own lab (12,13).

Degen and Williams observed in 1991 that a protein of 88 kDa associated with major histocompatibility class I heavy chain shortly after its translocation into the ER (14). This heavy chain-associated protein was initially termed p88 and it was insightfully proposed that it might help retain class I in the ER until it associated with β2-microglobulin and peptide, prior to its exit for
peptide presentation at the cell surface. That same year, Bergeron, Thomas and colleagues identified a number of ER membrane phosphoproteins from canine rough-ER derived microsomes (15). One of these proteins was homologous to the soluble ER calcium-binding protein, calreticulin, and was named calnexin. Calnexin and p88 were later demonstrated to be identical (16). Bergeron, Thomas and colleagues later demonstrated that a number of abundantly expressed glycoproteins (α1-antitrypsin, α1-antichymotrypsin, transferrin, C3, apoB-100 and α-fetoprotein) co-immunoprecipitated with calnexin in HepG2 cells and the glycosylation inhibitor tunicamycin abolished their interaction (17). Non-glycosylated albumin did not bind calnexin. These results lead to the proposal that calnexin may play a role in the quality control specifically of glycoproteins.

Earlier studies from Parodi and Cazzulo identified glucosylation activity initially in *Trypanosoma cruzi*, a parasitc protozoan that assembles and transfers unglucosylated Man9GlcNAc2 glycans onto nascent chains in the ER (18). Glucose was transferred onto the high mannose glycans by using uridine disphosphate-glucose (UDP-Glc) as a substrate. The transferring enzyme was later purified from rat liver microsomes and named UDP-Glc: glycoprotein glucosyltransferase (UGGT) (19). Purified UGGT was found to modify denatured substrates more efficiently than native substrates, providing a possible link between the structure of the maturing protein and the glycan composition (20). The presence of a reglucosylation activity that adds a single glucose to high mannose moieties in the ER helped to provide an explanation for an earlier observation with a temperature-sensitive mutant of the vesicular stomatitis virus, ts045, that encoded a mutant form of the viral G protein (21). VSV G existed persistently in a monoglucosylated state at the non-permissive temperature but when the temperature was reduced to the permissive temperature, the glucose was removed and the protein exited the ER, once again associating the glycosylation state of a protein with the structure of the polypeptide.

A major advance toward understanding the binding properties of calnexin was discovered by Hammond et al. who showed that binding to calnexin also required glucosidase trimming as interactions with calnexin were blocked by pre-treatment with the glucosidase inhibitors castanospermine or 1-deoxynojirimycin that supported the accumulation of glycoproteins in tri-glycosylated glycoforms (12). This indicated that not only were glycans necessary for binding to calnexin but also these glycans required some level of glucose trimming. Helenius and colleagues reconciled these and earlier results into a central model that described how the calnexin binding cycle is controlled by the composition of the glycan (12,13,22). The terminal glucose is initially trimmed in the ER by the membrane protein α-glucosidase I, followed by the trimming of the second glucose by the soluble heterodimeric glucosidase, α-glucosidase II, to generate the monoglucosylated substrate for calnexin. Binding to calnexin persists until the final glucose is removed, once again by α-glucosidase II. If the released substrate folds properly to its native state, it exits the ER for the Golgi in COPII vesicles. However, if it continues to display a non-native configuration, it is recognized by UGGT, which reglucosylates it to regenerate a monoglucosylated protein that can rebind to calnexin until it is deemed properly folded by UGGT. This model provides an explanation for: (i) how a carbohydrate-binding chaperone cycle can be mediated; (ii) the persistent presence of a non-native protein in the monoglucosylated state; and (iii) the important quality control role of the reglucosylation activity of UGGT in the ER.

The calnexin binding cycle model was initially tested using an *in vitro* translation system combined with rough-ER derived microsomes that permitted the accumulation of substrates in their tri-, di-, mono- and unglycosylated states to demonstrate that calnexin bound specifically to monoglucosylated proteins (23). Evidence was also provided to support a central tenet of the model that reglucosylation in the ER could direct rebinding to calnexin. The calnexin binding cycle was expanded to include the soluble paralogue of calnexin, calreticulin, that also bound monoglucosylated substrates (24). This novel method of chaperone binding shifted the focus to the glycan in directing the maturation, binding of chaperones and the trafficking of glycoproteins in the early secretory pathway.

### The ER Lectin Chaperone Network

#### The glucosidases

The deglucosylation events in the ER occur in a controlled sequential manner which is initiated by α-glucosidase
I that cleaves the outer most α-1,2-linked glucose (11) (Figures 1 and 3). α-Glucosidase I is a type II single-pass transmembrane protein with a large globular luminal portion that contains the catalytic domain and a short N-terminal cytoplasmic tail (25). The crystal structure of the soluble luminal domain of Saccharomyces cerevisiae α-glucosidase I, Cwh41p, has been solved (26,27). Human α-glucosidase I and Cwh41p share 24% sequence identity overall and their catalytic domains share ∼45%, thus Cwh41p may be used to model the human α-glucosidase I catalytic activity and substrate binding properties. Two acidic residues within the C-terminal domain of the Cwh41p catalytic site confer the protein’s catalytic activity and a number of conserved aromatic residues contribute to the protein’s highly specific substrate binding properties. α-Glucosidase I isolated from mouse fibroblasts was found in a complex with Sec61 as observed in proteomics studies of translocon-associated factors, which supports early intervention by the enzyme in the protein folding pathway (28). α-Glucosidase I null mutations in hamster cells, as well as in S. cerevisiae, are tolerated (29,30). Once the terminal glucose has been trimmed, the di-glucosylated glycan becomes a substrate for α-glucosidase II, a soluble ER resident enzyme that specifically cleaves the α-1,3-linked glucose moieties to generate deglucosylated glycans (11). Human α-glucosidase II is comprised of a large α-subunit (∼100 kDa) and a smaller β-subunit (∼50 kDa) that associate non-covalently (31,32). The α-subunit possesses catalytic activity and is retained in the ER by its association with the β-subunit, which contains an ER retention (KDEL) sequence (33). The β-subunit is not required for the catalytic activity of the enzyme; however, it appears necessary for its maturation, solubility and stability. Additionally, the β-subunit contains a mannose 6-phosphate receptor homology (MRH) domain, which has been proposed to bind the terminal mannose on the trimmed C-branch of glycans as well as the B-branch in order for the enzyme to efficiently act on the A-branch glucose (34,35). For translocating glycoproteins emerging into the ER lumen, the first glycan helps to recruit α-glucosidase II through binding to its MRH domain, supporting the more efficient cleavage of additional C-terminal N-glycans on the nascent chain (36). The function of α-glucosidase II depends on the presence as well as number of mannoses present on the glycan. The enzymatic activity of α-glucosidase II on the A-branch of the glycan is strongly influenced by the presence of mannoses on the adjacent B- and C-branches. In vitro data demonstrated a decrease in enzymatic activity that correlated to a decrease in mannose content on the targeted glycan (37). Like α-glucosidase I, α-glucosidase II null mutations are tolerated in hamster and mouse immortalized cell lines (38).

Calnexin and calreticulin

The type I membrane protein calnexin and its soluble parologue calreticulin share a similar general organization (Figure 2). Each is comprised of an N-terminal globular domain, a central proline-rich domain and a C-terminal domain. Human calnexin and calreticulin share 39% overall sequence identity (40).

The N-termini of the two proteins fold into globular domains consisting of anti-parallel β-sheets that create a single carbohydrate-binding site (39). The carbohydrate-binding site supports binding specifically to monoglucosylated glycans (13). Isolated calreticulin binds monoglucosylated glycans with micromolar affinity (41). The crystal structure of calnexin shows that the globular domain also contains a single bound calcium ion that accounts for the high affinity low capacity binding of calcium to the chaperones (39) (Figure 2). Calcium is important for stabilizing the lectin chaperones and regulating their binding to substrates (42,43). ATP also appears to modulate the conformation of the lectin chaperones, as they become more resistant to proteases in its presence, suggestive of them reaching more compact conformations (44,45). Recently, computational and mutagenesis approaches have been used to identify residues in the globular domain of calreticulin that contribute to ATP binding (45). ATP destabilized the substrate-calreticulin complex, indicative of it playing a role in regulating substrate binding. High affinity calcium binding by calreticulin is required for nucleotide binding and chaperone stabilization (45). Altogether the globular domain contains sites for binding to carbohydrates located on the substrate, and regulatory co-factors such as ATP and calcium.

The central Pro-rich or P-domain creates an arm that extends away from the N-terminal lectin domain. The P-domain of calnexin (140Å) is organized into a series of sequential motifs, comprised of four consecutive copies of
Figure 2: Structural and domain organization of calnexin and calreticulin. A) Linear cartoon representations of calnexin and calreticulin. Both proteins possess a cleavable ER-targeting N-terminal signal sequence (green), a lectin domain (orange) and a C-terminal domain (purple). Calnexin possesses a single-pass transmembrane domain (black) that couples calnexin to the ER membrane and supports a cytoplasmic C-terminal domain. Calreticulin is a soluble protein that is retained in the ER through its C-terminal KDEL sequence. B) Surface and ribbon representations of the crystal structure of the calnexin luminal domain (PDB: 1JHN) (39). The globular lectin domain (orange) forms a binding pocket to accommodate monoglucosylated clients and contains a calcium-binding site, with a single Ca\(^{2+}\) ion shown in cyan. The P-domain (gray) protrudes from the globular domain into a hook-like shape. Cyclophilin B and ERp57 bind at the tip of the P-domain.
motif 1 followed by four copies of motif 2 (46). The calreticulin P-domain is shorter than that of calnexin as it involves two sets of three consecutive repeats rather than four. Both P-domains are the site of interaction with two folding factors: ERp57, a protein disulfide isomerase; and cyclophilin B (CyB), a peptidyl-prolyl isomerase. Like PDI, ERp57 consists of four thioredoxin domains of which the two positioned at each termini possess the active CXXC motifs, while the central thioredoxin folds lack enzymatic activity (47). ERp57 contains a P-domain-binding region that localizes it to the vicinity of lectin chaperone substrates so it can assist disulfide bond formation and rearrangement of lectin chaperone-associated substrates. The interaction between the lectin chaperones and ERp57 is transient and it occurs in the absence of a client substrate (48,49). The region at the tip of the P-domain to which ERp57 binds overlaps with that of CyB binding, making these interactions mutually exclusive (50).

Calnexin and calreticulin also appear to bind client proteins in a glycan-independent manner (51). In vitro studies show that they can inhibit the aggregation of non-glycosylated substrates (52). The peptide-binding site seems to be located in the vicinity of the glycan binding site within the globular domain (53). While the affinities of calreticulin for glycosylated and non-glycosylated substrates appear to be similar, the kinetics are distinct. Binding to glycosylated clients is more rapid. Glycosylated and non-glycosylated clients support the formation of open and closed conformations, respectively, where the closed conformation involves the clamping down of the P-domain on the substrate stabilizing substrate interactions. The co-factor ERp57 also induces the closed conformation of calreticulin as probed using FRET measurements with purified components.

The C-terminal regions are divergent but they both contain ER retention signals. The membrane protein calnexin possesses a single C-terminal transmembrane region and a short cytoplasmic tail. Oriented in the cytoplasm is a di-Arg sequence that aids in ER retention, along with a number of phosphorylated residues that will be discussed below. At its C-terminus, the soluble calreticulin has a canonical KDEL sequence that supports ER retention and retrieval. This region on calreticulin also contains a series of acidic residues that support the low affinity ($K_d = 1 – 2 \text{mM}$) and high capacity ($\sim 20 \text{ mol Ca}^{2+} \text{ per mol protein}$) binding of calcium (54,55). Calreticulin acquires its calcium buffering properties through the C-terminal domain. The tail of calreticulin is intrinsically unstructured and only acquires a rigid and folded structure when bound to calcium.

**The glucosyltransferase protein folding sensor**

Non-native unglucosylated proteins are reglucosylated by UDP-gl: glycoprotein glucosyltransferase 1 (UGGT1) to regenerate monoglucosylated side chains that can support rebinding to calnexin and calreticulin (13,56). Once the glycoprotein is in a mono-glucosylated state, it also becomes a substrate for $\alpha$-glucosidase II, which competes with UGGT1 with opposing effects (57) (Figure 3). This positions UGGT1 to be a central gatekeeper in the ER that decides whether glycoproteins traffic onto the Golgi and beyond, or are retained in the ER for further assistance.

Human UGGT1 is a soluble ($\sim 170 \text{kDa}$) resident ER enzyme. At its C-terminus, it possesses a catalytic glucosyltransferase domain. This domain contains a DxD motif that is required for its catalytic activity (58,59). In an in vitro study, purified rat UGGT1 required calcium to transfer glucose onto deglucosylated glycoprotein acceptors, suggesting calcium dependence for the catalytic activity of UGGT1 (19). Similar to $\alpha$-glucosidase II, the catalytic activity of UGGT1 is more effective on glycans with high mannose content (20).

The N-terminal $\sim 80\%$ of UGGT1 contains what is proposed to be the important protein folding sensor domain. Bioinformatics analysis predicts that the N-terminal region is comprised of three consecutive thioredoxin-like domains (60). Thioredoxin folds are commonly associated with protein disulfide isomerase family members that play important roles in protein folding and quality control (47). However, none of the thioredoxin-like domains of UGGT1 contain the catalytic CXXC motif that shuttles electrons in disulfide formation, reduction or rearrangement reactions (60). Instead, these thioredoxin-like domains appear to act like the catalytically inactive domains found in many of the protein disulfide isomerase family members including PDI and ERp57. The three-dimensional structure of the third thioredoxin-like domain of UGGT1 has been solved and it involves a central $\beta$-sheet surrounded by a number of short $\alpha$-helices. It also possesses a distinct hydrophobic patch.
that might be involved in the recognition of folding intermediates or misfolded substrates. However, more studies are required to test and confirm this proposed function.

UGGT1 also has a homologue UGGT2 that shares 55% sequence identity with UGGT1 (61). The C-terminal glucosyltransferase domain appears to be active when swapped with the catalytic domain of UGGT1 or when tested in vitro against synthetic substrates (62,63). Currently, no biological role has been assigned nor natural substrates of UGGT2 have been identified.

**Functional Roles of the Lectin Chaperone Binding Cycle**

The lectin chaperones are multifunctional factors playing important roles in optimizing the efficiency of the protein folding program, minimizing deleterious side reactions such as aggregation, retaining immature or misfolded cargo in the ER, and possibly targeting aberrant proteins for degradation. Several experimental strategies have been used to tease out the role of the lectin chaperone cycle and their associated factors in the maturation and quality control of ER targeted client proteins. These approaches include following the maturation of lectin chaperone substrates in the absence of their assistance and comparing it to the natural pathway by using chemical inhibitors or cell lines that lack the individual components, as well as studies with purified components.

A main function of the lectin chaperone binding cycle is to promote efficient folding and assembly of glycoprotein substrates to their native states. Inhibition of α-glucosidases I and II with glucose analogues castanospermine (CST) or N-butyl-1-deoxynojirimycin (DNJ) circumvents the lectin chaperone binding cycle by accumulating substrates in their tri-glucosylated state. This is commonly associated with improper folding and reduced trafficking of glycoproteins from the ER as has been demonstrated for a number of substrates including hemagglutinin, neuraminidase, vesicular stomatitis G protein, class I histocompatibility complex and tyrosinase (64–69). Validated targets of calnexin and calreticulin assistance include endogenous and pathogenic membrane and soluble glycoproteins.

The use of glucosidase inhibitors simultaneously bypasses both lectin chaperones. To delineate the role of each chaperone individually, knockout cell lines have been used that are separately missing one of the chaperones or co-factors. Calnexin deficiencies led to more error prone folding of hemagglutinin in cnx−/− T lymphoblastoid cells than observed in crt−/− mouse embryonic fibroblasts.
The Calnexin Binding Cycle

Traffic 2016; 17: 308–326

315

(MEFs) (70). While hemagglutinin is known to interact with both chaperones (71), these results are indicative of hemagglutinin's reliance on calnexin for its proper folding. In contrast, studies in crr−/− cells have demonstrated the requirement for calreticulin in the proper processing and trafficking of MHC class I heavy chain and the bradykinin receptor (72,73).

The interaction between calnexin or calreticulin with substrates also promotes glycoprotein folding and degradation through its associated oxidoreductase, ERp57; or the peptidyl-prolyl cis/trans isomerase, CyB (48,50,74). As an oxidoreductase isolated ERp57 has been shown to catalyze the oxidation, isomerization and reduction of disulfide bonds (75). The catalytic activity of ERp57 is enhanced when it is associated with calnexin or calreticulin (76). The direct association of ERp57 with calnexin and calreticulin has been suggested to be important for bringing ERp57 and its substrates together, thereby enhancing local concentration of the isomerase for the catalysis of disulfide bonds. ERp57 knockdown was found to specifically delay secretion of the glycosylated substrates (77). Cellular ERp57 appears to oxidize HA post-translationally but not co-translationally, as in the absence of ERp57 only the post-translational oxidation processing of HA was compromised (78). For further discussion of the role of ERp57 in protein folding and quality control please see the article by Ellgaard et al., in this issue (79).

Much less is known about the roles of PPIs in the ER. CyB associates with the lectin chaperones through their P-domain on a site that appears to overlap with the ERp57 binding site (50). The immunosuppressive drug cyclosporine A, an inhibitor of cyclophilins, has been used to demonstrate that ER CyB helps prepare substrates for ERAD, as it delayed the turnover of glycosylated ERAD substrates (80). Isomerases associated with the lectin chaperones can assist the processing and trafficking of substrates associated with calnexin and calreticulin.

After release from lectin chaperones by α-glucosidase II trimming, unglucosylated substrates that do not reach their native state may be reglucosylated by UGGT1 (Figure 3). The re-established monoglucosylated state directs rebinding to calnexin and calreticulin. Using a cell based reglucosylation assay that relies on the use of a mutant cell line that transfers truncated carbohydrates to nascent chains, UGGT1 was found to modify a wide spectrum of proteins in live cells (81–83). The lysosomal protein prosaposin was determined to be an obligate substrate of UGGT1, as in its absence prosaposin poorly reached its proper locations but instead accumulated in intracellular aggregates (82). By supporting persistent chaperone binding, reglucosylation promotes substrate solubility and prevents aggregation (84). There is tremendous variability in the necessity of reglucosylation to aid the maturation of secretory pathway client proteins. Deletion of UGGT1 in mouse embryonic fibroblast cells has shown that there are three general types of glycoproteins (85). Class I proteins, such as VSV G, use only one round of lectin chaperone binding, leading to no alteration in folding rate or efficiency in the absence of UGGT1. Class II proteins, such as BACE501, require multiple rounds of binding, leading to early release from the ER and a decrease in folding efficiency in uggt1−/− cells. Class III proteins, such as hemagglutinin, exhibit a decreased rate of release from calnexin when UGGT1 was deleted, possibly because UGGT1 itself is required for their structural maturation or oligomerization. These results suggest that the role of the lectin chaperone binding cycle in glycoprotein folding is highly substrate dependent.

Persistent calnexin and calreticulin binding, like classical chaperones, act to slow the folding process (64,71,86). Since chaperone binding is mediated by N-glycans, substrate binding occurs in a region-specific manner (71). For larger multidomain proteins such as influenza hemagglutinin and neuraminidase, chaperone binding helps to direct the molecular choreography of the cellular folding pathway (69,71,86). N-terminal glycans recruit early co-translational calnexin binding (87), which can delay the formation of intramolecular and intermolecular disulfide bonds to create an efficient folding route (69,71,86). In the absence of lectin chaperone binding, accelerated folding and oligomerization can favor misfolding and aggregation. Trapping proteins on the chaperones delays folding, oligomerization and trafficking (23,69,83). These results suggest the possibility that the genetic code encodes an additional layer of complexity that determines the positioning of N-linked glycans to help with the recruitment of chaperones and the optimization of the cellular folding process.
An important role for the lectin chaperones is retaining intermediates, misfolded substrate or unassembled multimers in the ER. ER retention of non-native clients enhances proper glycoprotein maturation and prevents non-functional and potentially harmful glycoproteins from trafficking out of the ER (88). Glucosidase inhibition blocks glycoprotein entry into the calnexin binding cycle and led to an increased rate of RNAse secretion, as did the deletion of glycosylation sites (36). Glucosidase inhibition has also been shown to increase export of non-native proteins from the ER in the case of MHC Class I molecules (73). The cell line, MI8-5 CHO, which transfers glycans with no glucosesto glycoproteins, has facilitated numerous studies of the lectin chaperone binding cycle as glycoproteins can only enter the calnexin and calreticulin binding cycle after glucosylation by UGGT1. Treatment with glucosidase inhibitors in this cell line traps substrates on the lectin chaperones and delays glycoprotein secretion, as in the case of α-1-antitrypsin and prosaposin (83). Calnexin and/or calreticulin binding can also retain in the ER the monomers of incompletely assembled oligomers, as is the case of soybean agglutinin (89). UGGT1 can reglucosylate the folded monomers, directing them to re-associate with lectin chaperones until complete assembly is achieved. Collectively, these studies suggest that the ER lectin chaperones aid in retaining non-native glycoproteins in the ER, thereby promoting correct folding and assembly prior to ER exit.

While some eukaryotic cell lines can survive the knockout of calnexin, calreticulin, ERp57 or UGGT1, the deletion of these factors in mice causes severe phenotypes (90). Calnexin knockout mice exhibited decreased survival rates within 48 hr after birth, and survivors were smaller and displayed motor disorders such as unstable gait and truncal ataxia, as well as a loss of myelinated nerve fibers (91). Calreticulin deletion in mice is embryonic lethal at 18 days due to heart defects (92). ERp57 deletion in mice is embryonic lethal at day 13.5, possibly due to increased STAT3 signaling (93). Finally, UGGT1 deletion is typically embryonic lethal at day 13 in mice, although some mice survive to birth (94). The number and diversity of glycoproteins that rely on the lectin binding cycle for proper folding underscores its importance. Some of the obligate substrates of the lectin chaperone network appear to be required for development and survival.

**ER Organization of the Lectin Chaperone Network**

The function of the carbohydrate-binding chaperone network is in part controlled by the localization of its components within the cell. While the ER is comprised of a single contiguous membrane and lumen, it is divided into a number of functional regions that help support the efficient progression of maturing proteins and a number of other cellular processes (95–97). At first glance, the ER is separated into two morphologically distinct regions: the perinuclear rough-ER decorated with ribosomes; and the smooth ER that lacks ribosomes but extends throughout most mammalian cells to create a number of functional regions. The lectin chaperones are dispersed throughout both distinct areas (98,99). Calnexin and calreticulin have also been found in locations outside the ER, suggestive of them playing roles beyond the confines of the ER boundaries (100).

Membrane-bound ribosomes directed to the ER by the signal recognition particle (SRP) co-translocationally translocate nascent chains into the ER through the proteinaceous channel comprised of Sec61. An assembly line of translocon-associated factors awaits the emerging nascent chain to create a privileged environment that optimizes the efficient maturation of both membrane and soluble client proteins (101). The spatial organization of the factors is thought to temporally control access to the nascent chain with two large multimeric complexes that mediate N-glycosylation (OST) and signal sequence cleavage (signal sequence peptidase complex) having early access to maturing polypeptides. Calnexin and calreticulin associate with both ribosome-arrested polypeptides, as well as actively translating and translocating nascent chains (86,87). This indicates that the membrane-integrated α-glucosidase I that is a translocon-associated protein (28), and soluble α-glucosidase II also have early access to nascent chains. A single round of binding to the lectin chaperones appears to occur during translation as arrested chains were unable to be reglucosylated by UGGT1 unless released from the translocon by puromycin (81). Proteomics and morphological results indicated that UGGT1 does not accumulate in the rough ER (98,99,102). While the polysome-translocon complexes ensure spatial separation of nascent chains to minimize unproductive contacts that might lead to aggregation (103,104), binding
to the lectin chaperones provides an additional layer of protection during this vulnerable stage of maturation (64,65).

Maturing proteins progress from the rough to the smooth ER, where they continue to fold and assemble (if applicable). In the smooth or transitional ER, clients are interrogated or subjected to quality control evaluations. UGGT1 plays a central role in the evaluation process and appears to be located near ER exit sites as determined by immunolocalization as well as functional studies (102,105). Once a protein is reglucosylated, it is proposed to move from the calnexin-free vicinity of the ER exit sites toward a calnexin and calreticulin-rich region of the ER that required ~20 min for the thermosensitive tsO45 mutant of VSV G protein (105). Interestingly, while tsO45 VSV G is ER retained at the non-permissive temperature, its mobility was not retarded when compared to the mobility of the protein at the permissive temperature (106). This may be caused by weak and dynamic interactions between the glycoprotein and the lectin chaperones (micromolar) (41). The properly folded, deglucosylated and released substrate migrates to ER exit sites, is no longer recognized by UGGT1, and therefore permitted to leave the ER through COPII vesicles. Photobleaching studies with calreticulin-GFP; however, call into question the presence of a lectin chaperone-free zone as there was no region in the ER to which calreticulin-GFP did not have access (107).

Terminally misfolded or irreparably folded glycoproteins are efficiently reglucosylated by UGGT1 supporting persistent binding to the lectin chaperones (83). These clients are eventually targeted for turnover by the ER-associated protein degradation or ERAD process (108). Both calnexin and calreticulin have been found to accumulate upon proteasomal inhibition in a compartment with known ERAD factors such as ER Man I/MAN1B1, EDEM1, Derlin-1 and Os-9, as well as ERAD substrates (109,110). This compartment has been called the ERQC (ER quality control compartment) and has been proposed to be a concentrated area where ERAD takes place (Figure 4). As the ERQC lacks both UGGT1 and ERP57, it has been proposed that persistent binding to the lectin chaperones in the bulk ER may play a role in the delivery of aberrant substrates to the ERQC for subsequent dislocation and proteasomal degradation.

The ER directly contacts a number of organelles, including mitochondria (96) (Figure 4). Mitochondrial contact regions are termed mitochondria-associated ER membranes or MAM. As the ER is the major calcium store in the cell, it responds to metabolic needs through intracellular signals that can lead to rapid changes in free luminal calcium levels (111). ER calcium channels including the inositol 1,4,5-trisphosphate receptor and SERCA are enriched in the MAM where they help to control calcium exchange between the two organelles (112,113). The activity of SERCA is regulated by the lectin chaperones as they both inhibit SERCA-mediated calcium uptake into the ER (114,115). The SERCA activity appears to be modulated by the recruitment of the lectin chaperone-associated oxidoreductase ERP57 that reversibly controls the SERCA redox state and activity (116). The involvement of the lectin chaperones in SERCA regulation points to a role for calnexin and calreticulin in controlling fundamental cellular processes that extend beyond assisting protein maturation and quality control.
Figure 4: Cellular localization of the lectin chaperone network. In the rough ER (rER), which is characterized by the presence of ribosomes on the cytoplasmic side of the membrane and translocation machinery, α-glucosidase I and II sequentially trim glucose from newly added N-linked glycans generating monoglucosylated glycoproteins. Calnexin and calreticulin associate with actively translating and translocating proteins. As proteins continue to fold, they migrate to the smooth portion of the ER where they are interrogated by UGGT1 in the peripheral ER at ER exit sites (ERES). If proteins are deemed terminally misfolded, they are targeted for ERAD. Calnexin, calreticulin and misfolded proteins accumulate in the ERQC. The ER contacts the mitochondria at the mitochondria-associated membrane (MAM). Calnexin in particular is shuttled between the rER and the MAM by post-translational modifications based on the current needs of the cell. Calnexin is differentially phosphorylated (P — in yellow) on several residues within its C-terminal cytoplasmic tail, associating it with translating ribosomes and enriching it in the rER. Additionally, calnexin has been shown to be palmitoylated (in blue).

forms of post-translational modification that accentuate its many roles.

Calreticulin has also been found to act in locations outside the ER. While this proposal was originally met with skepticism, there is accumulating cell biological and functional evidence for calreticulin acting in the cytoplasm, the outer cell surface, and the extracellular matrix (129). Cytoplasmic calreticulin has been shown to bind to the cytoplasmic tail of integrins to stabilize ligand binding for calcium signaling and cell adhesion (130). Calreticulin has also been found on the outer surface of a variety of cells including melanocytes, fibroblasts, platelets, endothelial and apoptotic cells. Cell surface calreticulin supports cell spreading and migration (131,132). Calreticulin at the cell surface in association with phosphatidylserine acts as an ‘eat me’ signal to direct the phagocytosis of apoptotic or cancer cells by macrophages and neutrophils, as well as fibroblasts (133,134). Interestingly, the cell surface roles of calreticulin appear to contribute to its ability to accelerate the rate of cutaneous wound healing when added topically (135). Calnexin has also been found at the cell surface for a number of cells (136). These studies point toward an expanding influence of the lectin chaperones beyond the ER.

Translational Implications Involving the Lectin Chaperones

Calnexin and calreticulin bind and influence the maturation of a large number of proteins as they traffic through
the ER and many of these client proteins are associated with disease states. Modulation of the lectin chaperone binding cycle provides a potential avenue for therapeutic intervention for a variety of maladies. Modulation strategies can involve either the derailment of the maturation of proteins from pathogens, or the enhancement of essential endogenous activities that are associated with pathologies that involve diminished protein levels due to aberrant or inefficient maturation.

Enveloped viruses usurp the mammalian secretory pathway for the production of their membrane glycoproteins that serve essential functions in the viral life cycle. These viral glycoproteins rely on the endogenous glycosylation machinery and lectin chaperones to ensure their proper maturation. Circumventing the lectin chaperone cycle can lead to inefficient folding, and subsequent degradation of viral glycoproteins. Iminosugars that inhibit glucosidases have been explored as possible anti-viral agents for the treatment of enveloped viruses such as influenza, HIV and Dengue virus (137,138). Deoxynojirimycin (NB-DNJ) derivatives (Miglustat or Zavesca) and CST support the accumulation of glycoproteins in their tri-glucosylated state thereby abolishing the influence of the lectin chaperones on the maturation of the viral glycoproteins. Both hemagglutinin and neuraminidase fold inefficiently in the absence of calnexin and calreticulin assistance (64,69,71). Glucosidase inhibition has been shown to diminish influenza virus A replication in cell culture (139). The inhibition of glucosidase activity also caused structural alterations in Env or gp120, and inhibited HIV replication in cells (140,141). Clinical trials were initiated using NB-DNJ on patients with HIV but were halted due to side effects that included gastrointestinal distress, and the discovery that cataracts developed in rats upon NB-DNJ treatment (142). A DNJ derivative (N-9-methoxynonyl-DNJ or UV-4) is also a potent inhibitor of dengue virus infection in mice (138). Disrupting the production of mature and infectious virus by inhibiting the glycoprotein maturation process provides a potential mechanism for the development of broad-spectrum anti-viral agents directed toward enveloped viruses.

Cells are relatively tolerant of glucosidase inhibition under normal conditions. This includes treatment with chemical inhibitors, as well as the development of stable cell lines that lack either α-glucosidase I or II (30,38,143). In part, this is likely explained by backup compensating chaperone systems being in place; however, different outcomes can occur if cells are stressed and this phenomenon can be potentially exploited for therapeutic intervention. For instance, CST analogues have been shown to possess anti-tumor activity as iminosugar CST-derivatives inhibit the proliferation of breast cancer cells by promoting both cell cycle arrest and apoptosis (144). It is uncertain if this anti-tumor property is related to the disruption of the lectin chaperone binding cycle. Further investigation into the influence of glucosidase inhibitors on cancer cells is warranted.

There are a growing number of loss-of-functions diseases associated with proteins that traverse the secretory pathway (108). For this class of disease, the malady is often caused by a mutation in a secretory pathway client protein that disrupts the proper maturation of the protein in the ER resulting in its evaluation by the quality control process as aberrant and its subsequent targeting for degradation through the ER-associated degradation pathway. In some cases this involves an overzealous evaluation decision that supports the ER retention of a protein that possesses a slight alteration but is functional. Overexpression of calnexin has been shown to support the more efficient maturation or correction of some proteins in live cells including the pigment producing protein tyrosinase and the lysosomal protein glucocerebrosidase, proteins associated with albinism and Gaucher disease, respectively (145,146). A possible therapeutic strategy to extend cell culture findings to the clinic has been developed by the Kelly lab, which has used a thoughtful tactic of screening US Food and Drug Administration-approved drugs for their ability to augment the maturation of mutants associated with common lysosomal storage diseases in cell culture (147). They discovered that the L-type calcium channel inhibitors, Diltiazem and Verapamil, partially corrected the trafficking of two glucocerebrosidase mutations (N370S and L444P) in fibroblast cells from Gaucher patients. These drugs, approved for treating hypertension, increase the concentration of calcium in the ER by inhibiting ER ryanodine calcium efflux channels. Enhanced glucocerebrosidase maturation is proposed to be mediated by the higher calcium concentration supporting more effective assistance.
by the calcium-binding lectin chaperones calnexin and calreticulin. As there is a number of calcium-binding proteins in the ER, there is a number of activities expected to be influenced by increased calcium levels.

The ER can be viewed as a protein production assembly line where products entering the ER are folded and modified, before being packaged into COPII vesicles. In recent years, there has been a great increase in the interest in the use of secretory pathway products as biologics or biological compounds as therapeutic agents. These products include hormones, growth factors, enzymes, enzyme inhibitors, vaccines and monoclonal antibodies (148). One of the main hurdles for biologics is producing sufficient quantities of active proteins. The creation of cell lines that more efficiently produce properly folded secretory pathway client proteins could help to address this need. Cell lines with activated UPR (ATF4 or Xbp1 [with ER01 overexpression]) boosted the production of antibodies in CHO cells (149,150). Antibody secretion in CHO cells was also enhanced by PDI overexpression (151). Enhancing the involvement of calnexin and calreticulin might help the production of N-glycosylated biologics by accelerating their maturation within cells.

Myeloproliferative neoplasm is largely associated with mutations in Janus kinase 2 (JAK2) that increase signaling in the JAK/STAT pathway (152). Whole-exon sequencing was used to identify mutations found in the approximate one-third of patients that did not have a corresponding JAK2 mutation. Two independent studies discovered mutations in exon nine of calreticulin that were found in the majority of the non-mutated JAK2 myeloproliferative neoplasm patients tested (153,154). Interestingly, the most predominant mutations identified (deletion of 52 base pairs or insertion of 5 base pairs), which account for 85% of the mutations identified, both cause a similar reading frame shift that alters the C-terminus of calreticulin. Two alterations generated by the mutated C-terminus are: (i) the KDEL C-terminal ER retrieval/retention signal is absent; and (ii) many of the negatively charge residues that create the low affinity, high capacity calcium-binding property of calreticulin are either missing or changed to positive residues. These differences are expected to have a major impact on the localization and the important calcium signaling function of calreticulin. Further exploration is required to understand how these mutations are linked to chronic myeloid leukemia.

**Closing Remarks**

Major advances in our understanding of the protein maturation and quality control processes of the secretory pathway have occurred over the past two decades. Most notably, N-glycans have emerged as key maturation and quality control tags that signal the fate of the protein to which they are attached (10). Many details of the glyco-code have been solved and the players involved in encoding and decoding the protein fitness signals have been identified. Moreover, the glycan positioning on the maturing substrate has been recognized to play an important role in directing the molecular choreography for the protein maturation program of secretory pathway client proteins (71,86). Studies using additional substrates are now needed to elucidate how this additional layer of the genetic code, that dictates glycan positioning, helps to optimize the folding and maturation of glycoproteins.

The binding properties of the lectin and traditional molecular chaperones are seemingly the polar opposites-binding to attached hydrophilic modifications or hydrophobic patches on maturing or aberrant nascent chains, respectively. Bulky N-linked modifications are expected to be aqueous-exposed regardless of the conformation of protein so how can this modification be used to direct chaperone binding and evaluate protein conformation? The initial round of lectin binding, which is often co-translational, is determined by the level and rate of glucose trimming. α-Glucosidase II initiates binding as well as release by trimming the second and third glucose residues on the A-branch, respectively. Further studies are required to determine the mechanism of α-glucosidase II trimming and how its slow removal of the final glucose is regulated, as this plays a significant part in controlling the initial duration of lectin chaperone binding.

Though the lectin and traditional chaperone interactions are differentially mediated, the protein determinants that direct persistent chaperone binding or chaperone rebinding seem to be similar. Chaperone rebinding is directed by UGGT1, which appears to use a traditional chaperone-like recognition process to interrogate nascent
chains before reglucosylation, as UGGT1 modifies proteins with exposed hydrophobic residues (56). As UGGT1 is the central gatekeeper that determines if a protein exits the ER or is retained for further intervention, additional studies are required to understand how UGGT1 is responsible for evaluating the enormous number of proteins and conformations that exist in the ER lumen. Also of interest, is determining how misfolded proteins that are initially retained are eventually relinquished and targeted for ERAD. Are these proteins released after going through a certain number of reglucosylation-deglucosylation cycles or after a particular duration of time? Does trimming of A-branch mannoses by exo- or endomannosidase release a protein from reglucosylation? Does the spatial organization of the ER contribute to an iterative process for glycoprotein maturation and quality control? If so, how is the spatial organization maintained or shifted under stress conditions? While we have witnessed many advances in how glycans direct the trafficking of nascent chains over the past 20 years, future studies will continue to provide further resolution to these important fundamental cellular processes and how they can potentially be modulated for therapeutic intervention.

Acknowledgments

This work was supported by the National Institutes of Health under award number GM086874 (to D.N.H.); and a Chemistry-Biology Interface Predoctoral training grant (T32 GM008515 to L.L.).

References


95. Lynes EM, Simmen T. Urban planning of the endoplasmic reticulum (ER): how diverse mechanisms segregate the many functions of the ER. Biochim Biophys Acta 2011;1813:1893–1905.


